LEABHARLANN CHOLÁISTE NA TRÍONÓIDE, BAILE ÁTHA CLIATH | TRINITY COLLEGE LIBRARY DUBLIN
Ollscoil Átha Cliath | The University of Dublin

Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

The Application of Digital Signal Processing to the Modelling and Alleviation of Auditory Impairment.

Derek B. O'Reilly

Department of Electronic and Electrical Engineering
Trinity College
Dublin

Thesis submitted for the degree of Ph.D.

September, 2000

Declaration

I, the undersigned, declare that this thesis is entirely my own work, except where otherwise accredited, and that it has not been submitted for a degree at any other university or institution. I furthermore agree that the Library may lend or copy this thesis upon request.

Signed: Derek O'Reilly

Date: 16/2/2001

Summary

This work is concerned with the development of digital signal processing models of the unimpaired and impaired human auditory system with a view towards the development of signal processing strategies for the alleviation of auditory impairment.

The unimpaired auditory model consists of three primary sections modelling the outer, middle and inner ears, with particular emphasis on the inner ear to facilitate sensorineural impairment modelling. The outer and middle ear models are designed to superimpose the typical threshold of audibility curve on stimuli presented to the model, specifically providing attenuation for very low and high frequencies, thus setting the frequency range of audibility for the model. The inner ear model consists of several sub-models emulating the mechanical motion of the basilar membrane, the resonance of motion between the basilar and tectorial membranes, the mechanical to neural transduction of the inner hair cells (IHC) and feedback within the inner ear. Of particular importance is the tuning of the inner ear model arising from adaptive level dependent pseudoresonance, which is built up through the cascade of filter sections. The adaptive nature of the cascade depends on the stimulus level, the stimulus frequency content and the outer hair cell (OHC) activity within the model. The net result of these interactions is to incorporate highly selective, level dependent tuning into the inner ear model comparable to that found in the auditory system. The model takes as its input, acoustic stimuli and generates as its output, the neural firing on auditory nerve going to the brain.

The impaired auditory models focus on modelling two sensorineural impairments: recruitment and frequency smearing. The defects that cause these impairments in the auditory system are introduced into the unimpaired auditory model to emulate the auditory system suffering from these impairments. The reason for emulating the impaired auditory system is to facilitate comparison between the impaired neural activity generated by the impaired model and the unimpaired neural activity generated by the unimpaired model. Such models are extremely useful in the development of compensation strategies as they provide access to the responses of the auditory system (in its various forms) and facilitate objective analysis of the effectiveness of compensation strategies. This work details the development of a compensation mechanism to alleviate the effects of recruitment in the impaired model to illustrate how these models may be used in the development of digital signal processing strategies for incorporation into digital hearing aids.

This work also presents suitable methods for the experimental parameterisation of the impaired auditory models in this work on the basis of impaired subjects so they can emulate real instances of each impairment.

The development of each model is followed by an examination of their responses to various stimuli to illustrate the important features of each. The primary results show that the responses of the models follow closely with experimental data and the known responses of the auditory system, providing accurate emulations of the auditory system in its various forms. The responses of the compensated recruited auditory model illustrate that the compensation mechanism is capable of restoring compression in the impaired model to within normal levels and that when compensated, the effects of recruitment are significantly reduced restoring the growth of loudness in the impaired model to within normal levels. Overall this work shows that digital signal processing models can provide a very effective platform from which auditory impairment may be alleviated and that the models in this work provide an ideal environment in which to develop such strategies.

Extensive simulation studies and experimentation with various stimuli are presented to support the theoretical development of the models in this work.

Acknowledgements

No research is done in isolation; I wish to acknowledge the advice and assistance received from many others in completing this work.

Firstly, I would like to thank my supervisor, Dr. Brian Foley, for his support and encouragement throughout the development of this work and for providing me the freedom to explore such a fascinating topic.

Secondly, I am indebted to the Trinity Foundation and 3Com, for providing the financial support with which to pursue this research.

I would also like to acknowledge the technical assistance and expertise extended to me by the staff of the Department of Electronic and Electrical Engineering in TCD, especially Robbie Dempsey, Conor Nolan and Bernadette Clerkin, who were always available to provide support when needed. Thanks are also due to Liam Dowling for his insightful discussions on the beneficial applications of digital signal processing.

A special note of thanks must go to the Trinity College Library staff who were invaluable for their efforts in tracking down obscure books and papers. I would also like to thank Dr. Dermot Furlong for providing access to his extensive JASA collection.

I would like to express my sincerest thanks to Bernard O'Reilly, Mark Fynes and Jonathan O'Donovan for their patience in reading sections of this thesis and providing invaluable comments throughout the writing process of this work.

To my postgraduate compatriots Mark, Toman, Sarah, Dawn, Sparks and Will, and my dear friends Yvonne, Carmel and Bob for providing the friendship that made both undergraduate and postgraduate life in college some of the best years of my life. And to all the crew from AAP and the Printing House, my sincerest thanks.

Finally, I would like to thank Bernard, Rosemary, Fiona, Michael, Frank, Doreen and Leo for their continual support, encouragement and love, without which it would never have been possible.

to Bernard & Rosemary.

Table of Contents

Chapter 1 - Introduction 1

1.1 Human Hearing 1

1.2 Current Approaches to Correcting Hearing Impairment 1

1.2.1 Assessment of Auditory Impairment 2

1.2.2 Limitations of Recent Hearing Aid Technology 3

1.3 Using Digital Signal Processing in Audiology 3

1.4 Modelling the Human Auditory System 6

1.4.1 Benefits of Auditory Models 6

1.4.2 Past Approaches to Auditory Modelling. 7

1.4.3 Recent Approaches to Auditory Modelling. 7

1.4.4 Current Approach to Auditory Modelling 8

1.5 Modelling Impairments of the Human Auditory System 9

1.5.1 Past Approaches to Impairment Modelling and their Limitations 9

1.5.2 Current Approaches to Impairment Modelling and their Advantages 10

1.6 Scope and Structure of the Present Work 11

Chapter 2 - Background 14

2.1 Introduction 14

2.2 The Human Auditory System 14

2.2.1 The Outer Ear 15

2.2.2 The Middle Ear 15

2.2.3 The Inner Ear 18

2.2.4 The Perception of Sound 20

2.2.5 The Complete Auditory System 22

2.3 Impairment of the Human Auditory System 23

2.3.1 Recruitment 23

2.3.2 Frequency Smearing 24

2.4 Summary and Conclusions 26

Chapter 3 - Modelling Survey 27

3.1 Introduction 27

3.2 Unimpaired Auditory System Modelling 28

3.2.1 Macromechanical Auditory Modelling 28

3.2.2 Micromechanical Auditory Modelling 32

3.2.3 Hybrid Auditory Modelling 34

3.2.4 Middle and Outer Ear Modelling 37

3.2.5 Auditory Modelling Discussed 39

3.3 Impaired Auditory System Modelling 40

3.3.1 Modelling the Sensorineurally Impaired Auditory System 40

3.3.2 Modelling the Effects of Sensorineural Impairments 45

3.3.3 Impairment Modelling Discussed 50

3.4 Recruitment Compensation System Modelling 51

3.4.1 Recruitment Compensation by Linear Amplification 51

3.4.2 Recruitment Compensation by Amplitude Compression 52

3.4.3 Recruitment Compensation Discussed 54

3.5 Summary and Conclusions 55

Chapter 4 - Modelling the Human Auditory System 57

4.1 Introduction 57

4.2 Development of the Outer Ear Model 58

4.3 Development of the Middle Ear Model 62

4.4 Development of the Inner Ear Model 65

4.4.1 Modelling the Mechanical Motion of the Basilar Membrane 66

4.4.2 Modelling the Resonance between the Basilar and Tectorial Membranes 82

4.4.3 Modelling the Mechanical-to-Neural Transduction of the Inner Hair Cells 84

4.4.4 Modelling Feedback within the Inner Ear 92

4.4.5 The Complete Inner Ear Model 94

4.5 The Complete Auditory Model 95

4.5.1 The Outer and Middle Ear Sub-Model Dependencies 96

4.5.2 The Inner Ear Sub-Model Dependencies 97

4.5.3 Model Initialisation 98
4.5.4 Complete Model Implementation and Operation 98
4.6 Simulation of the Auditory Model 99
4.6.1 The Outer and Middle Ear Response 99
4.6.2 Tuning Curves of the Basilar Membrane Model 103
4.6.3 Tone Burst Response of the Auditory Model 110
4.6.4 Compression in the Auditory Model 112
4.6.5 Ramped and Damped Responses of the Auditory Model 115
4.6.6 Response of the Auditory Model to Synthetic Speech 119
4.7 Summary and conclusions 121

Chapter 5 - Recruitment 125

5.1 Introduction 125
5.2 Modelling Recruitment 126
5.3 Simulation of the Auditory Model suffering from Recruitment 128
5.3.1 Tuning Curves of the Recruited Basilar Membrane Model 129
5.3.2 Neural Response of the Recruited Model to Rising Stimuli Level 131
5.3.3 Tone Burst Response of the Recruited Model 134
5.3.4 Neural Response of the Recruited Model to Synthetic Speech 136
5.3.5 Compression in the Recruited Model 138
5.4 Identification of Recruitment using Distortion Products 138
5.5 Comparison with other Recruitment Models 141
5.6 Summary and conclusions 145

Chapter 6 - Frequency Smearing 147

6.1 Introduction 147
6.2 Modelling Frequency Smearing 150
6.3 Simulation of the Auditory Model suffering from Frequency Smearing 152
6.3.1 Tone Burst Response of the Frequency Smeared Model 152

6.3.2 Response of the Frequency Smeared Model to Synthetic Speech 159
6.4 Comparison with other Frequency Smeared Models 162
6.5 Summary and Conclusions 166

Chapter 7 - Recruitment Compensation 168

7.1 Introduction 168
7.2 Compensation Function Development 170
7.2.1 Generation of Band-Limited Sub-Signals 171
7.2.2 Generation of the Compensation Measures for the Sub-Band Signals 191
7.2.3 The Complete Compensation System 200
7.3 Simulation of the Recruitment Compensated Auditory Model 201
7.3.1 Neural Response of the Compensated Recruited System to Rising Stimulus Level 201
7.3.2 Tone Burst Response of the Compensated Recruited System 203
7.3.3 Compression in the Compensated Recruited System 204
7.3.4 Neural Response of the Compensated Recruited System to Synthetic Speech 205
7.4 Other Recruitment Compensation Systems 207
7.4.1 Linear Amplification with Automatic Gain Control 207
7.4.2 Amplitude Compression with Equalisation 208
7.4.3 Parametric Compression of Speech Parameters 211
7.4.4 Linear versus Non-linear Processing 212
7.5 Summary and Conclusions 214

Chapter 8 - Model Characterisation 216

8.1 Introduction 216
8.2 Characterisation of the Unimpaired Auditory Model 216
8.3 Characterisation of the Recruited Auditory Model 217
8.3.1 The Alternate Binaural Loudness Balance Test 218
8.3.2 Loudness Growth in 1/2-Octave Bands 220
8.3.3 Distortion Product Otoacoustic Emissions 223
8.4 Characterisation of the Frequency Smeared Auditory Model 228

8.5 Characterisation of the Recruitment Compensation System 232
8.6 Summary and Conclusions 233

Chapter 9 – Summary and Conclusions 235

9.1 Introduction 235
9.2 Summary of Models and Results 235
9.2.1 The Unimpaired Auditory Model 235
9.2.2 The Recruited Auditory Model 238
9.2.3 The Frequency Smeared Auditory Model 239
9.2.4 The Recruitment Compensated Recruited Auditory Model 241
9.3 Overall Conclusions 243
9.4 Critical Evaluation 245
9.5 Suggestions for Future Work 249

References 255

Appendix 269

A1 The Noble Identities 269
A2 Power Complementary Functions 269
A3 The Method of Bisections and the Secant Method for Root Solution. 270

Chapter 1 - Introduction

1.1 Human Hearing

As a race humankind enjoys the ability to hear and be heard. It is a luxury many of us take for granted in our everyday lives, allowing us to communicate freely and enjoy the acoustic delights of the world around us. Our ability to communicate and cooperate as a sentient race has allowed us to achieve marvels beyond anyone's dreams, turning the science fiction of now into the science fact of the future. It is an ability that has played an enormous role in facilitating the evolution of modern society. This evolution has, however, made it possible to create modern living environments that increasingly tax the very senses that have allowed us create them. Only when we notice the adverse effects of these environments on our senses can we begin to fully understand the enormity of losing such treasured gifts. Unfortunately, the development of a hearing impairment is commonplace in the environments we live and work in today; what is even worse is there are people who are born with hearing impairments and never experience the marvels of crystal clear sound. In the very young, hearing impairments markedly affect the development of a person as a fully communicating being and in future life can ultimately reduce the quality and productivity of their lives.

All, however, is not lost; humankind, being an inventive and innovative race, has evolved to a stage where hearing impairments can be alleviated to an extent. As our technology progresses we are becoming increasingly capable of compensating for the reductions in the efficiency of the human body due to age, disease and our environment. This Chapter introduces the topics under consideration, assuming some technical familiarity on the part of the reader while leaving detailed technical explanations for future Chapters.

1.2 Current Approaches to Correcting Hearing Impairment

Currently, there are two means of providing alleviation of hearing impairment, one through the use of surgical procedures and the other through the use of hearing aids. The choice of which method is used depends both on the particular impairment and on the severity of the affliction. In general, many conductive hearing losses caused by

abnormalities of the outer ear, middle ear or both can be corrected surgically. However, sensorineural hearing loss caused by some abnormality of the inner ear or the hearing centres of the brain is extremely complicated to alleviate surgically. In general surgical procedures on the inner ear are only viable under the most extreme cases of hearing loss, due to the risk and the collateral damage involved in performing such deep surgery. The preferred option for alleviating sensorineural hearing loss is through the use of hearing aids.

1.2.1 Assessment of Auditory Impairment

Before any strategy can be applied to assisting an individual, the particular impairment and its possible causes need to be identified as accurately as possible. This is carried out by firstly developing a case history for the individual to get an idea of the background behind the problem. This procedure also helps to identify individuals for whom medical treatment should be sought, particularly in the case of an impairment due to infection. Prior to performing any audiometric tests, otoscopy is performed to assist in identifying abnormalities that may require medical attention. It also provides dimensional details of the ear for characterisation of the subject. Tuning fork tests can then be used to quickly identify the nature of the hearing loss; there are several such tests but the most widely used are the Weber and Rinne tests. These tests, however, are not good at determining the degree of hearing loss. The standard method of assessing the degree of hearing loss is typically pure-tone audiometric analysis, which identifies the types of loss present and degree of loss across the audible frequency range. The results of pure-tone audiometric analysis allow broad characterisation of hearing impairments. Figure 1.1 shows a graphical representation of the classifications recommended by the British Association of Teachers of the Deaf [1].

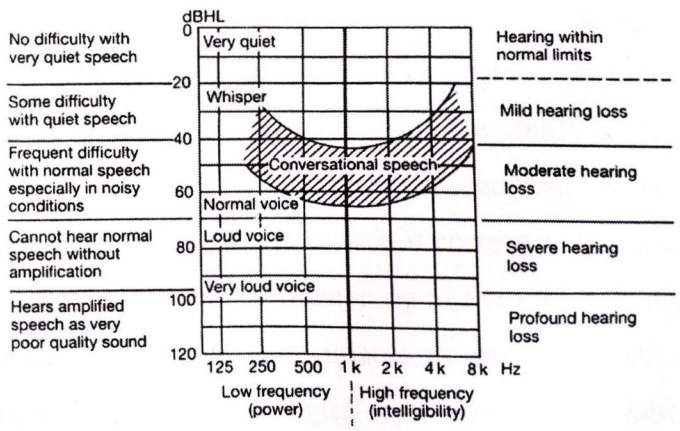


Figure 1.1 Degree of hearing loss and its effect on hearing speech.

Pure-tone audiometry separates conductive and sensorineural hearing loss and identifies the threshold of a patient's hearing at particular frequencies. The data from pure-tone audiometric analysis form the basis for the characterisation of a hearing impairment. The assessment of auditory impairment is discussed in greater detail in Chapter 8.

1.2.2 Limitations of Recent Hearing Aid Technology

Up until very recently the technology employed in hearing aids was quite simple in functional terms, usually comprising of frequency selective linear amplification with some automatic gain control. It is well known that these hearing aids provide only minimal benefits to the hard of hearing [2-7]. The reason for this is that there are many types of sensorineural hearing loss that do not respond to this kind of treatment. These complicated impairments have to be identified using specialised tests to differentiate between cochlear and retrocochlear hearing disorders in conjunction with standard tests to identify the site and cause of these hearing losses. Merely amplifying all acoustic signals above the auditory threshold so the impaired individual can hear them does not correct for many of the cochlear disorders that arise. The problem is that many of the algorithms previously used in hearing aids are not capable of providing alleviation for the abnormally rapid growth in loudness or the loss of frequency selectivity and sensitivity that typically occur when sensorineural impairment is present. Greater detail regarding the limitations of hearing aid technology is discussed in Chapter 7 in conjunction with specific sensorineural hearing impairment and the development of a compensatory algorithm.

1.3 Using Digital Signal Processing in Audiology

Of particular interest in this work are sensorineural impairments associated with damage to the cochlea; being the most important part of the auditory system it is also the most delicate and easily damaged. When parts of the cochlea are damaged they introduce extremely complex impairments that not only reduce the ability of the impaired to hear a sound but also reduces their ability to distinguish the content. If the auditory system is unable to suitably breakdown a sound and encode it correctly, then the auditory centres of the brain will have difficulty decoding the meaning of the sound. Many older hearing aids typically employing analog processing are unable to provide suitable alleviation for the sensorineurally impaired. This is where digital signal processing (DSP) can play a major role. DSP techniques have come a long way over the years allowing efficient complex

signal manipulations in both the time and frequency domains, manipulations that can be used to provide better mechanisms for alleviating hearing impairment [8]. It is a testament to the technology that most of the main manufacturers of today have incorporated DSP into their more advanced hearing aids. One of the world's leaders in hearing aid research and design, Siemens Audiological Engineering Ltd., has this to say about digital hearing aid technology

> *"Just because a hearing instrument will be digital does not mean it will be superior. The true heart of digital technology will not lie in the flexibility of fitting to today's amplification targets, but will rely upon the ability to provide a real end user benefit, and, in particular, the ability to significantly improve speech recognition (both in quiet and in noise) over current instruments".*
>
> Siemens Audiology Group [9].

The advent of digital hearing aids has seen a significant increase in the capabilities of hearing aids. However, that is not to say non-digital instruments do not have a place; there are many disorders that can be tackled using non-digital and less advanced digital devices. One of the main drawbacks with the latest digital hearing aids is the cost, which places many of these devices out of reach of those that need them most. In many cases it is only through the intervention of State Health Boards that these devices can be acquired. It is especially noteworthy that the National Health Service in the United Kingdom has in the last year made moves towards making digital hearing aids available to their dependants.

The primary goals of digital hearing aid technology is to provide better solutions to those problems currently alleviated with older technologies and begin tackling those problems that current instruments cannot tackle. According to recent studies by health care institutions such as the German Green Cross, nearly every fourth person living in industrialised countries suffers from hearing problems. Particularly alarming is the fact that the incidence of hearing impairments among the young continues to increase. The reason for this development is the permanent overstressing of the hearing mechanism. With this expanding number of impaired individuals, the varieties and severity of impairments is becoming increasingly large and a technology is required that can adapt to cater for this increase. It is in this kind of environment that the flexibility of digital signal processing excels.

Currently an enormous body of knowledge on DSP exists with which to begin addressing many of the current problems. So why are there so few hearing aids harnessing these powerful DSP techniques and why are there fundamental impairments that can only be minimally alleviated?

There are a few reasons but the most outstanding problem in this area is the lack of engineering (DSP) models for describing particular impairments and the need for a means of experimentally parameterising those models on the basis of the impaired individual. These DSP models of the auditory impairments are required to fully characterise the impairments in signal processing terms to facilitate successful development of DSP pre-processing techniques. When the information they present to the ear is processed by the impaired auditory system it results in the generation of neural information that better approximates the neural information unprocessed stimuli generate in the unimpaired ear. The best way to develop these DSP pre-processing techniques is to fully characterise the impairments they are trying to correct. Some of the reasons for this are

- The impairment model will identify in DSP terms what it is the ear is doing incorrectly.
- The characterisation of the impairment itself will provide important information as to what type of approach may best tackle the impairment.
- These models will allow objective testing of how appropriate a corrective technique may be at alleviating an impairment and will assist in modifying the corrective technique to provide the best possible solution.
- The models can be fine tuned to emulate a specific case of impairment.
- DSP solutions provide a means to deal with multiple problems.

Being able to create accurate DSP models of impaired systems and developing DSP algorithms that can compensate for these problems is only half of the solution. The models have to emulate real systems and unless these models can be made to emulate a particular individual's variety of an impairment, the corrective solutions cannot be tailored to provide optimum alleviation. The exploration of clinical tests to characterise these DSP models is therefore just as important as the models themselves. This work considers not only the development of models for auditory impairments but also a means of parameterising those models on the basis of an individual's impairment.

1.4 Modelling the Human Auditory System

There are many reasons both philosophical and scientific why one would want to model the human auditory system. By their very nature biological systems are among the most complicated systems in our world today. The development of models that accurately emulate the function of such systems would provide us with important tools with which to acquire a better understanding of how they function and more importantly malfunction. Understanding how the human auditory system functions enables us to prevent damage to a vital sense by highlighting its weaknesses and facilitates the development of a compensatory mechanism when it does suffer from impairment.

1.4.1 Benefits of Auditory Models

Being able to accurately model the human auditory system would provide the ability to objectively and non-destructively experiment with this system, with a view to probing its limitations and its ability to respond to a multitude of environments. It would also allow exploration of how defects in this system can affect our perception and comprehension of our environment and the people in it. A simple example of the limitations of subjective testing is the classic problem faced by audiologists today: trying to get an individual to accurately discriminate between different amplitude sounds of similar frequency. The problem with this is that the perceived loudness of a sound is a very subjective thing. A similar test when considered in the context of an accurate model of the human auditory system simply would not have such a problem. This of course assumes the model emulates the function of the system with a sufficient degree of accuracy. However, it is the acquisition of this sufficient degree of accuracy that is the Holy Grail.

It is for reasons such as these that man constantly tries to map the unknown with the known so that he may understand it. This goal is not a new one. When considering modelling of the human auditory system it can be traced back to the first century BC with the postulations of the Roman poet and philosopher Lucretius. He proposed a model involving little grains of sand in the inner ear responding to different tones. Today his model would be considered a "particle theory" of sound, anticipating in a most amusing manner the phonons of modern quantum acoustics. These theories, as fantastic as they sound, embody concepts that today form the basis of our understanding of sound [10].

1.4.2 Past Approaches to Auditory Modelling.

When we consider the process that has been taking place over the years in modelling the auditory system, a sensible approach of divide and conquer has been taken towards the subject. The system as a whole up until recently has been too complex to model as a complete system. This has been the case not only due to technological limitations but also because we simply did not know what was taking place within the system. To try and gain insights into whole subject matter, researchers have taken small parts of the complete system and tried to accurately model how they operate. Although these models may have limited use on their own, in terms of the system as a whole, they provide key information needed to understand how the complete system functions.

A clear example of this can be seen when we consider probably the most important part of the auditory system, the cochlea. The cochlea is the primary organ responsible for hearing yet it has really only been since the development of digital computers that we have begun to be able to try and model this most important part of the auditory system. However, it is a massively complex structure with many component parts, one of which is the inner hair cells. Jont Allen produced one of the most accepted basic models of the inner hair cells only in 1983 [11], after the work of Davis in 1956 [12]. This model, of a very small part of the cochlea, has provided us with many insights into how the transduction process takes place within the cochlea. This is only one of the many important models of a section of the overall system that has allowed us to understand what is taking place in the rest of the cochlea.

1.4.3 Recent Approaches to Auditory Modelling.

The models of the component elements of the auditory system that have been built up in the past serve as the building blocks used today to develop more complete and accurate models of larger sections of the complete system. It is only within the last decade that the technology has become available with which to build complete models of the auditory system. When considering a complete auditory model, it is meant in the sense of a model that can be presented with an input as a person would be, and produce as an output the neural information that would be sent to the person's brain on hearing the input. Indeed even today there are still few models that accurately simulate the complete process. Even still, all of these complete models have to make certain assumptions to fill in the gaps for

what is not yet fully understood. In truth, most of the complete models created so far are still limited by the knowledge upon which they are based. This knowledge, however, is continually evolving and advancing. Therefore it is necessary to constantly update these models to incorporate the latest advances and continually question any assumptions that have been depended upon. Where possible, a move should be made away from depending upon assumptions to proving whether or not these assumptions have been correct and thus base the models on proven information. In fact a stage has been reached where the models that have been created using different approaches to achieve the same goal are turning up different results. Ultimately, all models should provide the same results, if the auditory system is sufficiently understood and modelled. The results, however, imply that a complete understanding of the auditory system is yet to be attained. Only when models produce consistent results that are sufficiently similar to what is expected will the models become more than just learning tools.

1.4.4 Current Approach to Auditory Modelling

One of the main goals of this work has been the development of a model of the complete human auditory system, based on current known fact and informed assumption so as to produce a model that closely matches the results from experimentation and measurement. The development of this work in itself, however, has limitations. For ethical reasons, the direct measurement of many of the responses of the human ear cannot be performed. However, the use of other species with very similar auditory systems has been used to provide suitably altered information that is assumed to give a good indication of what would be expected if these measurements were to be carried out on humans. Such assumptions can be seen in most of the models today. However, with the advances in technology, the dependence on such techniques is diminishing rapidly. When choosing an approach to take for modelling such a complex biological system this limitation must be acknowledged, but where possible, avoided. The successful creation of such a complete model would provide many advantages, particularly in the context of impairment. Having a signal processing description of how the ear functions would allow us to objectively measure the neural output of the unimpaired auditory system, to any number of test inputs and thus provide an objective reference from which auditory impairment could be analysed. This has been one of the major considerations throughout the development of the model in this work. Ultimately, an input-output model of the human auditory system is developed which models the complete auditory process from the air borne sound waves in

the outer ear through to the fluid borne sound pressure waves in the inner ear and on into the neural firing patterns on the nerve fibres leading to the brain.

1.5 Modelling Impairments of the Human Auditory System

The primary endeavours over the past few years have focused on the development of models of the human auditory system with a view to emulating the particular functions taking place in this system when it is in its unimpaired state. That is to say when the ear being modelled does not suffer from any type of defect. In fact

> *"there has been very little discussion of the effect of different forms of cochlear damage on cochlear mechanics per se. In meetings devoted to cochlear mechanics over the past 11 years, for example, only Allen (1990) and Ruggero and Rich (1990) addressed issues related to hearing loss."*
> Walt Jesteadt 1997. [13].

It is really only in the last decade or so that research groups around the world have started in earnest to develop models of auditory impairments. The lack of detailed signal processing representations of auditory impairments has been a serious impediment to the application of DSP to the alleviation of hearing impairment. The models of impairments that have appeared, however, provide great insight into how the auditory system malfunctions when it suffers from hearing impairment.

1.5.1 Past Approaches to Impairment Modelling and their Limitations

Most of the literature that has been produced tends to focus on modelling the particular impairments themselves, giving detailed descriptions of how signals are altered or treated differently due to the presence of an impairment. These in themselves prove invaluable when trying to uniquely identify problems, their possible causes and ultimately their final effects. A very important function provided by such models is in making the links between what is happening to the signals and which parts of the biological auditory system could be impaired. However, such models operating in isolation do tend to ignore the fact that the impairment is a malfunction of a much larger system. There are many interactions and external influences that isolated models do not consider, not least among these is nature's ability to try and compensate for an impairment.

Another of the major problems with isolated impairment models is that they are generally very difficult to characterise in terms of a patient's problem. Basically most of these models do not allow for the characterisation of real biological impairments. They can simulate the effects of an impairment and how the signals are altered but cannot characterise how the whole auditory system is malfunctioning so as to manifest a particular impairment. It would be more correct to say that these models are in effect emulating the manifestations of the impairments rather than the impairments themselves. It is partly this characterisation problem that has prompted the recent subtle change towards modelling the complete auditory system and introducing impairments into these models, as they would appear in the real system.

1.5.2 Current Approaches to Impairment Modelling and their Advantages

There are potentially many benefits arising from modelling the auditory system and introducing impairments into these models. The models that are being created can more closely match what is actually taking place within the auditory system and provide a more realistic representation of what is really taking place inside an impaired ear. Such a framework would provide an ideal situation in which to emulate the complex interactions that take place within the biological system. Similarly, many of the secondary effects that occur within the auditory system due to malfunction of the system can be taken into account when providing an overall model of the hearing process. The resulting information that can be extracted from these models will provide a much more realistic view of what is taking place throughout the entire system. This information will help test any assumptions that have been made both in the model and in how particular impairments have been introduced into the models when compared with real systems.

Complete models that are capable of being impaired so they manifest the same symptoms as occur in the impaired biological system would have great use in many areas of the rehabilitation process. Possibly one of the most useful areas would be in the pre-evaluation of the effectiveness that surgical procedures would have on a patient's hearing. The changes made by a surgical procedure could be emulated in a suitably modelled impaired auditory system and direct comparison between the corrected model's output and the normal system's output could provide a measure of how successful the surgical technique might be at correcting the particular defect. This kind of information would be most advantageous both to the patient and the surgeon. However, it would be more

desirable to avoid surgical procedures particularly when dealing with sensorineural impairments due to the possible damage caused by performing such surgery. In general, the use of hearing aids is the predominant means of providing assistance to the hearing impaired and a significant hurdle is the configuration of these devices. In situations like this impairment models would be most useful because they would allow quantitative evaluation of the effects of different configurations.

Other areas of use would be in ascertaining how effective new hearing aid strategies are at providing relief to their users and also how they can be improved upon. Of particular interest in this work is the use of such models to assist in the development of DSP algorithms for the alleviation of hearing impairment, in particular sensorineural hearing impairment. Most sufferers of these impairments do not fall into the category suitable for surgery. However, at present there are limited means of providing relief from these impairments.

1.6 Scope and Structure of the Present Work

The primary objectives of this work are:

1. The development of a complete model of the unimpaired human auditory system.
2. The development of a model of the human auditory system suffering from recruitment.
3. The development of a model of the human auditory system suffering from frequency smearing.

Considering these objectives, the complete unimpaired auditory model is an input-output system capable of taking raw sound as its input and producing the neural firing on the auditory nerve fibres leading to the brain as its output. This model incorporates the functions of the outer, middle and inner ears with particular emphasis on the functions of the inner ear to facilitate exploration of sensorineural hearing impairment. The model does not, however, consider sensorineural impairment due to disorders of a retrocochlear nature, as its output is the neural firing delivered to the auditory nerves and does not include models for the auditory nerves themselves. In terms of modelling sensorineural impairment two particular impairments, recruitment and frequency smearing, are modelled through the

introduction of defects into the complete unimpaired auditory model similar to those defects occurring in the biological system that cause these impairments.

The secondary objectives of this work are:

1. The development of a model for the compensation of recruitment in an auditory model.
2. The exploration of methods for parameterising the models developed in this work.

Having developed models of the unimpaired and impaired auditory system, one of the secondary objectives is to develop a model for the alleviation of recruitment in the auditory model suffering from recruitment. Specifically, this compensation mechanism must cater for the increased auditory threshold and the abnormally rapid growth in loudness associated with recruitment. Finally, the impairment models and the compensation model require parameterisation to be able to successfully emulate the systems they represent. To this end, the final objective of this work is to explore methods for the parameterisation of the models developed. Once parameterised, impaired models are capable of modelling an auditory system suffering from a particular impairment and, using suitable characterisation methods, the compensation system can be parameterised to alleviate recruitment.

A prerequisite for all the models is they are computationally efficient, as even with the numerical processing power of computers today, complex models become extremely demanding as their level of detail increases. This prerequisite is especially true for the recruitment compensation model as ultimately such algorithms must be capable of operating in the confines of a hearing aid where there are significantly greater constraints.

The structure of the remainder of this thesis is as follows:

Chapter 2 presents a background on the physiology of the human auditory system, its function and malfunction. This Chapter is specifically designed to re-acquaint the reader with the fundamentals of auditory processing and to identify the nature of the impairments under investigation, recruitment and frequency smearing.

A survey of the recent approaches used to model the unimpaired auditory system, recruitment, frequency smearing and recruitment compensation is presented in Chapter 3.

One of the primary goals of this work is the development of a model of the complete auditory system from the outer ear to the auditory nerves. Chapter 4 details the development of this model along with simulation results of the model for various stimuli. Comparison is drawn with experimental work carried out on various species with auditory systems similar to the human auditory system, with other prominent models and with the psychoacoustic responses of the human auditory system.

The other primary goals of this work are aimed at developing engineering (DSP) models of the sensorineural auditory impairments of recruitment and frequency smearing. Chapters 5 and 6 deal with the development of models of these impairments and simulation results from the impaired models are presented for various stimuli. The validity of these models is discussed in relation to other isolated models and psychoacoustic experimental data.

Having developed a model for recruitment, Chapter 7 details the development of a compensation system to alleviate the effects introduced by recruitment. This compensatory mechanism is used in conjunction with the auditory model suffering from recruitment and the results of simulating the compensated recruited auditory system is presented for several stimuli. The suitability of this mechanism is discussed in relation to other mechanisms and the known responses of the unimpaired auditory system.

Chapter 8 deals with the problems involved in parameterising the impairment models in terms of impaired individuals and explores tests capable of acquiring the necessary information to perform the required characterisation. This chapter also examines the methods for acquiring the necessary information from the recruitment auditory model to parameterise the recruitment compensation system.

The thesis finishes with a summary and critique of the work presented including general conclusions and suggestions for future expansion and development of this work.

Chapter 2 - Background

2.1 Introduction

The models presented in this work and the works of others are fundamentally computational systems attempting to capture the essence of the operations and responses of the biological auditory system. Before considering the current state of auditory and impairment modelling or examining the developmental endeavours of this work, it is essential to be familiar with the underlying biological systems and their interactions. It is therefore the purpose of this Chapter to provide a suitable introduction to the physiology of the human auditory system, its function and malfunction. The level of detail provided is not intended to be a microscopic analysis of the human auditory system but is sufficiently detailed to acquaint the engineering reader with the fundamentals of auditory processing.

2.2 The Human Auditory System

The human auditory system can essentially be divided into three main sections, the outer ear, the middle ear and the inner ear. It is subdivided in this way because within each section the acoustic signals are transported in a different medium. In the outer ear signals are air borne, in the middle ear the signals are mechanically borne (via the ossicular chain) and in the inner ear the signals are fluid and electrically borne. Figure 2.1 illustrates a schematic representation of the human auditory system.

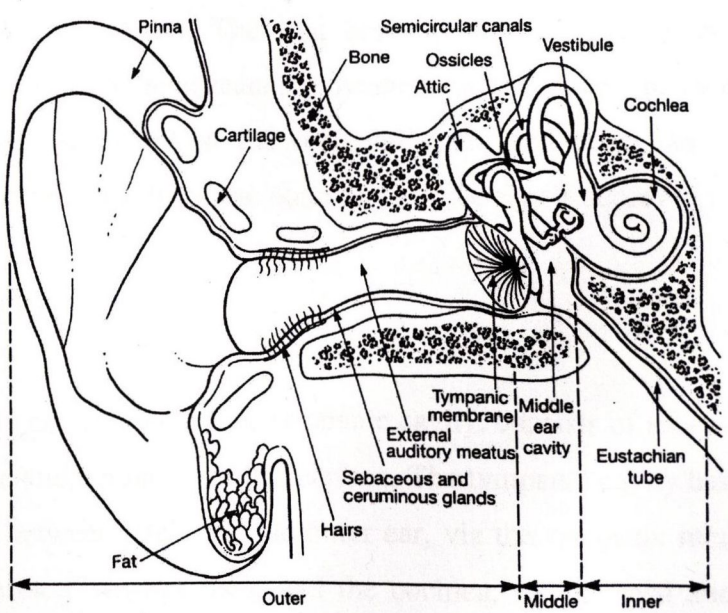


Figure 2.1 Schematic of the Human Auditory System.

Between each of the sections there has to exist an interface to transfer the signals from medium to medium. Each section will now be considered.

2.2.1 The Outer Ear

The outer ear consists primarily of two parts, the pinna and the external auditory meatus (ear canal), which terminates at the air-to-bone interface with the middle ear. The main function of the pinna is to funnel sound into the external auditory meatus; besides this, complex convolutions in the pinna provide cues for horizontal and vertical localisation in space [13.1]. The sound directed by the pinna is in the form of pressure waves in the air, which receive minimal frequency dependent pressure amplification due to the shape of the pinna.

The external meatus has two main functions. The cerumen and hairs along its outer lining prevent foreign bodies reaching the tympanic membrane and, being curved, it protects the eardrum and middle ear from trauma. It also acts as a resonance tube with a sealed end. It is generally in adults about 2.5cm long and thus has a resonant frequency around 3kHz. This provides between 10-15dB of amplification for signals with a frequency content around 3kHz.

Having passed the pinna and through the external meatus the sound then impinges on the tympanic membrane which is part of the air-to-bone mechanical interface between the outer and middle ears. The compressions and rarefactions of the sound waves move the tympanic membrane in unison. The first bone in the ossicular chain, the malleus, is connected to the tympanic membrane; movement of the tympanic membrane therefore causes sympathetic movement in the ossicular chain and hence an interfacing action between the air and bone media of the outer and middle ears is achieved.

2.2.2 The Middle Ear

The middle ear, centred on the tympanic cavity, consists of an air-to-bone interface, the ossicular chain and a bone-to-fluid interface. The tympanic cavity has access to the air-to-bone interface between itself and the outer ear, via the tympanic membrane and to the bone-to-fluid interface between itself and the cochlea, via the oval and round windows. Figure 2.2 illustrates a schematic of the tympanic cavity.

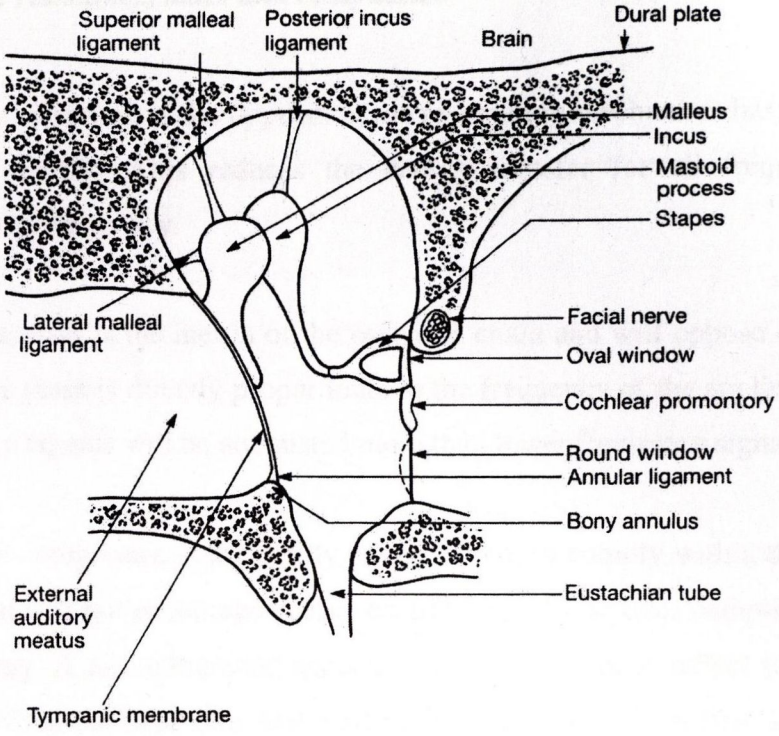


Figure 2.2 The Tympanic Cavity.

The malleus is connected to the tympanic membrane and to the incus. Movement of the tympanic membrane causes movement of the malleus, which in turn effects a force through the incus and onto the stapes. The stapes closes the oval window, which is the entry point into the cochlea. Movement of the stapes in the oval window sets up pressure waves in the cochlear fluid. These pressure changes at the oval window are relieved by sympathetic motion of the round window.

The tympanic cavity itself can be thought of as an air-to-fluid interface and effectively acts to minimise the transmission loss between the outer and inner ears due to the impedance mismatch between air and the fluid (perilymph) that fills the cochlea. The effective mechanical impedance of air is approximately 41.5Ω and of the fluid is approximately 140,000Ω. Without the tympanic cavity, the surface of the fluid in the cochlea would reflect approximately 99.9% of the sound waves impinging on its surface. The ossicular chain and its connections to the two interfaces of the tympanic cavity prevent this reflection but in turn add their own impediments to the impedance matching process they perform. These impediments are, acoustic resistance, acoustic mass and acoustic compliance.

2.2.2.1 Acoustic resistance, mass and compliance

Acoustic resistance is the opposition the responding mechanism has; it is analogous to mechanical friction. This reduces the energy transfer for all frequencies and is independent of the frequency.

Acoustic mass is the inertia of the ossicular chain and will oppose energy transfer. The effect of the mass is directly proportional to the frequency of the applied signal, hence higher frequency signals will be attenuated more than lower frequency signals.

Acoustic compliance is the ability of the system to comply with a driving force. A very stiff system will not encourage energy transfer, equally an over compliant system will absorb the energy. A non-compliant system, however, will act to reflect too much of the driving force. When the ligaments that control the ossicular chain stiffen, low frequencies are attenuated more than higher frequencies.

Depending on the conditions, each type of impedance will dominate in affecting a signal passing through the interface. Thus the middle ear has to perform amplification of the signal to overcome the media mismatch and the impediments it imposes itself. There are two factors that introduce this amplification into the system; the areal relationship and the lever action.

2.2.2.2 Gain of the middle ear

Gain resulting from the areal relationship arises because the area of the tympanic membrane is greater than that of the footplate of the stapes by a factor of approximately 21:1. However, since only two thirds of the tympanic membrane is mobile this ratio reduces to about 14:1, resulting in a magnification of the footplate force by about 23dB. Gain resulting from the lever action arises because the manubrium of the malleus is 1.3 times longer than the process of the incus, thus a lever advantage of 1.3:1 exists, resulting in a further 2.5dB of gain. The combination of these two mechanisms provides an overall gain of approximately 25.5dB within the middle ear prior to application of the force to the oval window. This gain allows the middle ear to overcome the problem of the huge impedance mismatch between the outer and inner ears and efficiently transforms the air borne signals in the outer ear into fluid borne signals in the inner ear.

2.2.3 The Inner Ear

In terms of the auditory process, the inner ear consists of the bony labyrinth, the membranous labyrinth and the organ of Corti. It is contained wholly inside the bone of the skull inside a membranous sac filled with a positively charged fluid, called endolymph. Between the sac and the bone of the skull there is a negatively charged fluid, called perilymph. A potential difference therefore exists across the membrane of the sac. The importance of this potential difference between these two similar fluids will become apparent, see Section 2.2.4.2.

2.2.3.1 The bony labyrinth

The bony labyrinth is a continuous cavity that communicates functionally with the middle ear via the oval and round windows. It begins with the cochlea which spirals around a central bony pillar called the modiolus for two and three-quarter turns. A narrow plate of bone, the spiral lamina, projects into the cavity of the tube within the cochlea, along its entire length, partly dividing the tube in two. The basal end of the cochlea opens into a cavity called the vestibule, which has several openings. Two of these connect to the middle ear, one closed by the round window and the other by the oval window. It is through the vestibule that the fluid borne pressure waves enter the cochlea.

2.2.3.2 The membranous labyrinth

The membranous labyrinth forms a series of communicating sacs and ducts within the bony labyrinth, two of which connect to the duct of the cochlea (utricle & saccule). This duct extends along the bony tube lying on the spiral lamina and stretches across to be attached to the outer wall by the spiral ligaments. This forms a triangular shaped channel with its apex at the modiolus, effectively creating three chambers. The middle chamber formed by the duct is filled with endolymph and the upper and lower chambers, the scala vestibuli and the scala tympani, are filled with perilymph. The upper and lower chambers connect at the apex of the cochlea via a small gap, the helicotrema, and terminate in the vestibule on the oval and round windows respectively. The upper wall of the duct is Reissner's membrane and the lower wall, the basilar membrane. See Figure 2.3.

2.2.3.3 The organ of Corti

The organ of Corti is a structure of cells that forms the sensory organ of hearing and is situated within the scala media on the basilar membrane. See Figure 2.3. It contains two rows of cells, called hair cells, which have fine hair-like structures (stereocilia) projecting from their upper surfaces. There is an inner row of single cells and an outer row of three or four cells separated from each other by Corti's arch. The inner cells have between seventy to one hundred stereocilia in two rows forming a double 'V' with the apex pointing away from the modiolus. The outer cells have more stereocilia arranged in three rows to form a wide triple 'W' again with their apices pointed away from the modiolus. Not all the stereocilia are the same length; the longest are at the apices and gradually reduce in length as they move away from the apices. The reticular lamina lies on top of this structure and above this, a gelatinous membrane, called the tectorial membrane. The longest stereocilia pass through the lamina and are embedded in the tectorial membrane. There are roughly 3000 to 4000 inner hair cells and about 10000 to 12000 outer hair cells in a normal undamaged auditory system.

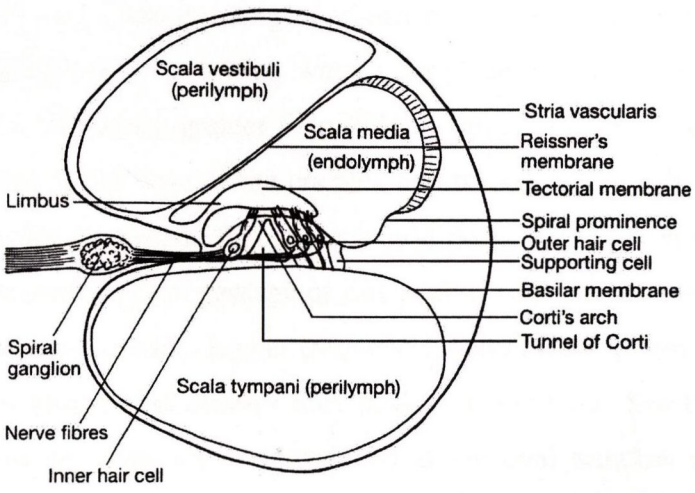


Figure 2.3 Cross-section of the Cochlear Duct.

The organ of Corti is widest at the apex and gets progressively narrower towards the base. The cochlea receives its nerve supply from the eighth cranial nerve and has both afferent (carrying information to the brain) and efferent (carrying information from the brain) nerves. There are less efferent than afferent fibres with the efferent nerves servicing the inner and outer hair cells equally. Each inner hair cell, however, has about 20 afferent nerve fibres synapsing with it, which counts for 90 - 95% of the 30000 to 40000 afferent nerves leaving the cochlea. The remaining 5 - 10% synapse with the outer hair-cells, each

fibre having to be heavily branched and serving up to ten outer hair-cells. It is not completely understood why the brain is able to communicate with the hair cells via the efferent nerves.

2.2.4 The Perception of Sound

The perception of sound is essentially a two-stage process once the sound pressure waves have passed from the outer ear to the inner ear. The organ of Corti is the sense organ, found within the cochlea, in which sound is analysed into its component parts before being electrically coded and sent to the brain. The function of the cochlea can be thought of as a process of transmission followed by transduction. Transmission being the transfer of acoustic energy from the oval window to the organ of Corti while transduction is the process by which this sound energy is converted to electrical impulses to be sent along the auditory nerve fibres to the brain.

2.2.4.1 The transmission process

The basis for our understanding of transmission comes from the theories of Von Békèsy [14]. He proposed a 'travelling wave theory' describing how movements of the oval window with a frequency greater than 20Hz would set up a wave along the basilar membrane, beginning at the base of the cochlea and travelling towards the apex. At some point along the basilar membrane this wave would peak, the location of this peak being highly frequency dependent. The passage of this peak is impeded by the varying mass and stiffness of the basilar membrane. Lower frequencies pass further down the membrane and peak near the apex. Higher frequencies thus peak near the base. See Figure 2.4a. Figure 2.4.b illustrates how pressure waves introduced at the oval window are relieved at the round window.

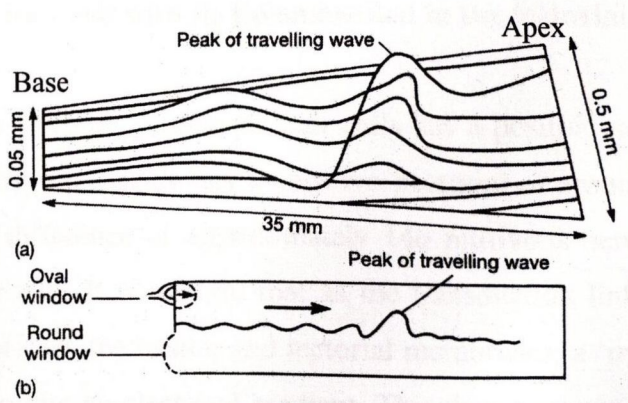


Figure 2.4 Von Békèsy travelling wave on the basilar membrane.

The tuning of the basilar membrane is important for frequency discrimination and the outer hair-cells play an important role in selectively stiffening the basilar membrane, thus sharpening its response by 'vibration amplification'. This can be seen in the production of otoacoustic emissions, which are by-products of the outer hair-cells' activation and can be measured in the outer ear. At low intensity stimulation, individual hair cells are stimulated and as the intensity level rises the membrane is displaced more causing groups of hair-cells to be stimulated reducing the fine-tuning of the system.

2.2.4.2 The transduction process.

Unlike the basilar membrane, when the wave travels along the tectorial membrane, it causes the membrane to move forward and upwards and backwards and downwards. The stereocilia that are imbedded in the tectorial membrane are bent by this shearing motion towards the tallest stereocilia and then away. Adjacent stereocilia are joined by a transduction link at their tips and tension on these links is alternately tightened and released. See Figure 2.5.

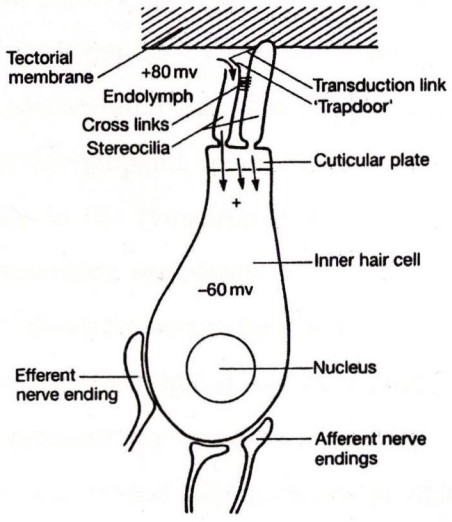


Figure 2.5 A hair cell with its tip embedded in the tectorial membrane.

At rest, the endolymph outside the hair cells has a positive potential of around 80 millivolts and the perilymph inside has a negative potential of around 60 millivolts. This results in a potential difference of approximately 140 millivolts between the outside and the inside of the hair cell. It is thought that as the transduction links tighten, caused by relative movement between the basilar and tectorial membranes, a *'trap door'* opens at the top of each cilia setting up an electrical gradient. This then causes an ionic interchange in

which the hair cell becomes more positively charged stimulating the release of neurotransmitters at the base of the cell which fire the nerve impulse causing a wave of depolarisation to pass down the length of the axon. This nerve impulse, called the action potential, is a few millivolts and is fixed. Higher intensity sound produces more firing action potentials but not larger potentials. It is also known that nerve fibres themselves show a frequency selectivity and have a spontaneous firing rate from 0 - 150 per second. Firing of the nerves occurs only at particular phases of the waveform, allowing a temporal regularity in response to a periodic stimulus. The overall effect is that sound waves in the fluid of the inner ear are converted to electrical impulses on the nerve fibres going to the brain where they are perceived as the sensation of hearing.

2.2.5 The Complete Auditory System

Having considered the components of the auditory system, it is important to realise that the different sections do not operate in isolation but are part of a complete system. In simple terms the auditory system can be thought of as an interface converting sound pressure waves into electrical signals going to the brain. The complete process starts with the sound pressure waves being guided by the pinna into the external auditory meatus, which channels the sound to the outer ear side of the tympanic membrane. When the pressure waves interact with the tympanic membrane they cause it to move. In the middle ear the malleus is connected to the tympanic membrane and when the sound pressure waves in the air cause the tympanic membrane to move, the tympanic membrane causes the malleus to exert a force down the ossicular chain to the footplate of the stapes. The footplate sits in the oval window and when it moves it introduces pressure waves into the fluid of the cochlea. These pressure waves cause travelling waves to propagate down the length of the basilar membrane whose displacement is maximal at specific frequency dependent locations. The basilar membrane is therefore able to separate the frequency components of the incoming stimulus and suitably excite the inner hair cells. Excitation of the inner hair cells causes them to fire neural impulses along the nerve fibres leading to the brain. This encodes the pressure waves into electrical impulses the brain is capable of interpreting. This overall process is clearly a simplified view of the many complex interactions taking place in the auditory system, however, it suffices to provide an overall view of the operations taking place. Greater detail is considered during the development of the auditory model, see Chapter 4.

2.3 Impairment of the Human Auditory System

It can be seen that the auditory system is a very complex system containing many components, each performing highly specialised functions. If this system were perfect then the reason for modelling it would be purely academic. However, like all biological systems it is not perfect and suffers from impairment. In fact, given the number of components in the system and the number of possible pathological processes that can take place in any part of the ear it is surprising that we can hear at all. These pathological processes can be classified under the following headings [15].

- Traumatic, i.e. from injury;
- Infective, or from micro-organisms;
- Tumours;
- Iatrogenic, caused by physician or surgeon, by use of inappropriate medication, surgical accident or unsterile operation;
- Idiopathic, or of unknown cause
- Physiological disorders, including degenerative effects of ageing;
- Congenital, or from birth, whether inherited or caused by environmental factors;

In general any of the above pathologies can affect any part of the auditory system and when they do, the resulting hearing impairments are classified under two main headings: *conductive hearing loss* and *sensorineural hearing loss*. Conductive hearing losses are losses resulting from a physical blockage of sound transmission to the inner ear and typically involve problems associated with the outer and middle ears. Sensorineural hearing loss on the other hand involves problems with the cochlea and auditory nerve. It may also be associated with lesions of the central auditory pathway. However for the purpose of this work only sensorineural impairments of a cochlear nature are considered, specifically the sensorineural hearing impairments of *recruitment* and *frequency smearing* are considered.

2.3.1 Recruitment

Recruitment is best explained in terms of the normal operation of the auditory system. When the normal ear is presented with various levels of stimuli it experiences a linear relationship between the increase in intensity and the related increase in perceived

loudness. However, damage to the outer hair cells in the cochlea results in loss of compression around the affected area of the basilar membrane causing an increase in the threshold of hearing at the frequency responded to by the damaged region and causing the rate of growth of loudness with increasing stimulus level to grow more rapidly.

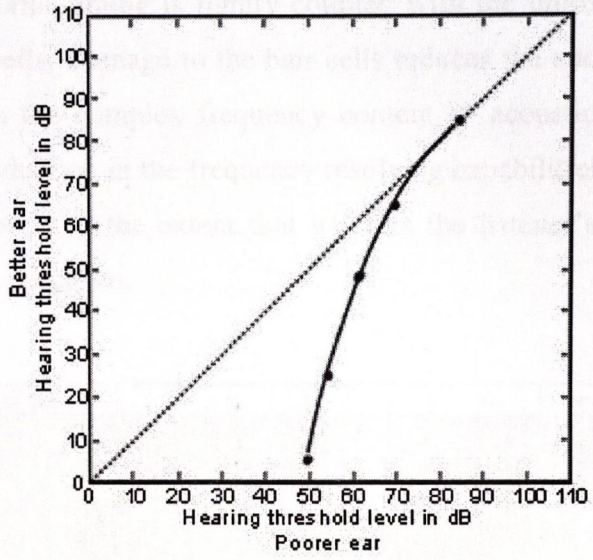


Figure 2.6 Characteristic Recruitment Curve.

Considering the plot in Figure 2.6, the straight line represents the equal loudness growth for two normal ears. The curve represents the growth of the recruited ear relative to a normal ear. Clearly the loudness of a tone at 50dBSPL in the impaired ear is approximately the same as the loudness of a 5dBSPL tone in the unimpaired ear, showing the increase in auditory threshold in the impaired ear. Increasing the intensity in the unimpaired ear by 10dB (e.g. from 10 to 20dBSPL) requires only an increase of approximately 2dB in the impaired ear, showing that the rate of growth of loudness is abnormally large. It is this manifestation of the loss of compression that is called recruitment. In signal processing terms recruitment can be generally described as the non-linear attenuation of sound in the time-domain and more precisely defined as a non-linear loss of dynamic range compression [16].

2.3.2 Frequency Smearing

Considering the many different types of sounds we hear every day it is quite a remarkable feat for the ear to provide us with the ability to distinguish between them and to be able to derive interpretable information from these stimuli. Of particular importance to mankind is the ability to hear speech and most importantly for the ear to be able to

accurately respond to the complex frequency domain resonances within that speech. It is this frequency selectivity that allows us to distinguish sounds and properly code the information for the brain to interpret. The precise frequency selectivity of the auditory system is due to the fine-tuning of the basilar membrane and its selective stimulation of the inner hair cells. This fine-tuning is tightly coupled with the unimpaired operation of the inner and outer hair cells. Damage to the hair cells reduces the auditory systems ability to accurately respond to the complex frequency content of acoustic stimuli. Among other things the resulting reduction in the frequency resolving capabilities of the auditory system affects speech perception to the extent that it limits the listener's access to the formant patterns in the speech spectrum.

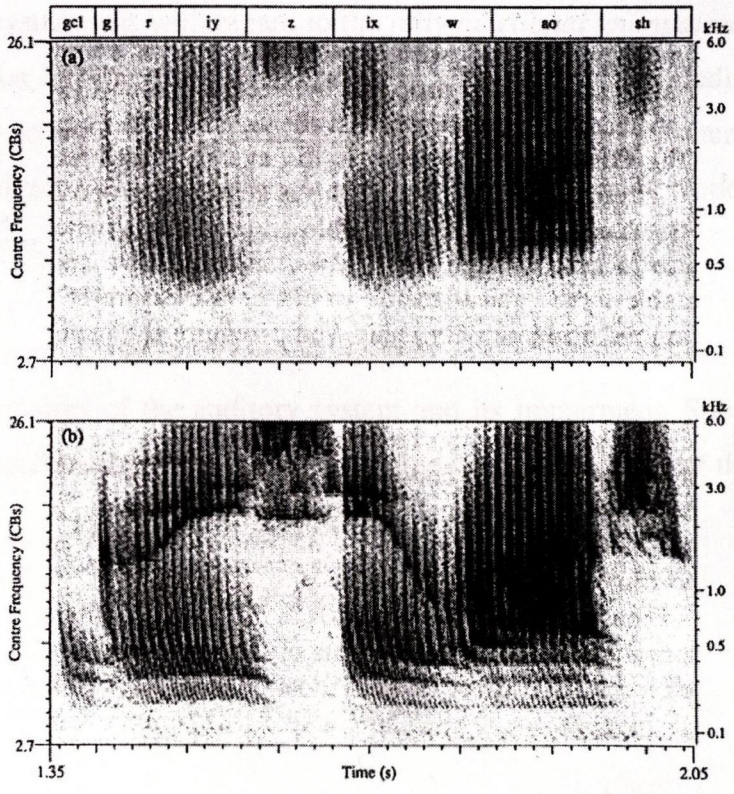


Figure 2.7 Characteristic effects of frequency smearing [17].

Considering the plot in Figure 2.7, the lower plot shows a cochleogram of the response of an unimpaired auditory system to the utterance "greasy wash" in which the formant content and the formant transitions are clearly visible. The upper plot shows the same response with severe hair cell damage clearly illustrating the impaired system's inability to respond to the complex content of the stimulus. This reduction in selectivity gives rise to the impairment of frequency smearing which in signal processing terms can be

described as a process whereby the impaired ear fails to accurately respond to the frequency domain resonances in natural speech.

2.4 Summary and Conclusions

This Chapter has identified the key elements of the anatomy, physiology and operation of the human auditory system with a view towards acquainting the reader with the fundamentals of auditory processing. The information presented highlights the primary aspects of the auditory system as detailed in [15]; however, significantly greater detail on the human auditory system and its operation may be found in [18], [19] and the audiological and medical literature in general. It is important to note that the information on the auditory system was with regard to the *fully functional unimpaired auditory system* and is of particular importance when considering impairment of the auditory system. This was followed by some background on impairment of the auditory system and focused on two sensorineural impairments of particular relevance to the current work: *recruitment* and *frequency smearing*.

The information presented here is designed to provide the reader with an overview of the primary features of the auditory system and its impairment. Subsequent Chapters deal more specifically and in greater depth with the particular aspects of these topics.

Chapter 3 - Modelling Survey

3.1 Introduction

The previous Chapter has presented the operations taking place in the human auditory system, operations that must be addressed when applying digital signal processing to modelling the human auditory system, its impairments and compensatory mechanisms for such impairments. Before examining the development of these models it is necessary to consider the environment out of which this work has grown. Clearly there is a vast quantity of both published and unpublished work in the broad areas of acoustics and signal processing, significant portions of which are relevant to the topics being considered. However, it is neither feasible nor particularly useful to delve into a historical account of all the developments that have taken place to date in this type of survey. That is not to say that this continuing developmental process is unimportant; on the contrary, previous work is an extremely important foundation for both current and future developments. It is not, however, the aim of this work to provide a detailed historical account of this lengthy evolution.

This Chapter specifically focuses on providing a survey of previous works that are most appropriate to the current work and examines both the advantages and disadvantages of the various approaches. The objective is to familiarise the reader with the current state of developments in the area with regard to the work in this thesis. Even under these constraints there is a significant body of work to consider. This survey has therefore been divided into three sections as follows.

- Unimpaired Auditory System Modelling
- Impaired Auditory System Modelling
- Recruitment Compensation System Modelling

It is difficult to draw precise lines of distinction between the sections as there are many interdependencies between them, particularly between modelling the unimpaired and

the impaired auditory system, but where possible each section minimises unnecessary comment on other sections.

3.2 Unimpaired Auditory System Modelling

On examination of the available work on modelling the auditory system two separate approaches can be identified; a macromechanical approach and a micromechanical approach. Typically auditory models have fallen into either one or the other category. However more recently hybrid models have begun to emerge that try to harness the benefits provided by both approaches. Whether a model falls into one category or the other depends on how the cochlear partition is modelled. For clarity it is best to consider each approach and its subsections in isolation prior to examination of the hybrid of these approaches.

3.2.1 Macromechanical Auditory Modelling

When a model is classified as a macromechanical model of the cochlea, this typically means that the model is focused on modelling the fluid dynamics of the cochlea, including fluid interactions with the flexible cochlear partition. These types of models do not however concern themselves with the interactions of the structures within the cochlear partition [20]. At the heart of these types of models lies a set of equations for the dynamics of the fluid in the cochlea in 1,2 or 3 dimensions, the solution of which typically provides the basilar membrane velocity and pressure (the number of dimensions depends on how the geometry of the fluid channels is represented). The solution of these functions is, however, non-trivial and approximations can be used to simplify the process. These approximations take the form of either a *short wave* or a *long wave* approximation. In the short wave approximation the wavelength of the travelling waves on the basilar membrane is assumed to be short compared to the *height* of the cochlear duct and in the long wave approximation the wavelength is assumed to be large compared to the *width* of the cochlear duct. The first formal macromechanical models of the cochlea used the short wave approximation [20] to reduce the complexity of the expression for the propagation constant in the dynamics equations. Unfortunately, the resulting solutions are only meaningful around the resonance place on the basilar membrane which greatly reduces the usefulness of such models. These types of models, however, lead to the more successful long wave approximation, which reduces the dynamics equations to that of equations for a 1-dimensional non-uniform

transmission line [10]. The various macromechanical models can, however, exist in two different forms, that of *classical* and *non-classical* models. The main category of macromechanical models is therefore divided into classical and non-classical models within which there are models employing the short wave approximation, the long wave approximation and no approximations. Clearly the most complex and general model will be that of a 3-dimensional macromechanical model where no approximations are used in its solution [21], [22].

3.2.1.1 Classical macromechanical models

The first of the primary groups within the category of macromechanical models is the classical group. In these types of models of the cochlea, the dynamics of the cochlear partition at a certain location are described by a local parameter [23], [24]. Many models in the literature are described as classical because they have the property that the mechanical reaction of the cochlear partition at a location depends *only* on the movements of the basilar membrane at that location. Considering linear models (all variables are linear functions of frequency), the mechanics of the partition are completely described by the driving-point impedance of the partition and since the basilar membrane mechanics dominate, this is typically called the basilar membrane impedance. This impedance is a function of both basilar membrane location and frequency, and exhibits resonance for a specific frequency at a specific location that is itself dependent on the frequency; therefore the model can map frequency to place on the basilar membrane [25]. For this to be the case the basilar membrane impedance and its inverse (the basilar membrane admittance) must be a *causal* function and thus can be built up with realisable elements. This is one of the main distinguishing features of these types of models.

A typical classical macromechanical model of the cochlea consists of two fluid filled channels separated by a moveable partition that is completely described by its local mechanical impedance. The short and long wave approximations can be used to solve the set of equations describing this scenario along with the more general case, without approximation. In the case of classical models the basilar membrane impedance function is specified prior to solution of the dynamics equations, is independent of these equations and must be specified as a causal function (the admittance function must also be causal). This is a limitation of classical models; it is not always possible to realise the impedance function as a driving-point impedance and a response requiring such an impedance

function cannot be obtained in a classical model with a realisable basilar membrane impedance [26]. Another significant limitation on classical models is that they must neglect longitudinal stiffness in the basilar membrane as the modelling approach ensures the characteristics of every location are independent of surrounding locations. It has, however, been shown that minor amounts of longitudinal coupling destroys the main character of the response of classical models [27].

A large body of work can be found in the literature covering detailed development of classical models and a recent analysis of the characterising equations and the limitations of this class of model can be found in [28], [24], [25] and their associated works. It is also worth noting that analysis of recent experimental data [29] has shown that the typical effective basilar membrane impedance shows clear signs of containing non-realisable components. This would imply that the cochlea does not exactly function according to a classical model, as classical models *must* have realisable impedance functions.

3.2.1.2 Non-Classical macromechanical models

The second group of models are the non-classical macromechanical models of the cochlea in which the dynamics of the cochlear partition are *not* determined by a local parameter [25]. This means that the mechanical reaction of the cochlear partition at one location is additionally controlled by the basilar membrane movements at other surrounding locations. Given that there are several ways in which to model the effects remote sites have on the local dynamics of the cochlea, this group of models encompasses a wider variety of models than the classical group. Several of these models have recently appeared using approaches such as dual interacting wave-propagating structures [30], models incorporating response enhancement by outer hair cells remotely stimulated from where they deliver their effect [31], [32] and 3-dimensional hydroelastic models in which the basilar membrane is treated as a hydroelastic plate [33], to name a few. The development of the non-classical group from the classical can be traced to the expression of the basilar membrane impedance. In a similar manner to classical models, the idea of basilar membrane impedance is retained in non-classical models by defining it in terms of the quotient of the fluid pressure (above the basilar membrane) to the basilar membrane velocity, both of which are only known after solution of the model equations. This non-classical impedance is called the *effective* basilar membrane impedance to differentiate it from the classical basilar membrane impedance. Considering the effective impedance

depends on the solutions to the equations in these models, it can only be determined after the equations are solved because it is not determined solely by local mechanical properties [25], [26]. The effective impedance therefore may not be realisable as the impedance function and/or its inverse may be non-causal. This method of determining the effective impedance does, however, incorporate all the dynamics of the cochlear partition that were used in formulating the model. The distinguishing feature of this group of models is therefore that the impedance is not specified (or known) before solution of the model equations and that the resulting impedance function may be non-causal.
Note: It is not necessary to determine the impedance in non-classical models in order to use them but it is possible to calculate the effective impedance if required.

Typically the form of non-classical models is the same as classical models in that they consist of two elongated fluid filled channels separated by a flexible partition. However, in non-classical models the characterising equations only express properties of the fluid in the channels and their interactions, along with specifications of the mechanics of the organ of Corti. They do not incorporate (or need) a specification of the basilar membrane impedance. It can, however, be shown that in general terms, the response of a classical model which has as its basilar membrane impedance function the effective basilar membrane impedance of a non-classical model, is equal to the response of that non-classical model [26]. This holds provided the two models have the same structure and geometry, the solutions obey the same boundary conditions and the effective impedance is causal. This has been proven for long wave, 2-D and 3-D models. The benefits of non-classical models arise in the far greater flexibility in the form of the effective basilar membrane impedance, which allows the models reproduce the non-minimum phase responses seen in experimental results [34], [29], [25]. Non-classical models have also been shown capable of incorporating longitudinal rigidity without detrimentally affecting the character of the model response, which is important in a realistic representation of the responses of the basilar membrane and its interactions with the cochlear fluids [33], [35].

Increasing numbers of models and reviews of these types of models can be found in the literature. The recent articles by deBoer [22], [21], [25], [26] and Allen [20] (among others) provide a detailed review of the macromechanical modelling approaches along with references to other reviews and models of this type.

3.2.2 Micromechanical Auditory Modelling

In contrast to macromechanical models of the cochlea, micromechanical models of the cochlea focus on modelling the interactions of the structures within the cochlea, specifically the basilar membrane, the tectorial membrane and the organ of Corti [20]. The primary aim of these types of models is to explain the tuning and nonlinear compression of the cochlea from physical principles. Examination of this group of models reveals a large variety of models having different objectives and levels of detail depending on their intended use. However, as the level of information on the physical parameters of the auditory system increases, a greater level of convergence can be seen among this group of models. Typically at the heart of these models lies some form of mechanical network emulating the elements of the inner ear and their interactions. Having specified the mechanical networks, expressions for the various responses can be determined for these networks. A typical example would be the foundation model in [20] for which general expressions are developed that relate the shear force on the hair bundle and the shear velocity between the tectorial membrane and reticular lamina to the basilar membrane pressure and velocity. Having considered this group of models and the various forms and approaches they take, two clear subsections emerge depending on how the fine-tuning of the auditory system is modelled. There are two competing views as to how the fine-tuning occurs giving rise to passive and active models. The category of micromechanical models is therefore subdivided into active and passive models.

3.2.2.1 Passive micromechanical models

The first of the primary groups within the category of micromechanical models is the passive group. Models fall into this group if their underlying explanation of cochlear mechanics adheres to the idea that there are no sources of mechanical energy at the stimulating frequency within the cochlea. The most successful of these types of models is the resonant tectorial membrane model in which the cochlear partition transforms the broadly tuned frequency response of the basilar membrane into a sharply tuned frequency response at the hair cell bundle through resonant tuning of the tectorial membrane to a frequency half an octave below the resonant frequency of the basilar membrane [20]. The fine-tuning therefore arises because of the introduction of a spectral zero half an octave below the resonant frequency of the basilar membrane [36]. The parameters of this type of

model can be selected so that the responses closely match experimentally measured neural threshold tuning curves.

A typical starting point for a passive model represents the organ of Corti, the tectorial membrane and the basilar membrane by their respective masses. The tectorial membrane is coupled to both a rigid body and the organ of Corti by both compliance and damping elements. The basilar membrane is also coupled to a rigid body by a compliance and damping element and is directly connected to the organ of Corti. This type of approach typically emulates the geometry of a radial cross-section of the cochlear partition and can be described by a set of equations representing the interactions of the various elements. Solution of these equations allows analysis of the responses of the model. Clearly more complicate representations of the elements of the auditory system are possible and yield more complex equations describing the cochlear partition interactions. Models based purely on this form have, however, gone into disfavour because they are unable to adequately account for basilar membrane tuning and because it is believed that the outer hair cells act as a source of mechanical energy in the auditory system playing an important role in the fine-tuning of the cochlear partition.

3.2.2.2 Active micromechanical models

The second group of models in the category of micromechanical models is the active group. The underlying idea for models within this group is that sources of mechanical energy do exist at the stimulating frequency within the cochlear partition [20]. Models of this type generally adhere to the form of the cochlear amplifier model in which the sensitivity of the basilar membrane vibration, to low-level sounds, increases due to an active mechanism within the cochlear partition and results in increased frequency selectivity of these vibrations [37]. The cochlear amplifier approach operates by adding mechanical energy from the outer hair cells to the travelling wave near the place of maximum response, in response to a stimulus. This energy is then reabsorbed in other places along the cochlear partition. These types of models can very successfully reproduce the neural threshold tuning curves measured experimentally. Typically the addition of the mechanical energy is modelled by inclusion of regions of negative impedance in the cochlear partition and when the travelling wave encounters these regions they receive an increase in power. The negative damping model in [38] is one of the more recent successful variations of this type of model.

A typical active model generally takes the basic form set out for passive models and incorporates an added mechanism to generate the cochlear amplifier effect. A set of equations is used to describe the interactions of the elements in the mechanical networks and solution of these equations provides access to the various responses of the models. It must, however, be noted that the exact source of the mechanical energy behind the cochlear amplifier effect is unknown but a growing body of support favours the outer hair cells as the source of the mechanical energy as it has been shown that they respond mechanically to electrical stimulus. Where active and passive models differ greatly is in their interpretation of the responses of a damaged cochlea. In passive models it is attributed to alteration in the basilar membrane stiffness whereas in the active model it is attributed to reduction in the cochlear amplification.

Examination of the literature reveals a wealth of passive and active models along with a growing interest in active micromechanical models. More recently, significant effort has gone into the development of hybrid micromechanical models incorporating negative damping and a resonant tectorial membrane, with great success [39], [40], [38], [41]. Further details about micromechanical models can be found in [20] and associated references.

3.2.3 Hybrid Auditory Modelling

Another addition to modelling the auditory system has evolved in the form of a hybrid between the macromechanical and micromechanical modelling approaches. Strictly speaking this category of models are not hybrids, in the true sense of the word. However, they have a firm basis in the macromechanical and micromechanical modelling approaches. It can be seen from examination of the macromechanical and micromechanical approaches that both have evolved from direct modelling of the physiology of the auditory system. These types of models therefore rely upon accurate measurement of the characteristics of physiological components of the auditory system. However, many of the characteristics that must be measured alter significantly post-mortem. This poses a significant problem for these types of approaches, as for ethical and moral reasons much of the required information cannot be measured in a live subject so that many of these parameters must be estimated from humans post-mortems and animal research data. As advances in measurement techniques have evolved along with an increasing knowledge of

the auditory system, these types of models have seen greater success and many of the required parameters can and have been relatively accurately assessed.

Where the hybrid approach differs from the micro/macro mechanical approaches is that many of the components of these models have *no* direct physiological basis whatsoever. Instead of directly modelling the physiology of a particular component of the auditory system, the *responses* of the component is modelled. In general, these models are designed to emulate the range of responses of the various components in the auditory system to the range of stimuli each component may receive. A typical example of this would be modelling the travelling wave on the basilar membrane as a wave digital filter representation of a travelling wave propagating along a 1-dimensional non-uniform transmission line [42], [17]. Clearly the basilar membrane and a set of wave digital filters are two very different entities but they do share similar behavioural properties particularly in the representation of how travelling waves propagate.

Where the similarity exists between hybrid models and macromechanical models is in the responses of macromechanical models. Solution of the dynamics equations used in macromechanical models typically yield information such as the basilar membrane velocity and pressure; these types of responses are what hybrid models emulate. A typical example would be modelling the travelling wave propagation along the basilar membrane and the resulting broad tuning in the responses of a filter bank. This form of hybrid model does not contain any information regarding the fluid dynamics or its interactions with the cochlear partition but can still produce the same effect as a travelling wave propagating along the basilar membrane due to fluid interactions with the basilar membrane. Similar correlations exist between hybrid models and micromechanical models. The interactions of the components in the cochlear partition in micromechanical models are derived from modelling the physiological components in the cochlear partition and developing equations regarding these interactions, solution of these equations allows exploration of their nature. The responses that occur due to these interactions between the difference components form the basis for hybrid models [43]. Typically hybrid models focus heavily on modelling the responses of the cochlear partition including the basilar membrane, tectorial membrane and the organ of Corti. However, there is generally little physiological information used or required in the specification of these models.

In general, hybrid models have a foundation in the responses of physiologically based models and in psychoacoustical evaluation of the human auditory system and its responses. This can especially be seen in the emulation of the fine tuning of the auditory system in these models which typically depends on physiologically measured neural tuning curves in animals and psychoacoustically measured tuning curves from notched noise experiments in humans. Just as for the other main categories of models, hybrid models can be subdivided into two main groups identified by how the tuning of the cochlear partition is emulated. Each group uses a different method for implementing tuning; one uses transmission lines to emulate the tuning of the basilar membrane, the other uses a filter bank approach.

3.2.3.1 Transmission line based hybrid models

The first group within the category of hybrid models based on a transmission line representation of the cochlear partition is the closest model to a direct hybrid of the macromechanical and micromechanical approaches. These models generally employ a lumped parameter transmission line to model the wave propagation within the cochlea and are closely related to the 1-dimensional macromechanical model of the cochlea using the long wave approximation in the solution of the model dynamics [44], [45], [46]. Recent versions of these types of models have incorporated the micromechanical resonances associated with the tectorial membrane and outer hair cell interaction with the basilar membrane and have become hybrid models in the true sense of the word [47], [39], [48]. Older versions of this type of model that did not incorporate the micromechanical interactions tended to have problems reproducing tuning curves but their later incorporation has solved many of these problems. These models do, however, have several drawbacks, namely they require large numbers of sections and small time steps making them very computationally expensive to operate [49]. Greater detail on these types of models can be found in [44], [50], [39] and [48].

3.2.3.2 Filter bank based hybrid models

The second group of models within the hybrid category are those employing some form of filter bank approach to modelling the wave propagation in the cochlea and the tuning of the basilar membrane and associated structures. Typically there are two possible forms of filter banks, either cascade or parallel. Cascade filter banks generate their fine-

tuning through pseudoresonance whereby active sections are cascaded together and the system achieves its overall sharpness through the build-up of the contributions of the sections in the cascade [51]. None of the individual sections are in general highly tuned and typically exhibit a low pass type response. The contributions of all sections up to a point generate the so-called auditory filters that represent the filtering effect that occurs when a travelling wave propagates along the basilar membrane [52]. In contrast, parallel filter banks rely upon a parallel bank of overlapped highly tuned filters where each filter is specifically designed to model the filtering exhibited by the basilar membrane [53]. These filter are generally highly tuned and typically exhibit a band pass type response, the simplest of which are the rounded exponential (roex) shaped filters [54], [55]. Both approaches quite successfully emulate the fine-tuning that occurs in the auditory system at significantly reduced sampling rates and numbers of filters. The computational load of these structures is therefore greatly reduced compared to transmission line based models.

There are increasing numbers of hybrid models appearing in the recent literature, greater detail on some of the more outstanding complete auditory models can be found in [55], [42], [17], [52], [49] and their associated references.

3.2.4 Middle and Outer Ear Modelling

From an examination of the information presented in the previous section, it can be seen that the various forms of the models have been segregated on the basis of how the inner ear is modelled. This is because the inner ear is responsible for the transduction of sound to neural excitation and is hence the site of possible sensorineural impairment, which is of particular importance in this work.

Considering the outer ear, its primary function is to channel sound to the tympanic membrane and can be considered to be an approximately cylindrical acoustic resonator providing broad gain around its first and second normal modes of vibration [42]. Models of the outer ear typically take the form of uniform lumped parameter transmission lines [42], [56], [57] or transfer functions introducing suitable gain at the resonant frequencies [54]. More often than not, the external ear consisting of the upper torso, head and pinna are not included in these models particularly in monaural signal processing applications.

Briefly considering models of the middle ear, they take the form of both physiological and response based models just as for inner ear modelling. Physiological based models are typically finite element models emulating the mechanical interactions of the components of the ossicular chain in the middle ear. These types of models are ideal candidates for the development of replacement components for the middle ear. However, they are not particularly useful when modelling the signal processing taking place in the middle ear. A significant portion of middle ear models used in signal processing models of the auditory system take the form of electroacoustic network models [56], [42], [59] or transfer function representations [54], [52], [49].

Electroacoustic network models aim to represent the effective middle ear impedance and reflectance in order to model the pressure delivered to the round window. These types of models are based on the physiological responses of the middle ear and incorporate the gain provided by the middle ear to overcome the impedance mismatch between the outer and inner ears. In general, these types of models represent with great detail the characteristics of the middle ear and typically reside in auditory models that accommodate the impedance of sound in air (in the outer ear) and in fluids (in the inner ear). The representation of the impedances of the middle ear facilitates attenuation of very low and very high frequencies as seen in the responses of the middle ear. In contrast, transfer function based models have no physiological basis and typically introduce into signals passing through them the effects the middle ear has on signals passing through it. This generally amounts to both low and high frequency attenuation along with minimal gain across the normal frequency range of hearing. These types of models do not emulate the losses of the middle ear or the impedance mismatch between the outer and inner ears. Therefore the typical gain provided by the middle ear is not incorporated into these models as the losses are not present. These types of models typically reside in response based auditory models in which it is accepted that the middle ear accommodates transfer of signals from air to a fluid with minimal mid-frequency loss and attenuation of both very low and very high frequency signals.

In general both types of middle ear models allow signal transfer from outer ear models to inner ear models, with the resulting inner ear signals being very similar. The primary difference between the two is in the level of detail each type of model incorporates. This is especially important when conductive hearing loss is to be simulated where precise control over all the parameters of the middle ear is desirable. However,

when concentrating on sensorineural hearing loss such a level of detail is not required and it is much more beneficial to alter a transfer function to mimic a reduction in the middle ear's capabilities rather than manipulate a myriad of parameters to generate the same effect. This is of prime importance in the current work as the focus is on sensorineural hearing loss and not conductive hearing loss, therefore highly detailed control over the parameters of the middle ear is neither desirable nor required. The significant reduction in computational complexity by using transfer function representations of the middle ear allows greater focus to be placed on the inner ear where it is more important when dealing with sensorineural impairment.

3.2.5 Auditory Modelling Discussed

Considering this section on unimpaired auditory modelling, in can be seen that all models follow one of two basic methodologies within which there are several options as to the form of the model. Either a model emulates the physiology of the auditory system or it emulates the responses of the auditory system, a subtle but important difference when deciding how to approach the complex problem of modelling the human auditory system. The decision as to which path to take can be made at a simplistic level by considering only the basic advantages and disadvantages of each approach and selecting the most advantageous option. However, this will not provide the optimum model. The most important consideration must be given to the ultimate use of the model. Obviously if the ultimate aim is to exactly replicate the operation of the auditory system this can only be done using some form of physiological model. However, if the aim is to provide access to the responses of the various components of the auditory system then the excess complications associated with physiological modelling can be avoided by developing models that characterise the modes of operation of the various components of the auditory system and how they respond to the stimulation they provide each other. This form of approach focuses on emulating the responses of the auditory system rather than how those responses are generated by the components of the auditory system and in this manner can avoid the problems associated with accurate measurement of inaccessible physiological parameters and their complex interactions. The most appropriate form of models to achieve this type of goal are hybrid models and this is the form chosen for the auditory model in this work, as the primary use being made of the auditory model here is to consider the responses of the auditory system when it is suffering from impairment.

This section on unimpaired auditory modelling has identified the various forms of models that exist and has highlighted some of their advantages and disadvantages. The following Chapter will deal with developing an auditory model and will identify the limitations of the various methods available and justify the choices made for the auditory model used in this work.

3.3 Impaired Auditory System Modelling

One of the main goals in this work is to model sensorineural hearing impairments, specifically recruitment and frequency smearing. Having examined the ways in which the unimpaired auditory system has been modelled, this section identifies how the impaired auditory system has been modelled in the literature. Considering the variety of sensorineural impairments and the equally numerous models emulating these impairments, this section focuses on those modelling approaches most appropriate to recruitment and frequency smearing. An examination of the literature reveals two categories by which to differentiate the various models. The first category encompasses models that emulate the impaired auditory system and the second encompasses models that replicate the effects of an impairment.

3.3.1 Modelling the Sensorineurally Impaired Auditory System

This first category identifies those models that emulate the sensorineurally impaired auditory system. The general approach is to model the unimpaired auditory system and introduce the impairment into the unimpaired model in order to generate a model of the auditory system suffering from a particular impairment. The process of introducing an impairment into the unimpaired model depends on how the unimpaired model has been designed and more specifically on the form of the unimpaired model. Comparison of the responses of the unimpaired and the impaired models allows analysis of the effects of a particular impairment.

Consider macromechanical and micromechanical models of the physiology of the auditory system. In order to introduce an impairment into these models the cause of the impairment must be identified (as in which components are malfunctioning) and then what is exactly wrong with these components (the source of the components malfunction). The ultimate cause of an impairment must therefore be linked to specific internal physiological

disorders of components of the auditory system in order to introduce the necessary malfunctions into the specific components of an unimpaired auditory model. The isolation and quantification of the necessary parameters to do this is difficult and in many cases cannot be done in live subjects. Examining the available information on impairment modelling, it can be seen that the majority of models avoid using physiological models because of the difficulty associated with parameterising these models. This is the main reason why most impairment modelling is carried out using hybrid models. Hybrid models of the auditory system focus on emulating the responses of various components in the auditory system and the interaction of the various responses. Introduction of an impairment into these types of models is a less complex procedure. The source of an impairment must still be isolated down to particular components malfunctioning but it does not require the cause of the malfunction of the components to be quantified, only the alteration in the responses of the components. Once this has been assessed, the methods used for generating the responses of the components in the unimpaired auditory model can be altered to reflect the changes that have taken place in the auditory system to generate a model that emulates the impaired auditory system.

3.3.1.1 Modelling recruitment

Although first observed back in 1936 by Mark Gardner, recruitment has until recently been poorly understood and as a result the primary focus in the past was not on modelling recruitment but on modelling the *normal* relationship between the intensity of physical stimuli and the psychophysical (loudness) response. The result of lengthy heated discussion and research on the topic over the years has resulted in agreement upon a cube root relationship and now any power-law relation between the intensity of a physical stimulus and the psychophysical response is referred to as Stephen's law [60], the history of which can be found in [61], [60] and associated references. It suffices to say that the fruits of these endeavours along with experimental results [62], [63, [64] have confirmed that recruitment occurs in the cochlea and is caused by damage to the outer hair cells. The outer hair cells are responsible for dynamic range compression on the basilar membrane. Damage to the outer hair cells causes loss of this compression and the cube-root loudness growth that starts with the nonlinear compression of the basilar membrane motion is thus altered causing an abnormally rapid growth in the loudness level [65], [66].

In terms of introducing recruitment into models focusing on the responses of the auditory system there are several ways in which this can be done. However, three broad methods exist which cover the range of approaches employed.

The first method relates to those models that have a power law relationship at their core. Typically, transfer through the outer and middle ears is represented by a fixed transfer function followed by an auditory filter bank to model the filtering of the inner ear, which generates an excitation pattern. This excitation pattern then feeds a power law relationship, which generates the resulting loudness. Many of the current models using this method have their roots in the model of loudness perception in normal-hearing subjects proposed by Zwicker [67], [68]. Examination of these models reveals two possible approaches by which recruitment can be introduced, a one-component approach and a two-component approach. In the one-component approach the effects of recruitment are introduced by either adding internal noise [69], [70] or by alteration of the absolute threshold in the power law relationship [71], [72]. This type of one-component approach assumes that hearing impairment is comparable to a masking condition in normal hearing subjects and causes a steeper than normal increase of the loudness function, the net effect of which is audiometric threshold shifts and recruitment. However, these approaches are unable to reproduce the reduced loudness summation[1] that should occur with impairment and requires a threefold increase of the auditory filter bandwidth to approach the measured values, which is not unreasonable [73]. It is important to note that this approach does *not* alter the exponent of the power law. In comparison, the two-component approach takes a different view of things and involves *increasing* the exponent in the power law relationship [73]. Increasing the exponent of the power law has the effect of reducing the compression, which results in steeper than normal increases in the loudness function thus introducing recruitment. In conjunction with alteration of the power law exponent, a frequency dependent attenuation is typically applied to the excitation level within each auditory filter to incorporate the selectivity loss that occurs with this type of impairment. Greater detail of these methods can be found [73], [69] and associated references.

The second method relates to those models that take a statistical approach to modelling the auditory system, usually focusing on the inner ear operations. In general,

[1] When loudness summation is reduced this means that the loudness of a sound of fixed overall intensity no longer increases with increasing bandwidth beyond a certain bandwidth, called the critical bandwidth for loudness.

these types of models take a lower level approach to modelling the auditory system than the first method. These models are typically built up by moving through the inner ear one inner hair cell at a time and characterising each location by randomly sampling the distributions appropriate to the characteristic frequency of each location [74], [75]. Typical parameters that would be characterised are the spontaneous firing rate, thresholds, slopes of rate-intensity functions, maximum discharge rates, tuning sharpness and firing rate adaptation. This builds up parameter distributions that are physiologically correct at each centre frequency across the entire neural population [74]. When these types of models are stimulated they generate neural excitation patterns by calculating the response of each fibre in the simulated auditory nerve based on the characteristic parameters at each location. It is known that loudness is related to the total neural spike count [76], [77] in the active neural population allowing representation of the loudness response of the auditory system to acoustic stimuli. Recruitment is typically introduced into these types of models by specifying a reduction in the spatial distributions of the degrees of outer hair cell function. These types of models focus on the neural firing on the auditory nerve fibres and as such do not incorporate direct models of the outer hair cells. However, the characteristic distributions used to parameterise such models of the auditory nerve are directly dependent on the degree of hair cell function. Reduction in the degree of outer hair cell function models outer hair cell damage, which alters the characteristic distributions of the model. This results in threshold shifts, changes in tuning properties and alteration of the suprathreshold responses [40] which alter the neural firing and hence the loudness. The overall effect is an increase in the threshold and an abnormally rapid growth in loudness.

The third method relates to those models that focus on the interactions of the components of the auditory system. These models typically take the lowest level approach to modelling by incorporating all the various elements of the inner ear and their responses. Transfer through the outer and middle ears, if included, is generally represented by fixed transfer functions or electric circuit models [55], [54], [42]. As with most models targeting sensorineural impairment modelling, the inner ear models are generally highly detailed, right down to the level of basilar membrane motion, tectorial membrane motion, outer hair cell contributions to compression and tuning and inner hair cell mechanical to electrical transduction [55], [42], [17], [52], [78]. Depending on the exact level of detail and particularly if the outer hair cell action is modelled, recruitment can be introduced into these types of models by specification of a reduction in the degree of outer hair cell function. This will have the effect of increasing thresholds, altering the tuning and

compression of the basilar membrane and alteration of the stimulation of the inner hair cells. The result of this will be changes in the inner hair cell firing and hence in the neural firing on the auditory nerve fibres with the net effect of an abnormally rapid growth in loudness and increased thresholds. If, however, the outer hair cell action is not explicitly incorporated into such a model, the effects of a reduction in the degree of outer hair cell function can be incorporated in a manner similar to that presented in the second method above. Specifically the tuning of the basilar membrane is highly dependent on the degree of outer hair cell function and its reduction broadens the tuning on the basilar membrane and reduces the compression. The factors directly controlling these characteristics, in a model without outer hair cell interactions, can be altered to incorporate the effects of outer hair cell malfunction. The models presented in [79], [80] use this type of approach where specification of a gain component for tuning curve tips is reduced causing a reduction in the compression and hence introducing recruitment.

3.3.1.2 Modelling frequency smearing

Modelling frequency smearing in auditory models has not seen as prolific a coverage in the literature as recruitment. However, it can potentially be introduced into any model that reproduces the frequency selectivity of the auditory system. The prime causes of frequency smearing are broadening of the tuning of the basilar membrane and a reduction in the degree of inner hair cell function. The net result of these defects is a reduction in the ability of the auditory system to accurately respond to the frequency domain resonances in acoustic stimuli.

This reduction in frequency selectivity causes broader excitation patterns and reduces the auditory system's ability to resolve the individual components of complex signals [81]. Specifically for signals such as speech, it results in blurring of the spectral envelope and reduces the auditory system's ability to detect and discriminate between formants causing reduced intelligibility [82]. Considering that the sharp tuning of the auditory system is dependent on the interaction of the outer hair cells with the basilar membrane and the resonance of motion between the basilar and tectorial membranes, if the degree of function of the outer hair cells is reduced in conjunction with a reduction in the resonance of motion between the basilar and tectorial membranes then an auditory model will experience broadening of its tuning and experience reduced frequency selectivity. This in turn will broaden the excitation to the inner hair cells, which in turn will generate a

broadened neural excitation, characteristic of an auditory system suffering from frequency smearing. The frequency smearing effects are further enhanced by a reduction in inner hair cell function.

Introduction of frequency smearing into an auditory model of the type containing a direct representation of a power law relationship would require direct manipulation of the auditory filter bandwidth to broaden its tuning. Models taking a statistical approach to modelling the auditory system would require alteration of the distributions describing the tuning at a specific characteristic frequency to broaden the tuning taking place in these models. Similarly the distributions governing the spontaneous and maximum firing rates would be altered to reduce the degree of inner hair cell function. Models taking a low level approach to emulating the auditory system would require a reduction in the degree of function of the inner and outer hair cells along with a reduction in the resonance of motion between the basilar and tectorial membranes to implement frequency smearing. Clear examples of the effects of such an approach can be seen in [17].

Despite this potential to model the auditory system suffering from frequency smearing, very little in the way of frequency smeared auditory models can be found in the literature. Frequency smearing has, however, received significant consideration in terms of reproducing its effects for experimentation with unimpaired listeners (see section 3.3.2.2).

3.3.2 Modelling the Effects of Sensorineural Impairments

The second category identifies those models that emulate the effects of a sensorineural impairment. These models are designed to simulate the effects of an impairment to provide an unimpaired listener with a realistic demonstration of the effects of a sensorineural impairment [83]. In general, these types of models do not fall into any of the macromechanical, micromechanical or hybrid categories of auditory models because they typically have neither a physiological basis nor a basis in the responses of the components of the auditory system. Instead these models focus purely on introducing, into an acoustic signal, the necessary alterations so that an unimpaired auditory system responds to it in a similar manner to that of an impaired auditory system stimulated with a similar unaltered acoustic signal. Where these types of models have their primary uses are in training unimpaired people to deal with people who suffer from auditory impairment, in

the subjective testing of auditory compensation systems and in the psychoacoustic exploration of auditory impairments.

3.3.2.1 Modelling the effects of recruitment

When considering the development of a model that can reproduce for unimpaired listeners the psychoacoustic performance of impaired listeners suffering from recruitment, these models must be able to introduce both the shifts in auditory thresholds and the abnormally rapid growths in loudness experienced by impaired listeners. Examination of the available data reveals three possible methods for the introduction of the necessary effects.

The first and most successful method involves the use of *dynamic level expansion* in which level dependent expansion is responsible for producing elevated thresholds and abnormal loudness growth. A condensed explanation of the theory behind this method is best considered in terms of the detailed work of Lum [83] and Duchnowski and Zurek [84]. Consider Figure 3.1, the solid line represents the loudness matching function for an impaired listener and the dashed line represents a similar curve for a normal listener.

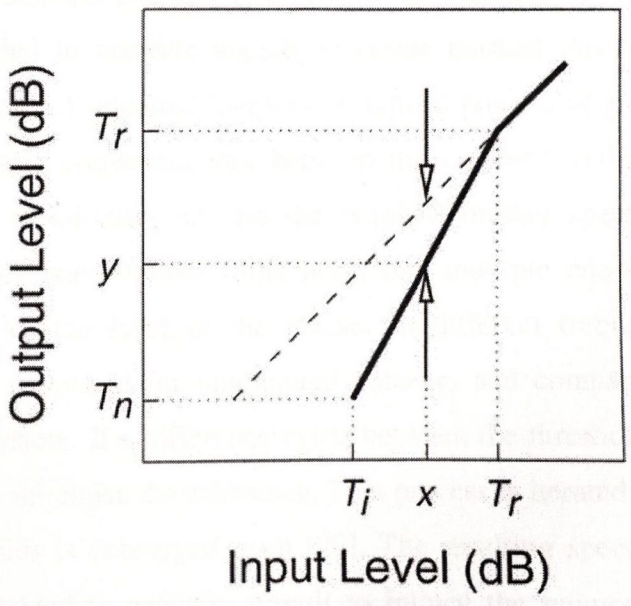


Figure 3.1 Typical input-output function of an expansion simulator.

According to the loudness matching function, the sound level x is equally loud in an impaired ear to the sound level y in an unimpaired ear. In order that an input signal presented to an unimpaired listener will be perceived at the same loudness as signal level x

for an impaired listener, it must be attenuated by the amount indicated by the arrows so that the output level is y. Above the threshold of recruitment (T_r) loudness is assumed equal and no attenuation is required; however, as the sound level is reduced below T_r the attenuation progressively increases. When the sound level gets below the impaired auditory threshold (T_i), the level of attenuation is arbitrary as the signal is inaudible being below the unimpaired auditory threshold (T_n).

Even though this method operates using level-dependent attenuation, it is classified as dynamic level expansion because the output level changes are greater than the input level changes. In these types of models the signal is typically divided up into multiple frequency bands each having their own input-output characteristic depending on the level of impairment at each specific centre frequency. Greater detail on the theory and particularly the implementation of this method can be found in [85], [86], [87], [88] and [89]. It is worth noting that several different versions of this method exist in the literature; in particular, this method crops up under the title of automatic gain control with equalisation. However, they are all versions of dynamic level expansion.

The second method involves the use of *additive noise* to induce the effects of elevated detection threshold and abnormal growth of loudness. In this method, spectrally shaped noise is added to acoustic signals to create masked thresholds for unimpaired listeners equal to those of impaired listeners. A typical process of generating the masking noise utilises the linear correspondence between the masker level in a critical band and threshold elevation to initially estimate the required masker spectrum. The masker is generated by splitting band-limited white noise into multiple one-third octave bands to allow specification of the level of the masker at different frequencies. The masking thresholds are then estimated for unimpaired listeners and compared with those of the impaired auditory system. If a difference exists between the thresholds, the masking level is suitably altered to minimise the difference. This process is iterated until a suitable match between the thresholds is converged upon [90]. The resulting spectrally shaped masking noise can then be added to acoustic stimuli to induce the required effects. In general, additive noise produces a simulation of speech reception performance that is reasonable [90], [7], [91], [92], [93]. However, the ever present noise is unrealistic for demonstration purposes and limits the level of hearing loss that can be simulated [84].

The third method involves using *centre clipping* to produce threshold elevation with recruitment. It entails introducing dead-zones into acoustic signals which have the effect of simulating various types of hearing loss i.e. flat-loss, sloping-loss etc. by altering the simulated thresholds of detection in various channels independently [94]. The overall effect is to cause elevated thresholds and an abnormal growth of loudness near threshold. However, the use of centre clipping introduces gross distortions that are audible to the unimpaired auditory system and which are not part of the abnormal loudness growth [84], hence this method is not widely used.

3.3.2.2 Modelling the effects of frequency smearing

The basic idea behind all models of the effects of frequency smearing is to broaden the spectra of test signals so that the excitation pattern produced in a normal ear resembles the pattern that would be produced in an impaired ear using unprocessed signals [81]. There have been several methods used to simulate the effects of frequency smearing. Primary among these efforts have been the use of vocoders whose output consists of overlapping bands of noise [95], the introduction of temporal jitter in an analog delay line [94], spectral smoothing in the frequency domain [96], [97], [81], [98] and time domain techniques involving modulation of the speech waveform by low-pass filtered noise [99], [100].

The most successful of these methods has been those employing spectral smoothing in the frequency domain. These models utilise analysis/synthesis systems in which the input acoustic signal is divided up into analysis frames by windowing and the short-term complex spectrum of each frame is generated using the Fourier transform. The spectrum is then smeared and converted back into the time-domain using an inverse Fourier transform. The resulting smeared frames are then recombined using an overlap add algorithm to generate the processed continuous output. Of particular interest are the different smearing processes used at the core of this method. One such approach, for sinusoidal stimuli, takes each power component (magnitude squared) of the short-term complex spectrum and replaces it with a weighted sum of the surrounding components, where the weighting function is the broadened auditory filter centred on the component, normalised by the equivalent rectangular bandwidth associated with that filter. The smeared power spectrum is then combined with a random phase spectrum drawn from a uniform distribution $(0^0 - 360^0)$. This has the effect of transforming the sinusoidal stimulus into a noise with

the desired magnitude spectrum; however, the subjective quality of the output signal is quite different to the original signal and is a limitation of this approach [81]. An enhanced version of this approach, applicable to speech stimuli, has been developed in which the original phase spectrum is preserved, tending to preserve the signal waveform rather than replacing it with noise. This is done by incorporating the inverse of the normal auditory filters into the weighting function to correct for the unimpaired subject's own auditory filters when generating the smeared power spectrum and using the original phase spectrum (from the short-term complex spectrum) when generating the smeared complex spectrum of the analysis frames prior to inverse Fourier transforming [98], [101], [85]. The effect of these enhancements is to produce excitation patterns in the unimpaired auditory systems that are a better approximation of the excitation patterns in impaired auditory systems. Another approach to spectral smearing involves calculation of the short-term spectral envelope based on the power spectrum. Having been projected onto a log-frequency scale the envelope is convolved with a Gaussian-shaped filter of fixed relative bandwidth, smearing the envelope. The equivalent rectangular bandwidth of the Gaussian filter determines the degree to which the envelope is smeared. The smeared envelope is then imposed on the original spectrum causing smearing of the spectral energy prior to inverse Fourier transforming. This approach to spectral smearing maintains the signal's phase and harmonic structure in the processed output and works well for both noisy and periodic signals [97]. In general the spectra of the output signals for these types of processes show reduced spectral structure compared to the input signals. However, there is typically greater spectral structure in the output than intended. It must also be noted that when voiced speech sounds are processed, even with high degrees of smearing, the resulting stimuli still sound voiced when the phase spectrum is coherent. When these stimuli are presented to unimpaired listeners they tend to perform better than would be expected with the degree of smearing that is introduced at a frame level. This can, however, be catered for by increasing the level of smearing.

Considering some of the other methods using noise bands, they typically filter the signal into bands and multiply each band by a bandpass noise [95], [100], [99], [102] or some variation on this theme. Typically the bandwidth of the noise determines the degree of spectral smearing; the bandwidth of the noise often increases with centre frequency to emulate the way the auditory filter bandwidth increases with frequency. Expansion beyond the auditory filter bandwidth introduces smearing. A major problem with these approaches however is that the multiplication with random noise introduces *spurious amplitude*

fluctuations that are uncorrelated with the input signal and are reflected in the excitation patterns they generate, making them less suitable for the purpose of testing compensation systems.

An examination of the available work in this area reveals a variety of ways in which the effects of frequency smearing can be introduced into signals; the sections above provide an overview of the more successful approaches used. Greater detail can be found on these methods in the associated references and in the literature in general, [102], [96], [99], [103].

3.3.3 Impairment Modelling Discussed

Just as for modelling the auditory system it is important to identify the ultimate use of an impairment model before deciding on how it should be developed. The sections above identify two primary categories in which impairment modelling can be classed along with several sub-categories. In general, impairment modelling takes the approach of either modelling the impairment or else modelling the effects of an impairment. Whichever approach is employed typically dictates the use of the model. Considering the uses of such models, it would be fair to say that models of an *impairment* are better suited to objective analysis of an impairment whereas models of the *effects of an impairment* are better suited to a subjective analysis. The reason for this is that the responses of an impaired and unimpaired model are quantifiable and therefore directly comparable whereas models of the effects of an impairment are open to subjective interpretation as their effects are not directly quantifiable (except under extreme cases e.g. audible/inaudible). That is not to say one is better that the other but that each is better suited to particular situations.

In terms of the current work, models of an impairment are more suitable as they allow quantification of the effects of an impairment and also facilitate quantification of the benefits a compensation mechanism may have in counteracting the effects of an impairment. Ultimately a compensation system would quantifiably match the responses of a compensated impaired auditory model to that of an unimpaired auditory model. This is only one test of the success of a compensation mechanism. Having accomplished this, such a mechanism could be tested in conjunction with a model of the effects of the impairment to subjectively examine the effectiveness of the mechanism. Presenting an acoustic stimulus, that has been processed to generate an impaired response in an unimpaired

auditory system, to an unimpaired individual in conjunction with the compensation system should evoke the same responses as the unprocessed sound in the same listener without compensation. If this is the case the compensation mechanism is a possible candidate for alleviation of auditory impairment. Clearly in this scenario both types of models fulfil a very useful role.

3.4 Recruitment Compensation System Modelling

The development of models of the unimpaired and impaired auditory system not only allows exploration of these systems but also facilitates the development of compensation systems for auditory impairments. This section specifically identifies approaches that have been taken towards the development of compensation systems for recruitment. The inherent complexity and variety of models of this impairment have led to distinct categories of approaches in the development of compensation mechanisms for this impairment. Two main categories exist, linear amplification with automatic gain control (AGC) and amplitude compression with equalisation [4].

3.4.1 Recruitment Compensation by Linear Amplification

Although not particularly successful at compensating for recruitment under real world conditions, linear amplification with AGC has been shown to be quite successful under controlled laboratory conditions where test stimuli are usually of relatively high overall intensity and limited dynamic range. The primary goal of such systems is to raise as much of the speech signal above the threshold of hearing as possible while applying AGC to keep signals below the threshold of pain [4]. The problem with this is that typically the loudness level is determined by only one or a few frequency bands and the overall gain setting for comfortable listening is such that other frequency bands are inappropriately amplified. More recently, linear amplification has been combined with frequency-response shaping in order to ensure all frequencies contribute equally to a signal's loudness [3]. The rationale behind this is that if signals are amplified so that all frequency bands contribute about equally to its loudness then this will maximise the amount of signal that is audible (averaged over a wide frequency range) and should provide the best speech intelligibility [104]. This provides better audibility; however, increasing the audibility by increasing the linear gain can only be done to a limited extent in the presence of recruitment. Once the signal goes above the increased absolute threshold, the rate of growth of loudness level

with increasing level is greater than normal and often catches up with that in the normal ear. This means that the range of sound levels over which sounds are audible *and* comfortable is much smaller for the hearing impaired thus limiting the range over which linear amplification is useful [3]. Under laboratory conditions these techniques work quite well (within limits) and frequently outperform other methods. However, in the presence of background sounds it can become impossible to apply sufficient linear gain to ensure the speech spectrum is above threshold while preventing the background from becoming unpleasantly loud [105], [85], [3]. Unfortunately in real world situations this is typically the case. Despite this, linear amplification is typically the benchmark against which other methods are tested as it is well understood and research into its usage is relatively mature.

3.4.2 Recruitment Compensation by Amplitude Compression

There are several different approaches to introducing amplitude compression into acoustic stimuli but they all have the same primary goal, to compress the normal dynamic range of hearing into the reduced range of the impaired subject so they experience sounds of comparable loudness to that of an unimpaired listener. There are varying rationales behind the use of compression systems but they all incorporate this desire to reduce the dynamic range of signals [8]. Considering the forms of compression that have been explored in the literature, they can be arbitrarily sub-divided into two differing categories, amplitude compression and parametric compression, each encompassing several different approaches to achieving a similar goal.

The first category focuses on those methods that use some form of amplitude compression to manipulate the dynamic range of a signal. These methods range from long-term average speech spectrum manipulation [6] to multiband processing strategies [106], [107], [108], [20], [109] with several variations of these methods in between. Considering long-term average spectrum manipulation techniques, the typical aim is to "place the long term average spectrum within the loudness contours of the impaired hearing at a position similar to its position within normal hearing" [6]. The rationale behind this is that low level signals should be placed within the lowest loudness contours of the impaired listener and high intensity signals should be placed within the highest contours. In that way all signal levels in the impaired listener are located in a similar relative (loudness) position to those in an unimpaired listener. The placement in the impaired system is relative to that in the unimpaired system because the spread of loudness contours over the dynamic range is

reduced and if the signals in the impaired system are in a similar relative position to those in a normal system then both systems should experience similar loudness levels for the same signal. Methods of this type focus on the signal intensity and applying suitable compression to the average signal level, with limited regard for variations due to frequency. In contrast, multiband processing strategies address both the intensity and frequency dependent nature of the compressive gain required to suitably alter the dynamic range of signals presented to an impaired system. These systems typically separate signals into various bands and apply a compressive gain to each band depending on the input signal intensity [106], [107], [108]. This approach allows for the fact that some bands may stimulate regions of the auditory system that are undamaged and require no compression while applying due attention to those bands that do interact with damaged regions of the auditory system and do require compression. In this manner signals in an impaired system can be made to more closely replicate loudness levels similar to those in unimpaired systems. Recent improvements in methods of this type also incorporate level dependence on the intensity within each band and not solely on the overall input signal intensity preventing low frequency high intensity signal content from dominating the compression gain being applied [110], [5]. Considering the importance of both intensity and frequency in the determination of the levels of compressive gain to apply, multiband processing systems are currently undergoing significant investigation.

The second category focuses on those methods that use parametric compression to manipulate the dynamic range of a signal. Typically these methods perform some form of frequency analysis, the output of which is used to characterise a model for speech. The various characterising parameters for the model are manipulated to introduce compressive gain into the signal and the components of the speech model are then used to synthesise the stimuli presented to the impaired listener. There are two sub-sections depending on how the analysis is performed, either using Fourier transforms [7] or wavelet transforms [4]. When the Fourier transform is the analysis employed, the output of the analysis is used to characterise a model of speech based on a sum of sinusoids with various amplitudes, frequencies and phases. Each component represents a frequency band and compression gain is applied to the amplitude parameters to alter the dynamic range of the stimulus. When the wavelet transform is employed, the output of the analysis is used to characterise a model of speech based on a sum of wavelets. The wavelet coefficients determine the signal characteristics and compression gain is introduced by manipulating the wavelet

coefficients. These methods are quite similar to multiband compression and generate similar results [111].

Generally speaking using amplitude compression to overcome the effects of recruitment performs about the same as linear amplification when the background noise is not fluctuating, is low level and the type of loss is a flat loss. These types of scenarios are, however, somewhat artificial as real world stimuli typically contain fluctuating background signals in competition with the desired signals and in many cases the type of loss is often sloping. Under these real world conditions compression has been shown to be more beneficial than linear amplification [3], [112], [8], [105]. This topic is revisited in greater detail in Chapter 7, Section 7.4 providing greater detail of the mechanisms involved. In terms of the hearing aids that use some form of compression, the reader is referred to a recent review by Dillon [8] which identifies and comments on how compression has been employed in many current hearing aids. It is worth noting that the common thread throughout this extensive review is that there currently is no one strategy that optimally compensates for the loss of dynamic range but that compression in its various forms can provide significant benefit to the hearing impaired.

3.4.3 Recruitment Compensation Discussed

This section has dealt with modelling recruitment compensation strategies of which there are primarily two. The most well known of these is linear amplification typically employing some form of automatic gain control to avoid providing stimuli at or above the threshold of pain and often with the added benefits of frequency-response shaping. The less well know strategy, but one that has grown significantly in its popularity, is the use of compression in its many forms, e.g. wideband compression, multiband compression and parametric compression. The decision as to which approach to take depends on the form of the impairment and the environment in which the compensation mechanism must operate. As indicated above linear amplification is not particularly useful for recruitment compensation in the presence of fluctuating background noise with sloping hearing losses; compression has shown itself to be far more versatile at accomplishing this task. Although there are several different hearing aids using compression of some type, they are all based on some form of the three possible compression methods mentioned; see [8] for an extensive review.

This section has identified the primary strategies employed in recruitment compensation and indicated some of their limitations. Chapter 7 will deal with the development of a recruitment compensation model and justify the choices made for the compensation system in this work.

3.5 Summary and Conclusions

This Chapter has provided a survey of the most appropriate works from the literature to the developments reported in this thesis. The areas include modelling the unimpaired auditory system, modelling the impaired auditory system and modelling recruitment compensation systems.

In the area of modelling the auditory system, three primary categories of models exist macromechanical models, micromechanical models and hybrid models. Macromechanical models are subdivided into classical and non-classical models each employing either no approximations, the short wave approximation or the long wave approximation and can be 1, 2 or 3-dimensional depending on how the geometry of the fluid channels is represented. Micromechanical models are subdivided into passive and active models. Both the macromechanical and micromechanical categories have physiological bases for their development. Recent modelling efforts have formed a third category of auditory models called hybrid models which have no physiological background but depend on modelling the responses of the components of the auditory system. There are two main approaches to models of this type involving transmission line based models and filter bank based models. In general, auditory models are defined by their specification of the inner ear especially when it comes to examination of sensorineural impairment. A brief introduction is also given to models of the outer and middle ears.

In the area of modelling the sensorineurally impaired auditory system there are two main categories of models, those that model the impairment and those that model the effects of an impairment. In both categories two sensorineural impairments are examined, recruitment and frequency smearing. Considering models of the impairments, recruitment models typically utilise one of three methods, power law relationships, statistical modelling or suitable impairment of the responses of components of an auditory model. Frequency smearing has not seen as great an exploration as recruitment but can be modelled through manipulation of the auditory filter bandwidth in auditory models. These

types of impairment models are best suited to objective quantification of the effects of impairments. Considering models of the effects of the impairments, these are designed to provide a normal listener with the experiences of an impaired listener. The effects of recruitment can be modelled using dynamic level expansion, additive noise or centre clipping and the effects of frequency smearing can be best modelled using spectral smoothing in the frequency domain or by using additive bandpass noise. These types of models are best suited to subjective analysis of the effects of impairments.

In the area of modelling recruitment compensation systems there are two primary categories of approaches, linear amplification and amplitude compression. Linear amplification is well known and is typically coupled with AGC; more recent expansion of these types of models include frequency response shaping. Although not ideally suited to recruitment compensation in real world conditions, these methods do work quite well under controlled conditions. Better suited to real environments are compensation systems using compression in its various forms. Compression can be subdivided into two sections depending on how the mechanisms reduce the dynamic range of the acoustic signals. The first using amplitude compression consists of those methods involving wideband and multiband compression and typically do not perform a low level decomposition of the acoustic stimuli. The second section consists of those methods that use parametric compression where the acoustic stimuli undergo low-level decomposition into wavelets or sums of sinusoids. Manipulation of the characteristic parameters resulting from the decomposition allows the introduction of dynamic range compression.

The objective of this Chapter has been to familiarise the reader with the current methods and practices in the area of auditory modelling and to provide an insight into the background behind the developments in this work. This survey is by no means a complete historical review of all the developments in the area of acoustics and digital signal processing but is intended to highlight the most appropriate information for the work in this thesis.

Chapter 4 - Modelling the Human Auditory System

4.1 Introduction

The development of a model of the human auditory system requires careful consideration of the information available from experimentation and from other models of the system. In particular, any model being developed must in some form or other correlate with the physiology of the system or its responses. The previous Chapter has dealt with the work of other individuals in the area and is a fair representation of the current work taking place in the field. In particular, this information gives an indication of the directions that must be taken in the development of models that are more and more capable of emulating the results being reported by experimentation on this system. Ideally, if all the necessary information were available, development of an engineering model that perfectly emulates the physiological processes taking place in the auditory system would be relatively straight forward; however, clearly the auditory system is not fully understood and the development of models is therefore quite complex. In part, this is one of the driving forces behind the development of models so a better understanding can be attained of how this system functions.

Considering the information presented in Chapter 3, Section 3.2, it can be seen that three different classes of models exist, macromechanical, micromechanical and hybrid models, each having its own benefits and limitations. Given the problems associated with accurately ascertaining precise physiological data and the fact that the current work is primarily focused on the responses of the auditory system, in both its unimpaired and impaired forms, a hybrid approach to modelling the auditory system appears to be the optimum direction in which to proceed in developing a model of the auditory system, considering the objectives of this work.

Having decided on the theoretical approach to pursue, a general specification of the layout for the model is required. Given the currently available information in the field, there are many layouts and approaches that can be taken. Of great importance to this work is a model layout that is accurate, flexible and computationally efficient. In particular, computational efficiency is important because complete auditory models are complex and have many interactive components, which require significant computational power in order to operate. If the computational efficiency of the model is not carefully considered during

its development the hardware requirements to operate such models may escalate to a point where it becomes impossible/infeasible to operate them. Considering the information presented in Chapter 3 and many other works, the most suitable approaches are those that use modular elements to build up the models. In particular the work of Kates [49], [52], Patterson and Allerhand [55], Giguère [17], [42] and Meddis [113], [114] have been instrumental in the development of the models in this work. The general form of the complete model focuses on developing a modular time-domain model using a cascade of sub-models to suitably emulate the responses of the outer, middle and inner ears. The sub-model development specifically targets emulation of the complex processes taking place in the inner ear. The reason for specifically focusing on the inner ear emulation is that ultimately a model is required that is capable of modelling sensorineural defects, impairments that specifically corrupt the functions provided by the inner ear. The components making up each of the sub-models are designed to be as self-contained as possible to facilitate ease of replacement as new information becomes available and for experimentation with other models. The sub-models themselves have been developed to incorporate the latest experimental measurements of the auditory process using suitable altered information where data has been derived from species other than human. In this manner a model has been developed that accurately emulates the processes of the human auditory system.

The remainder of this Chapter specifically deals with the development of sub-models for the outer, middle and inner ears followed by a discussion of the complete model, its dependencies, initialisation, implementation and operation. The responses of the auditory model are then considered for various stimuli and compared to those measured experimentally. A comparison will also be carried out between the results generated by this model and those generated by other prominent models in the field. The operation of the model and its performance will then be discussed along with some conclusions on the complete auditory model.

4.2 Development of the Outer Ear Model

When an auditory stimulus occurs in the environment around us the sound pressure waves in the air first meet the pinna of the outer ear where they are focused into the external auditory meatus and impinge on the tympanic membrane. This path from the outside world provides amplification of particular frequencies within those signals prior to

meeting the tympanic membrane. The effect of this amplification is to make the auditory system more sensitive to particular frequencies; specifically it reduces the threshold at which we begin to hear these frequencies. Measurement of this threshold at which different frequencies are heard clearly demonstrates the effects of this amplification. The general trend of this threshold of audibility has been generated from averaged measurements of a group of otologically normal young adults and is shown in Figure 4.1 [115].

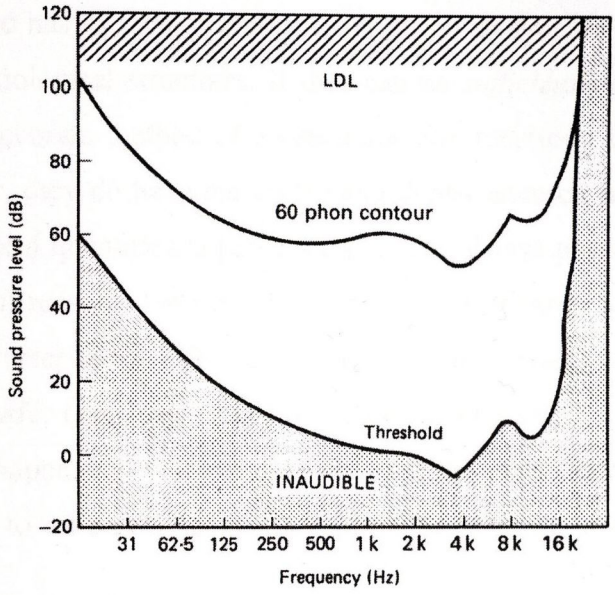


Figure 4.1. The human threshold of audibility curve.

The diagram shows that for very low and very high frequencies the auditory system is not particularly sensitive but for mid-range frequencies it is very responsive. This is important for humans as the significant frequency components of speech fall within this mid-range category. The outer and middle ears play an important role in the formation of the auditory threshold curve. The particular contribution of the outer ear (consisting of the pinna and external auditory meatus) in forming the auditory threshold curve are in the provision of enhancement of mid-frequencies through resonance effects, the maximum being approximately 10dB between 3-4kHz [115], [116].

When it comes to modelling the auditory function provided by the outer ear there are various ways in which this can be done. In particular, the outer ear has been modelled using analog networks in the form of uniform transmission lines [42], [56], electrical models [59] and transfer functions embodying the response characteristics of the biological structures [54]. These methods can be categorised into two groups. One group models the physiological structures and the other models the effects of those structures.

- The methods using transmission lines and electrical circuits generally try to model the physiological structures.

- The use of transfer functions aims at modelling the effects the structures have on signals passing through them and hence focuses on the *responses* of the physiological structures.

Each method has its own advantages and disadvantages. In favour of those models emulating the physiological structures, if they can be *sufficiently* parameterised they can provide a highly accurate method of representing the functions taking place within the outer ear. However, they do have the underlying dependence on being able to accurately measure physiological quantities, a process that is not always possible in live subjects for practical and ethical reasons. On the other hand, psychoacoustic tests can provide ample information to characterise the effects of these structures, results that can be embodied in the response of transfer functions. Modelling the effects the outer ear has on signals allows us model what is happening to the signals but *not* how these effects are being induced. Thus, this method to an extent ignores the actual physiology behind the functions it is performing.

In terms of developing a model of the complete auditory system, the representation of the outer ear function using transfer functions is far more appealing. The main reason for this is that it is far less numerically intensive to model the effects of a structure than the structure itself and also because many of the required parameters for physiological emulation of the outer ear can only be estimated. An important point to note is that we are more concerned with what is happening to the signals passing through the outer ear in signal processing terms than in how particular effects are caused. With this in mind the effects introduced by the outer ear are modelled using the transfer function $H_O(z)$ in Equation (4.1). This transfer function is based on the work carried out by Pflueger, Hoeldrich and Riedler [54] in which the magnitude response of the outer (and middle) ear filtering is developed from equal loudness contours.

$$H_O(z)=\frac{0.109z^7+0.109z^8}{0.2836-0.9879z+2.1396z^2-3.5548z^3+4.7251z^4-4.7251z^5+3.9295z^6-2.5359z^7+z^8} \tag{4.1}$$

It can be seen from the frequency response of this transfer function, in Figure 4.2, that the function provides the necessary gain over mid-range frequencies associated with the enhancement due to resonance that occurs in the outer ear.

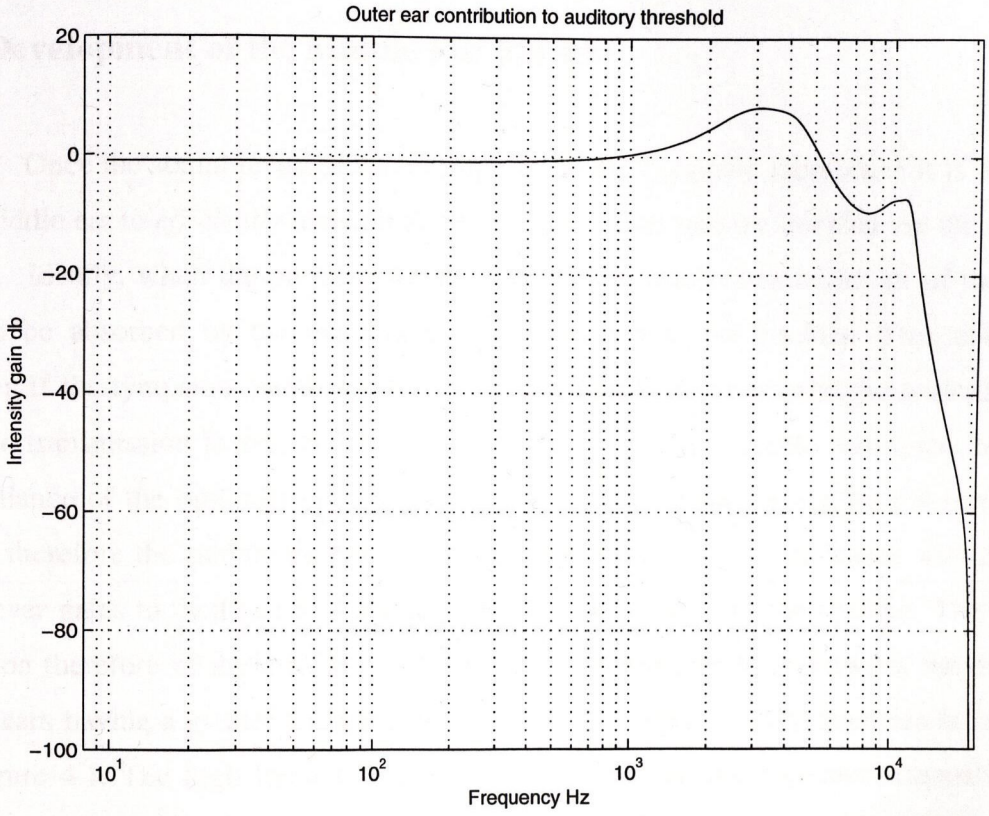


Figure 4.2 Frequency response of the transfer function $H_o(z)$.

It is worth noting that the outer ear on its *own* is not responsible for all the characteristics of the threshold curve. The middle ear plays a very important role as well. The high frequency attenuation of the transfer function $H_o(z)$ exhibits a response that results from the combination of the outer and middle ear responses. Strictly speaking the outer ear only provides damping for all higher-order modes below the ear-canal cut-off frequency resulting in plane wave modes of propagation of acoustic signals to the tympanic membrane [56]. It is the middle ear that is predominantly responsible for the high and low frequency attenuation and thus the elevated thresholds at high and low frequencies seen in Figure 4.1. A significant portion of the high frequency response should therefore be generated as part of the middle ear model. However, the specification of the outer ear model was created in conjunction with that of the middle ear. Taking this approach combines the high frequency response of the middle ear with the response of the outer ear and in the context of a complete auditory model is a valid arrangement. Realistically, the

combination of the outer and middle ear models should be considered together since the combination of the contribution of both the outer *and* middle ears form the general trend in the auditory threshold curve. This arrangement also proves very convenient in modelling terms. (See Chapter 4, Section 4.5.1).

4.3 Development of the Middle Ear Model

Once the sound pressure waves impinge on the tympanic membrane it is the job of the middle ear to *efficiently* transmit these pressure waves into the cochlea via the ossicular chain. Ideally, when the pressure waves reach the tympanic membrane all of the energy would be absorbed by the membrane and transferred to the cochlea. This could only happen if the tympanic membrane had the same impedance as air and the ossicular chain had no transmission losses, which is not the case [56]. The acoustic resistance, mass and compliance of the ossicular chain itself act to attenuate signals being passed through the chain therefore the middle ear must provide compensation for these losses with the areal and lever gains to facilitate efficient transmission of signals into the cochlea. The primary function therefore of the middle ear is one of impedance matching between the outer and inner ears having a greater tendency to efficiently pass mid-range frequencies as indicated in Figure 4.1. The high frequency content of signals experience greater attenuation than mid-range frequencies due to the acoustic mass, which is directly proportional to frequency. Similarly the low frequency content of signals experience greater attenuation than mid-range frequencies due to acoustic compliance, which is inversely proportional to frequency.

There have been other models of the middle ear that generally take the form of electroacoustic networks [42], [59], transmission line & electrical circuit models [56] and transfer functions modelling the middle ear responses [54], [52]. Using a similar argument to that of the outer ear, the use of transfer functions modelling the middle ear responses is used in this model. The current model incorporates two different models primarily emulating the low frequency properties of the middle ear because the high frequency properties are incorporated into the outer ear model. The combination of the outer and middle ear models together generate a response equivalent to the effects imposed by the outer and middle ears in the auditory system. Provided both the outer and either middle ear models are included in the complete auditory model, both the effects of the outer ear and the middle ear will be superimposed on signals passing through these models, as required.

The first model is based on work carried out by Pflueger, Hoeldrich and Riedler [54]. The transfer function suggested by this work is given by Equation (4.2) and provides only one degree of freedom in the specification of the transfer function.

$$H_{M1}(z) = \frac{z^2 - 2z + 1}{z^2 - 2Rz + R^2} \qquad (4.2)$$

R is a variable parameter use to characterise the transfer function. The transfer function in Equation (4.2) is cascaded with that of the outer ear model and the attenuation provided at low frequency is altered by varying the R-values. Figure 4.3 illustrates the frequency responses of the cascade of the transfer functions in Equations (4.1) and (4.2) for various values of R. Considering the characteristics of the threshold of audibility shown in Figure 4.1 and the equal loudness level contour measurements in [117], a value of R = 0.989 is used in the model to provide the best agreement with these experimental results. The R-value was arrived at by minimising the difference between the response of the combined transfer functions and the 0dBSPL equal loudness level contour (effectively the absolute threshold of audibility).

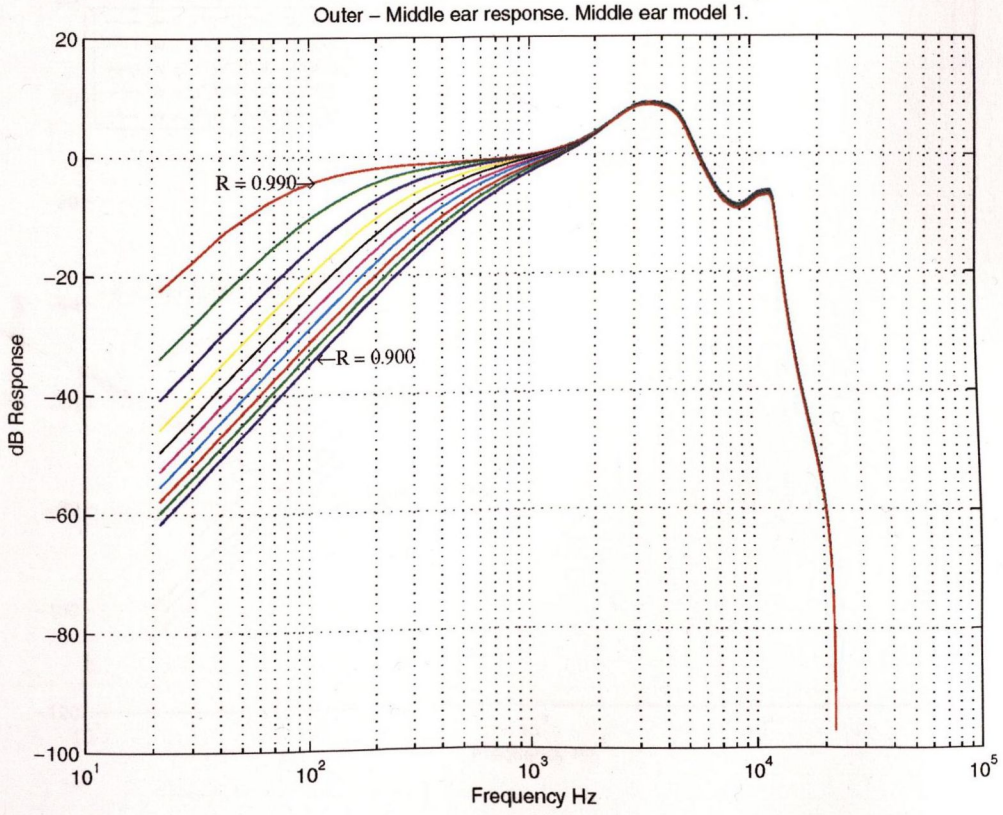


Figure 4.3. Cascade of outer-middle ear transfer functions for various R-values.

The second model is based on the work of Lynch, Nedzelnitsky and Peake [118] and documented by Kates [52]. The transfer function used is given in Equation (4.3) and provides two degrees of freedom for its characterisation.

$$H_{M2}(z) = \frac{z^2 - 2z + 1}{\left(1 + \omega/Q + \omega^2\right)z^2 + 2\left(\omega^2 - 1\right)z + \left(1 - \omega/Q + \omega^2\right)} \qquad (4.3)$$

Both ω & Q are variable parameters used to characterise the transfer function. The outer ear model is cascaded with the transfer function in Equation (4.3) and their combined frequency responses for various values of ω & Q are shown in Figure 4.4(a & b). Again considering the characteristics of the threshold of audibility shown in Figure 4.1 and the equal loudness level contour measurements in [117], setting $\omega = 2 * \pi * 4500/40000$ and $Q = 1.4000$ provides the best correlation with threshold data for mid-range frequencies. These values were arrived at by minimising the differences between the response of the combined transfer functions and the 0dBSPL equal loudness level contour.

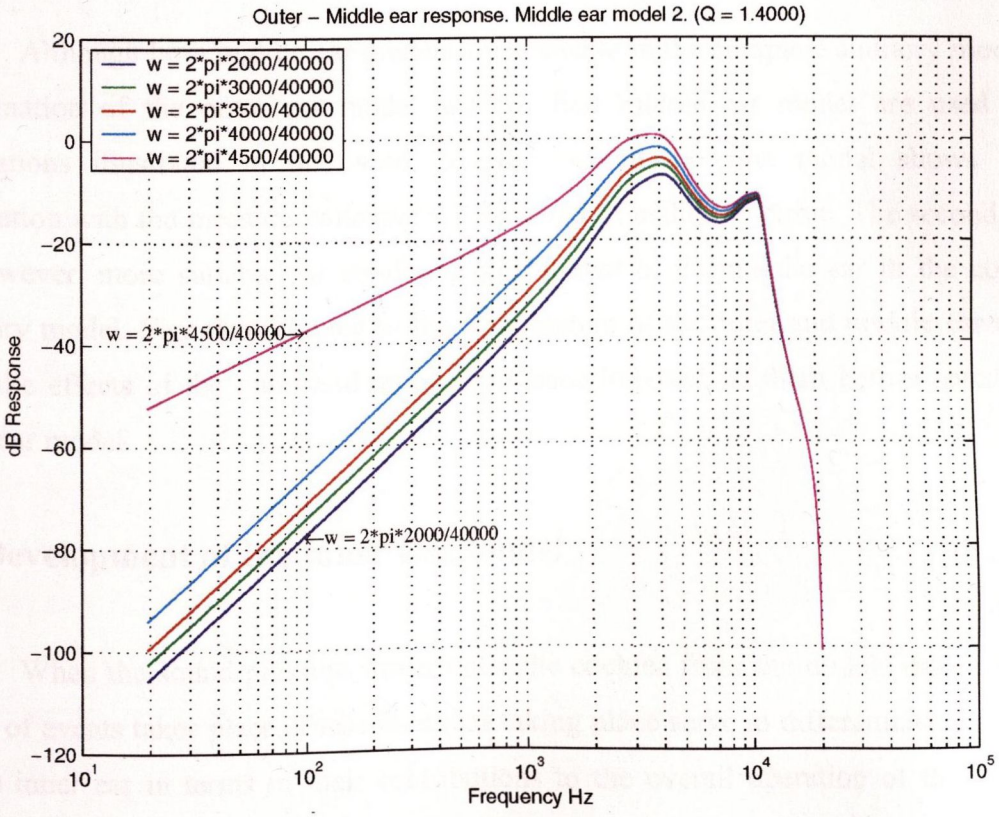


Figure 4.4a. Cascade of outer-middle ear transfer functions for various ω values.

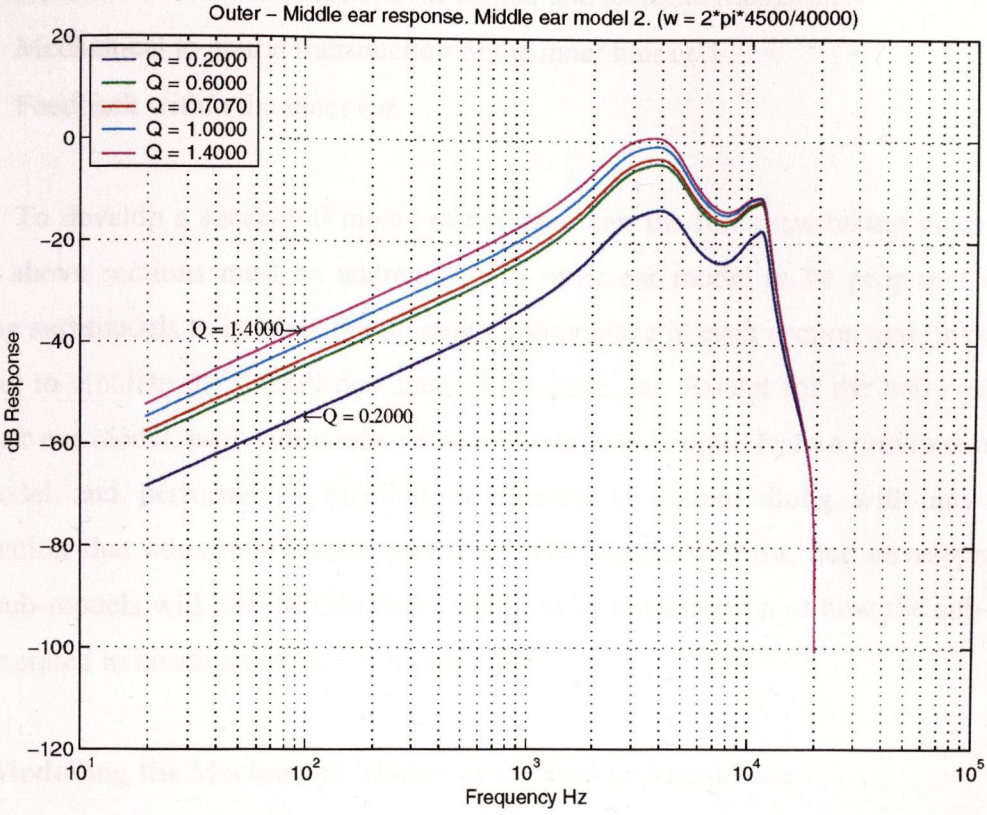


Figure 4.4b. Cascade of outer-middle ear transfer functions for various Q values.

Although both middle ear models are available in the complete auditory model, the combination of the outer ear model and the first middle ear model are used in the simulations illustrated in this work because overall the first model shows greater correlation with the measured *unimpaired* threshold of audibility curve. The second model is, however, more suitable for emulating impairment of the middle ear in the complete auditory model. Signals presented to the combination of the outer and middle ear models have the effects of the outer and middle ears superimposed on them before entering the inner ear model.

4.4 Development of the Inner Ear Model

When the sound pressure waves enter the cochlea from the middle ear, a complex series of events takes place. These functions taking place serve to differentiate the sections of the inner ear in terms of their contributions to the overall operation of the inner ear. These sections can be considered under the following headings:

- Mechanical motion of the basilar membrane.

- Resonance of motion between the basilar and tectorial membranes.
- Mechanical to neural transduction of the inner hair cells.
- Feedback within the inner ear.

To develop a successful model of the inner ear, the functions taking place in each of the above sections must be addressed. The inner ear model to be proposed involves creating sub-models to emulate the functions taking place in each section and linking them together to emulate the overall operation of the inner ear. Except for the entry point into the inner ear model, each sub-model takes as its input the output from a previous inner ear sub-model and performs its function to generate an output along with any control information that other sub-models may need. The justification for, and development of, these sub-models will now be described, followed by a description of how the sub-models are integrated to create an overall inner ear model.

4.4.1 Modelling the Mechanical Motion of the Basilar Membrane

Considering the mechanical motion of the basilar membrane, the travelling wave theory based on the work of Von Békèsy [14] and verified by Rhode [119] using the Mössbauer technique is the accepted form of motion for the basilar membrane. The encapsulation of this theory can be seen in the literature and can be described as follows. The stapes vibration within the oval window sets the cochlear fluid in motion inducing pressure differences across the basilar membrane. These pressure differences cause mechanical displacement of the basilar membrane in the form of travelling waves that propagate from the base to the apex of the cochlea. The basilar membrane widens as well as decreases in thickness along its length so that the amplitude of the wave begins to grow and, at the same time, the wavelength decreases as the wave propagates. However, because of damping, the wave reaches a point where it peaks and then quickly decays to zero. The location of this maximum amplitude depends on the frequency content of the wave. In this way, the spatial axis of the cochlea may be associated with a tonotopically ordered (i.e. frequency) axis [42], [53], [33]. The form of this mapping between basilar membrane place and frequency is non-linear and the response that occurs around a place of maximum displacement is highly non-linear although the response farther away takes a linear form. The non-linear response at the place of maximum displacement also has the effect of generating reflections which travel back up the cochlea and through the middle ear to generate otoacoustic emissions in the outer ear.

In the development of a model for this process these aspects of the motion need to be catered for; yet in many of the different approaches that have been taken this is not the case. This is largely due to the overwhelming complexity that arises. The majority of the approaches that have been taken loosely fall into one of the following categories; modelling the basilar membrane as

- 1,2 or 3-dimensional transmission lines. [26]
- A bank of parallel filters. [55]
- A hydro elastic system. [33]
- Wave digital filter representations of transmission lines. [42]
- Cascade filter sections taking advantage of pseudoresonance. [52]
- Non-linear models of frequency selectivity. [113]

On the whole these methods can be loosely categorised into those that use arrangements of filters and those that involve transmission lines. Each method has its own advantages and disadvantages. In the case of transmission line based models, they generally model the membrane itself as a transmission line and can provide accurate emulation of the motion of the basilar membrane once the necessary characteristic physical parameters are determined. However, the high sampling frequencies (100kHz) needed to get such models to function accurately require very large processing capabilities; also, many of the physical parameters cannot be directly measured. On the other hand, the filter-based approach looks at characterising the motion of the basilar membrane and not the membrane itself. Thus the overall structures used will have a response that emulates the travelling wave *pressure*. Such structures can accurately model the travelling wave motion but are more complicated to control and generally cannot directly model the reflected waves in the cochlea (a separate reflection path can be used to cater for reflections). They can however operate at much lower sampling frequencies and therefore are much less computationally demanding. In terms of being able to model reflections, the amount of reflected energy is minimal in a cochlea having no discontinuities in its physiological structure [49]; hence reflections will not significantly alter the forward travelling waves that are of primary concern.

In essence the difference in approaches is a subtle one. Where one approach models the membrane itself the other models the motion of the membrane. Considering a hybrid type approach has been selected for the overall model development, a filter-based method

would appear to best fulfil the model requirements, especially seeing as the responses of the auditory system are paramount in the current work. The decision on how to approach modelling the basilar membrane using a filter-based method must be considered in the light of some of the finer points of the travelling wave motion, namely, the physiological tuning of the basilar membrane, the mapping of frequency to place of maximum excitation on the basilar membrane, the stimulus level dependence of the tuning curves of the basilar membrane and basilar membrane compression. It is important to note that the initial model is not designed to cater for reflections. However, this topic is re-visited in the final Chapter.

4.4.1.1 Physiological tuning of the basilar membrane

In particular it is known that the motion of the basilar membrane is in the form of a travelling wave and the greater the displacement of the membrane the greater the stimulation of inner hair cells around the area of displacement. This allows us to deduce that there exists a mechanical tuning of the basilar membrane motion [120]; hence, frequency tuning curves can be derived for each point along the membrane based on physiological data and psychophysical masking data.

Clearly the basilar membrane is continuous and tuning curves exist for all points on the membrane; however, for digital simulation the membrane must be divided up into discrete sections. Each of these sections is located spatially on the basilar membrane and has a characteristic *centre frequency* at which it responds maximally and has an associated frequency tuning curve. These frequency tuning curves are mapped in the response curves of suitable transfer functions or cochlear filters as they are commonly described in the literature. Ideally the response of the cochlear filters should exactly replicate the tuning curves of the discrete points on the basilar membrane. The tuning curves themselves can be described as being relatively broadly tuned, responding highly non-linearly around their centre frequency and responding linearly further away from their centre frequency. They have a much steeper roll off on their high frequency side than on their low frequency side and in general their bandwidths are linearly proportional to their centre frequencies above 800Hz. For lower centre frequencies the tuning curves become gradually less tuned. Within a filter-based approach, the cochlear filters must reproduce these *frequency tuning curve characteristics* in their frequency response. The relatively constant tuning over most

of the cochlear length implies that the centre frequencies of the curves must be *logarithmically* mapped to basilar membrane location [53].

4.4.1.2 Mapping of frequency to place of maximum excitation on the basilar membrane

This logarithmic mapping of frequency to place of maximum displacement is vitally important in the development of an accurate model. The direct measurement of these functions for ethical reasons cannot be carried out on living humans. However, experimentation with animals and fresh human cadavers has provided sufficient information for the development of such functions. In the literature there has been significant work carried out in this area and in particular the work of Liberman [121] and Greenwood [122] has provided successful functions that can be applied to the human auditory system. These functions allow us to accurately map basilar membrane locations with their frequencies of maximal displacement. The basilar membrane model must use such mapping functions to make the transformation between the spatial and spectral domains for the cochlear filters.

4.4.1.3 Stimulus level dependence of the tuning curves of the basilar membrane

The mechanical tuning of the basilar membrane provides the ability to distinguish, with high resolution, the frequency content of signals propagating along the basilar membrane. Considering the dynamic range of hearing going from 0dBSPL (0.00002Pa) to 120dBSPL (20Pa), there are many different stimuli that can be presented to the auditory system that fall both inside and outside this range. Stimuli that fall below this range are inconsequential as we cannot hear them but stimuli in the upper sections of this range and above can be quite harmful. As the stimulus level entering the cochlea rises the basilar membrane displacement increases in unison and the greater the displacement of the basilar membrane the more inner hair cells are activated thus reducing the frequency selectivity of the auditory system. This means that as the stimulus level rises the tuning curves broaden with the added effect of causing the location of maximum displacement to increase, resulting in a reduction in the centre frequency of the tuning curve. The filter representation of the tuning curves must be able to take these phenomena into account. Extremely loud sounds can damage the inner hair cells and even cause the basilar membrane to tear. The effects of basilar membrane tearing and the extreme stimuli levels

required to cause such effects are not considered in this work, in general only stimuli that fall within the normal dynamic range of hearing and their effects are considered.

4.4.1.4 Basilar membrane compression

It is known that the basilar membrane has a non-linear frequency response from the measurements of Rhode [119]. These experiments show that the amplitude of basilar membrane vibrations grows non-linearly with increasing stimulus level and can be quantified by calculating the ratio of the basilar membrane vibration to malleus vibration. If the mechanics of the cochlea were linear, this ratio would be level independent but it is not. This type of relationship means that the unimpaired cochlea fundamentally embodies a *compressive non-linearity* and is the reason why the inner hair cell tuning curves are not identical to either the isodisplacement or the isovelocity tuning curves for the basilar membrane [20]. It can be calculated that the inner hair cells at best can only physically code a range of approximately 63dB, the practical range is actually much less. This means that on their own they cannot code the 120dB dynamic range of acoustical signals the ear can respond to and that the non-linear motion of the basilar membrane must be providing compression within the mechanics of the cochlea. The outer hair cells are the source of the basilar membrane non-linearity but it is unclear what precisely controls the non-linearity. It is known that the non-linearity is related to the outer hair cell length changes, controlled by the outer hair cell membrane voltage. This voltage is determined by shearing displacement of the hair cell cilia by the tectorial membrane and the most likely cause of non-linear basilar membrane mechanics is due to changes in the micromechanical impedances within the organ of Corti which result from outer hair cell length changes [61]. This compressive nature of the basilar membrane due to outer hair cell intervention has to be incorporated into the cochlear filters; hence they must be capable of receiving a feedback control signal from the outer hair cell representation within the inner ear model.

4.4.1.5 Structural model of the basilar membrane

Having considered the intricate details of the travelling wave motion with a view towards modelling it using a filter-based approach, strategies employing banks of parallel band pass filters [54], [55] and using a cascade of filters employing pseudoresonance [51], [123] were evaluated. The method using a cascade of filters employing pseudoresonance was finally selected because individual parallel filters were unable to efficiently reproduce

the fine-tuning found at higher frequencies in the auditory system. The essence of pseudoresonance can be best explained in terms of the cascade itself.

The cascade consists of a number of section filters, $H_i(z)$, that together model the basilar membrane motion. The frequency tuning curve and hence the cochlear filter response at any point in the cascade depends on all the section filters up to that point. It is this build up of the contributions of all the sections up to a particular point that is encapsulated in the term pseudoresonance.

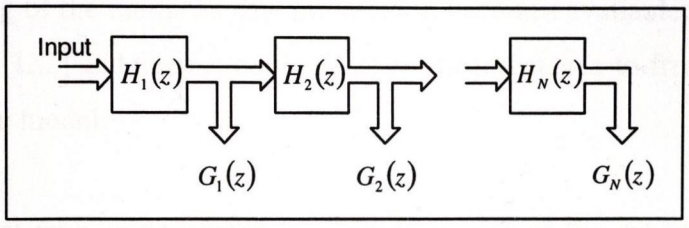


Figure 4.5. Diagrammatic representation of a cascade.

In signal processing terms this can be reduced to time domain convolution of the section filter impulse responses up to the point in question or more simply the z-domain product of the section filter transfer functions. Figure 4.5 shows a diagrammatic representation of the cascade approach as explained in [52]. When considering a signal processing expression for the frequency tuning curve at any point, *N*, in the cascade we are referring to the absolute frequency response curve of the resultant z-domain transfer function $G_N(z)$ given by.

$$G_N(z)=\prod_{i=1}^{N} H_i(z) \qquad (4.4)$$

$H_i(z)$ is the z-domain transfer function of a section filter in the cascade. It is important to differentiate between section filters and cochlear filters. The section filters, $H_i(z)$, locally characterise a section of the basilar membrane but the overall frequency tuning curve for a particular point on the basilar membrane depends on the cochlear filter, $G_N(z)$, which is the cascade of all the section filters up to that point.

4.4.1.5 Spectral alignment of cochlear filters on a spatial basilar membrane

Using this cascade approach identifies the structure being used to model the basilar membrane motion but not how or where in frequency the section filters are placed. In this work the range of frequencies responded to by the human auditory system is specified to lie between 50Hz and 20kHz; the effective/useful range is, however, much less. This means that the complete length of the basilar membrane must respond to frequencies within this range, based on some non-linear translation function. At the outset, the model was designed to be modular so different functions could be easily tested and also to facilitate upgrading of the model as new information becomes available. With this in mind both Lieberman's [121] and Greenwood's [122] cochlear distance-to-frequency maps were implemented in the model.

The original map as developed by Lieberman was of the cat frequency mapping. Although the cat and human basilar membrane are similar, the cat's frequency range is specified from approximately 96Hz to 60kHz, far above that of any human. To cater for this, the mapping has been scaled by a factor of 3 to cover the range 32Hz to 20kHz. The resulting mapping function based on normalised cochlear distance x is given by Equation (4.5), [52]. A plot of this function can be seen in Figure 4.6

$$f(x)=160\left(10^{2.1x}-0.8\right) \qquad (4.5)$$

The Lieberman function provides centre frequencies for points with normalised distance x from the apical end of the basilar membrane, i.e. the low frequency end.

The Greenwood map is a more generalised function but is designed to provide cross species maps. The function essentially requires only empirical adjustment of a single parameter to set the upper frequency limit, while a slope parameter can be left constant if cochlear partition length is normalised to unity or scaled if distance is specified in physical units. The resulting human mapping function based on a 0.035m long basilar membrane is given by Equation (4.6), [124]. A plot of this function can be seen in Figure 4.7.

$$f(x)=f_0e^{-kx}-f_1 \qquad (4.6)$$

where $k = 4.837$, $f_0 = 20800Hz$ and $f_1 = 145.5Hz$. The Greenwood function being used provides centre frequencies for points with normalised distance x from the basal end of the basilar membrane, i.e. the high frequency end. The frequency range provided by this function is approximately 20Hz to 20kHz.

Linearly dividing the basilar membrane along its length and using either of these functions can provide the frequency map needed to locate the cochlear filters spectrally within the cascade. All that remains is the specification of the section filters themselves.

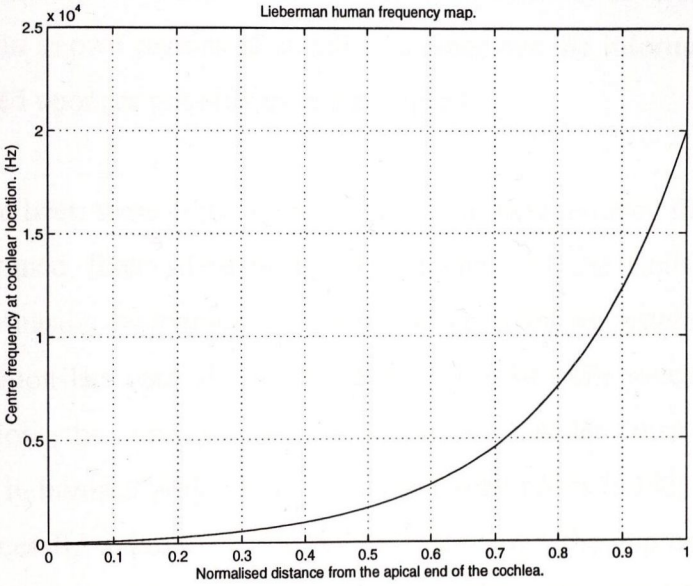


Figure 4.6 Graph of the altered Lieberman frequency map.

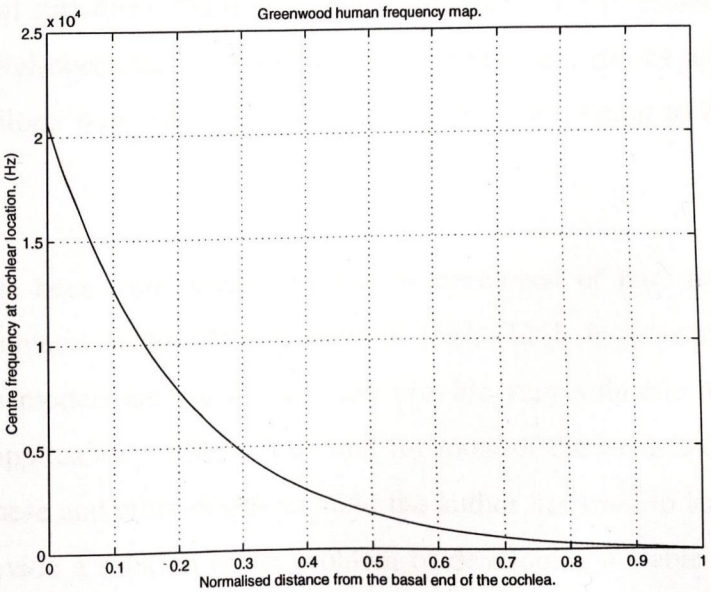


Figure 4.7 Graph of the Greenwood frequency map.

4.4.1.6 Section filter specification

The specification of the section/cochlear filters in this model, as in all filter-based approaches, is vital to the successful emulation of the basilar membrane motion. Many of the models that use filter-based approaches tend to have problems in their representation of the cochlear filters. These models try to directly equate the basilar membrane filter shapes with tuning curves resulting in sharp resonances in individual sections. The problem with these sharp resonances is that they cause instabilities in computational models and analog circuit representations of active basilar membrane feedback systems. Models with problems such as these have to have their domain of applicability severely restricted so that they operate within known regions of stability as otherwise the information they produce cannot be depended upon, or possibly even reproduced.

There have been some filter representations that have avoided these problems. The use of broadly tuned filters developing the sharpness of the tuning curves through interaction of the basilar membrane excitation and an adaptive system gain control in a cascade configuration has been shown by Kates [52] to be quite successful, although on closer examination the tuning curves are much broader than those measured psychophysically in humans [49]. The use of wave digital filters in [42] and [17] is a good example of a successful approach to the problem. However, the approach is underpinned by a transmission line model resulting in high sampling rates, models containing large numbers of wave elements and the use of non-real wave quantities to avoid the problem of unrealisable digital structures. These models similarly show sharp resonances at low input levels and the level-dependent gain of the basilar membrane curves arises from locally resonant effects alone thus ignoring the outer hair cell contribution to basilar membrane function.

There have been some successful models developed of late that have addressed some of these problems using different methods [55], [124]. In general though once the limitations of the models are known they can provide very valuable information within their domain of applicability. This can be said for most of the models documented in the literature. With these and other details in mind the author has tried to learn from the work of others and provide a solution to the problem of developing a stable basilar membrane model incorporating

- Computational efficiency.
- Close adherence to psychophysical tuning curves measured in humans both spatially & spectrally.
- Cochlear filters that are stimulus level dependent.
- A realistic compressive non-linearity.

When considering the specification of a particular section filter it must be considered in the context of the cascade as a whole since each section filter is only a link in the chain. It is the combined effect of all preceding sections up to a particular section that defines the overall cochlear filter response for a specific point on the basilar membrane. The only non-dependent section in the cascade is the first section, as it alone must uniquely define the cochlear filter response for the first spatial point on the basilar membrane. The specification of the centre frequency for the first cochlear filter from one of the location-to-frequency mapping functions will be the centre frequency for the first section filter. Sections further down the cascade will *not* have this property. In general, the section filters generating the frequency response curve at a point do not have centre frequencies corresponding to the centre frequency of the cochlear filter response they are generating. This implies that solution of the section filter coefficients reduces to an operation of finding suitable centre frequencies for the section filters so their combined effect generates the required cochlear filter responses at the centre frequencies specified by one of the mapping functions.

This requires the specification of a general section filter having a suitable response that can be spectrally shifted to have any required centre frequency. Specification of the first section filter will be the easiest to perform yet it is the most crucial as all other section filters will be based upon it.

Many different types of transfer function have been explored based on information in the literature. However, their limitations, some of which have been described earlier, are all too apparent. In particular, current models are largely based on linear or quasilinear filters, but the auditory system embodies a non-linearity around the point of maximal displacement leading to sharp resonances using these methods. To avoid these problems the development of non-linear sections was decided upon, as these could potentially emulate the psychophysically measured responses of the human auditory system more accurately and could avoid the problem of sharp resonances. Ideally, cochlear filters were

required that could provide a response the same as that of a transmission line, based on the cochlear wave equation but avoiding the problems associated with using transmission line representations.

A novel approach based on the theories of Kates [52], [49], Peterson and Bogert[46], Zwislocki [125], Schroeder [50], Eysholdt and Mellert [126], E. Ambikairajah, N.D. Black, R.Linggard [127], Allen and Fahey [65], Giguère and Woodland [42], [17] and Patterson and Allerhand [55] was developed to solve this problem. Rather than trying to model the basilar membrane response as a complete transmission line, each section filter is made emulate an *isolated section* of a 1-dimensional transmission line loaded by the impedance of the adjacent section. The section filters model the response of isolated 1-dimensional transmission line segments avoiding the problems associated with complete transmission line representation. A suitable combination of these sections together, as described earlier, forms the cochlear filters, the response of which models the entire transmission line response.

The work in [49] takes a similar approach but is limited. The sections used do not allow for the outer hair cell interaction with the basilar membrane, affecting the dynamic compressive nature of the system. Also, deriving the section filter control signal *exclusively* from the motion of the cochlear partition prevents adaptation of the *cochlear* response to suitably match the psychophysically measured tuning curves in the human auditory system; in particular, it adversely affects the non-linearity of the cochlear filter response at the centre frequency and its linearity at remote frequencies. The importance of these features has only recently been confirmed by Meddis [113].

Making the parameters in the section filters that characterise the transmission line sections doubly dependent dynamic variables can solve these problems. Firstly, making the section filter coefficients dependent on the peak region stimulus level links the section filter response to the stimulus level and, secondly, making the local characteristic impedance of the section filter dependent on the outer hair cell function makes the link with the overall micromechanical impedance required by the cochlear filters. Such characterisation of the section filters very neatly conforms to the current theory of basilar membrane excitation and compression, namely that the excitation and compression are dependent on both the stimulus level and outer hair cell activity and that they are active mechanisms.

Figure 4.8 depicts an equivalent analog circuit of a section of a 1-dimensional transmission line. The transfer function of each section filter represents one of these transmission line sections. Each section is loaded with the input impedance of the adjacent section in accordance with Eysholdt and Mellert [126] and E. Ambikairajah, N.D. Black, R.Linggard [127]. Their work proves that replacing the input impedance of the adjacent section with an approximate characteristic impedance has a minimal effect on the transfer function from the base to the apex of the basilar membrane. They also derived a transfer function for such a section consisting of a pair of high-Q poles and zeros along with a pair of low-Q poles and zeros to accurately emulate a segment's response.

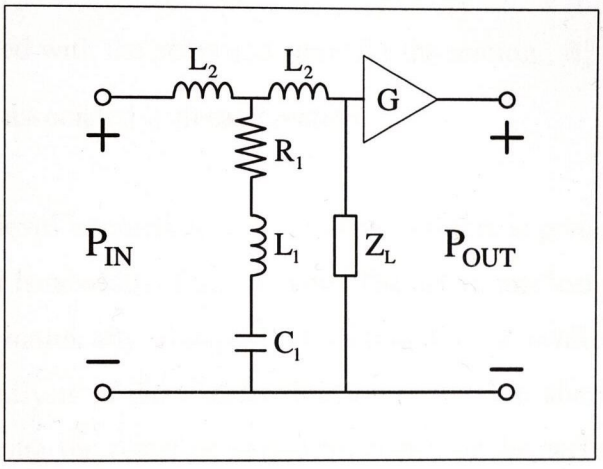


Figure 4.8 Isolated 1-dimesnional transmission line segment.

When considering such a transfer function in terms of computational efficiency, it can be reduced somewhat based on the observation that around the characteristic frequency the effect of the high Q poles and zeros will dominate the effects of the low Q poles and zeros. Therefore the transfer function can be simplified to a pair of high Q poles and zeros [49]. Combining these ideas with the modifications above defines section filters that can

1. *locally*, on their own, emulate a dynamic isolated 1- dimensional transmission line segment loaded with a dynamic characteristic impedance.

2. *globally*, in cascade, emulate a complete dynamic transmission line with micromechanical impedances just as occurs in the basilar membrane with the necessary associated dynamic cochlear filter frequency response curves.

This provides the key ingredients needed to develop section filter transfer functions so that when suitably parameterised and cascaded, they provide cochlear filters that will

fulfil the requirements as documented. In terms of the transfer function being used it takes the general form as shown in Equation (4.7).

$$H(z) = G \times \frac{\left(b_o^2 + \frac{b_o}{Q_o} + 1\right) + 2\left(b_o^2 - 1\right)z^{-1} + \left(b_o^2 - \frac{b_o}{Q_o} + 1\right)z^{-2}}{\left(b_p^2 + \frac{b_p}{Q_p} + 1\right) + 2\left(b_p^2 - 1\right)z^{-1} + \left(b_p^2 - \frac{b_p}{Q_p} + 1\right)z^{-2}} \qquad (4.7)$$

where b_o & b_p are the frequencies of the zeros and poles of the section and Q_o & Q_p are the Q factors associated with the poles and zeros for the section., $b_o = 1.15b_p$, $Q_o = 1.5Q_p$ and G is a gain factor associated with each section.

The stimulus level interaction with the section filters is predominantly involved in adjusting the gain and bandwidth of the sections. The active mechanism involves adjusting the pole Q which dynamically changes the section filter's centre frequency gain and bandwidth. Simple analysis of the transfer function shows that altering the Q factors also has the effect of altering the resulting centre frequency of the section. When the section filters are in cascade and as the stimulus level rises the gain of the sections around the place of maximum excitation is reduced, the filter bandwidths increase and their centre frequencies reduce resulting in cochlear filters that follow closely with experimental results.

The outer hair cell action when completely functioning provides compression across the cascade of approximately 2.5:1 in agreement with experimental measurement [119]. This ensures both a linear growth in perceived loudness relative to growth in input stimulus level and a non-linear compressed increase in basilar membrane displacement with increase in input stimulus level. The compression has the effect of compressing the absolute auditory range of approximately 0dBSPL to 120dBSPL into a range of approximately 0dBSPL to 48dBSPL. The outer hair cells invoke their action by altering the range limits of the Q factors of the section filters. When the outer hair cells are fully functional the Q factors have access to their complete range. At this point we are only concerned with developing a model of the fully functional auditory system; the effects of outer hair cell malfunction will be dealt with in Chapter 5.

When the filters are initially not being stimulated they are in their maximum response state. The Q_p values must be specified for this state before solutions for the characteristic values required by the filter sections can be generated. The function in Equation (4.8) is used to generate the Q_p values for the section filters.

$$Q_p = Q_{\text{Ref}} + Q_x * X \qquad (4.8)$$

where $Q_{\text{Ref}} = 2$ is a reference value for the Q factor and Q_x is the increase in Q required when you progress a normalised distance X down the cochlear partition. The cochlear partition is linearly divided up in the spatial domain just as for the frequency mapping functions. Using the above equation, a Q_p value can be assigned to each section depending on how far down the partition it is.

All that remains for specification of the section filters is a method for calculating the section filter coefficients. This essentially involves solving for the frequency of the poles of the section filter transfer functions. The easiest section to solve for will be the first section.

An approach similar to that suggested in [52] is used to solve for the pole frequencies. When the transfer function of a section filter is examined, the centre frequency of such a section is defined as the frequency of maximum response within the overall response curve. To calculate the centre frequency of the response curve of the first section filter you have to numerically solve for the appropriate root/frequency of the 1st order frequency derivative of the squared discrete time Fourier transform of its section filter transfer function. This can be mathematically expressed as the roots of Equation (4.9) with the pole frequency of the section set to some initial value.

$$\frac{\partial}{\partial \omega}\left(\left(H_1(z)\big|_{z=e^{j\omega}}\right)^2\right) = 0 \qquad (4.9)$$

The process of numerically solving for the pole frequency involves convergence of the initial value to a value that produces a root at the desired cochlear filter centre frequency. Based on the form of the transfer function, two roots can be found, one a local maximum and the other a local minimum. The root representing the centre frequency will

be the one that is lower in frequency given the form of the high frequency roll-off of the transfer function. Setting the pole frequency based on the solution to this equation will locate the resulting section filter so that it will have the required centre frequency. Obviously the section filter and cochlear filter will have the same centre frequency for this case of an *isolated* section. However, when we consider the first section filter it is not entirely valid to do so in isolation. The reason for this is that the resulting centre frequency of the cochlear filter frequency response will be shifted by transfer functions in series with the inner ear model, namely the middle and outer ear transfer functions. These other functions must be taken into account when calculating the section filter's pole frequencies so the resulting cochlear filters have the desired centre frequencies as calculated by the mapping functions. An approach is needed that caters for transfer functions external to the cascade.

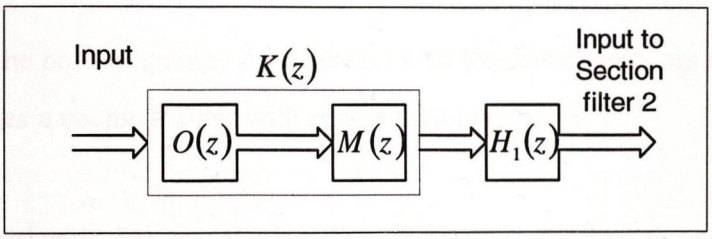


Figure 4.9 Outer, Middle and Inner ear transfer function cascade.

Figure 4.9 depicts the z-domain transfer functions representing the outer ear model, $O(z)$, the middle ear model, $M(z)$ and the first section of the inner ear cascade model, $H_1(z)$. Combining the outer and middle ear transfer functions into $K(z)$ effectively leaves two transfer functions in cascade representing the outer ear, the middle ear and part of the inner ear, with $K(z)$ being *fully* parameterised. The only unknowns reside in $H_1(z)$. Trying to directly apply the method of derivatives to the product of these transfer functions as before becomes too unwieldy. Instead of considering products of transfer functions it is more beneficial to consider sums by taking the natural logarithm of the cascade of $K(z)$ and $H_1(z)$ using the property of logarithms that $\ln(xy)=\ln(x)+\ln(y)$. Most importantly this procedure does *not* alter the location of the roots of the transfer function resulting from the cascade of $K(z)$ and $H_1(z)$. Thus the problem reduces to the *special* solution of.

$$\frac{\partial}{\partial\omega}\left(\ln\left(\left(K(z)\big|_{z=e^{j\omega}}\right)^2\right)\right)\Bigg|_{\omega=\omega_c}+\frac{\partial}{\partial\omega}\left(\ln\left(\left(H_1(z)\big|_{z=e^{j\omega}}\right)^2\right)\right)=0 \qquad (4.10)$$

A *special* solution of Equation (4.10) is required in that the derivative associated with $K(z)$ is evaluated at a value of ω equal to the desired cochlear filter centre frequency, ω_c . This value is added to the derivative associated with $H_1(z)$ and numerical solution of the resulting equation converges on a pole frequency so that the lowest root has a value equal to the first cochlear filter's centre frequency (ω_c) as specified by the place to frequency mapping function. This pole frequency fully characterises the first section filter, as it will specify the spectral location required for the poles of the section filter. All other parameters for the section filter are derived from this value to produce the required tuning curve response. This method also provides an elegant approach for solving the pole frequencies of the other section filters in the cascade. The solution can be generalised as follows.

To solve the pole frequency for a section k so that the cascade up to (and including) that point generates a cochlear filter with centre frequency ω_{ck} :

- Evaluate Equation (4.11) to calculate an offset value due to the presence of transfer functions prior to the section of interest.

$$\Phi_{\omega_{ck}} = \frac{\partial}{\partial\omega}\left(\ln\left(\left(K(z)|_{z=e^{j\omega}}\right)^2\right)\right)\Bigg|_{\omega=\omega_{ck}} + \sum_{i=1}^{k-1}\frac{\partial}{\partial\omega}\left(\ln\left(\left(H_i(z)|_{z=e^{j\omega}}\right)^2\right)\right)\Bigg|_{\omega=\omega_{ck}} \qquad (4.11)$$

Note: The pole frequencies for sections up to the unsolved section will already have been calculated and thus they will be fully characterised.

- Numerically solve Equation (4.12) so its lowest root is at the k^{th} cochlear filter frequency from the mapping functions.

$$\Phi_{\omega_c} + \frac{\partial}{\partial\omega}\left(\ln\left(\left(H_k(z)|_{z=e^{j\omega}}\right)^2\right)\right) = 0 \qquad (4.12)$$

The frequency used in the k^{th} section filter to locate the root of Equation (4.12) at the k^{th} cochlear filter frequency, from the mapping functions, will be the pole frequency for the k^{th} section filter.

Note: The method of bisection followed by the secant method for root solution is used to converge on the appropriate pole frequency.

Using the above method it is possible to generate the coefficients for all the section filters in the cascade. The current model uses a total of 120 section filters. The input to the inner ear is passed to the first section filter and on down through the cascade. The output from the section filters is the emulated basilar membrane travelling wave pressure. Each section filter also takes inputs from a feedback mechanism to control the stimulus level dependence of the sections and the compression provided by the outer hair cells. This mechanism is described later in Section 4.4.4.

4.4.2 Modelling the Resonance between the Basilar and Tectorial Membranes

The sound pressure waves moving down the cochlea not only stimulate motion of the basilar membrane but also cause motion of the tectorial membrane. Relative motion between the basilar and tectorial membranes causes stimulation of the stereocilia protruding up into the tectorial membrane. This gives rise to a resonance of motion between the two membranes which introduces certain spectral phenomena into the frequency response of the cochlea. In particular, one octave below the frequency of maximum excitation there is a downward notch in the response [128] and a resulting phase shift of π radians [129], [130]. The work of Allen [120], [20] and Kates [52] has lead to the development of a second filter for the section filter outputs to introduce these spectral effects into the basilar membrane response prior to stimulation of the models for the hair cells. The form of the second filter used here is similar to that of Allen and Kates, providing a notch in frequency one octave below the section centre frequency, a phase change of π and gain at the centre frequency. The extra gain at the centre frequency arises from observations in [128] that the gain around the centre frequency increases with frequency.

In general, the poles of the transfer function are located near the section centre frequency and the frequency of the zeros is located 1 octave below that of the poles. The sharpness of the resonances is adjusted by setting the Q for the zeros to be twice as large as the Q for the poles. The general transfer function for the second filter is given by Equation (4.13) in terms of the pole frequency and pole Q.

$$S_k(z)=\frac{\left(\omega_p{}^2Q_p+\omega_p+4Q_p\right)+2Q_p\left(\omega_p{}^2-4\right)z^{-1}+\left(\omega_p{}^2Q_p-\omega_p+4Q_p\right)z^{-2}}{4\left(\omega_p{}^2Q_p+\omega_p+Q_p\right)+8Q_p\left(\omega_p{}^2-1\right)z^{-1}+4\left(\omega_p{}^2Q_p-\omega_p+Q_p\right)z^{-2}} \quad (4.13)$$

To match the observed gain at centre frequency as seen in [128] the maximum Q factor for the poles is given by the frequency dependent equation.

$$Q_p = 2(1+2f) \quad (4.14)$$

where f is frequency given in kHz. Equation (4.14) and hence (4.13) are significantly different to those proposed by Allen and by Kates because this second filter has to match up with the new section filters proposed in this model but the principle behind the use of this transfer function is still the same.

In terms of the characterisation of the second filter, it only needs the specification of the pole frequency to uniquely specify an instance of this transfer function. To ensure the peak response of the second filter is at the section centre frequency, the pole frequency is numerically solved for using an approach similar to that used to solve for the section filter coefficients. This involves numerically solving for the pole frequency so the *upper root* of the first order frequency derivative of the squared discrete time Fourier transform of the second filter is at the section centre frequency. That is, a value for (ω_p) must be found that makes the upper root of Equation (4.15) lie at the section centre frequency.

$$\frac{\partial}{\partial\omega}\left(\left(S_k(z)\big|_{z=e^{j\omega}}\right)^2\right)\Big|_{\omega=\omega_{ck}} = 0 \quad (4.15)$$

The reason the upper root is sought is that the lower root isolates the local minimum of the response whereas the upper root isolates the local maximum of the response; this can be seen when considering the overall response of such a transfer function. In general, each section in the cascade requires a second filter located at its section centre frequency with the output from each section filter being the input for the section's second filter. The output of the second filters will be the required basilar membrane pressure response incorporating the spectral effects identified by experimentation.

4.4.3 Modelling the Mechanical-to-Neural Transduction of the Inner Hair Cells

The mechanical motion of the basilar and tectorial membranes stimulates the inner hair cells resulting in the firing of nerve impulses along the neurons synapsing with each hair cell. The complete biological process as described earlier performs the function of converting the mechanical motion within the cochlea into electrical signals going to the brain. This process is called *mechanical-to-neural transduction.* The inner hair cells performing this operation reside along the length of the basilar membrane so that each section of the cascade requires an inner hair cell model to react to the local stimulation of that section. Each hair cell model needs to take as its input the local basilar membrane pressure and produce as its output the local neural firing pattern associated with its location in the cascade. The combination of all the hair cell model outputs, from all the sections in the cascade, generates a coded neural map of the cochlear response to the sound pressure waves propagating down the cochlea. This cochlear response encodes the input stimulus both spectrally and temporally. This map can be considered a bio-electrical representation of the stimulus entering the inner ear.

Development of a model for transduction requires consideration of the nature of the processes taking place. The motion of the cilia alters the conductance of a biological membrane at the surface of the cilia thus modulating the ionic current flowing into the inner hair cell. When the cell becomes sufficiently charged it causes the release of synaptic transmitters which migrate into the cleft of the cell causing the associated afferent nerve to fire. When the nerve fires, electrical impulses (spikes) travel along the nerve to the brain. The electrical potential of these spikes does not alter but the rate at which these spikes are generated is stimulus level dependent. Even when the ear is not being acoustically stimulated the acoustic nerves fire *spontaneously.* The rate at which they fire depends on the type of fibre. There are three types of fibres with differing spontaneous rates, (SR) [131]:

- Low Spontaneous Rate $SR \leq 0.5$ spikes/s
- Medium Spontaneous Rate $0.5 < SR < 18$ spikes/s
- High Spontaneous Rate $SR \geq 18$ spikes/s

Generally, high-SR and medium-SR fibres tend to threshold near auditory threshold and exhibit clear saturation behaviour. In contrast low-SR fibres have thresholds ranging

up to 40dB above auditory threshold. However, their firing rates grow in response to stimulus levels beyond their nominal saturation level. Typically during strong steady-state acoustic stimulation the firing rates can reach an average *saturation* level of around 150spikes/s. When a strong stimulus is presented to the auditory system the firing rate can reach *peak instantaneous* levels of up to 1000spikes/s. Rates much higher than this are not possible due to the refractory period of a nerve after each firing. This refractory period arises because it takes a finite amount of time for the neural transmitter chemicals to build up to the level at which a neural impulse can be fired. Any stimulation of the inner hair cell causes an initial high response that decays to an *adapted* firing rate. The literature records an initial *rapid* adaptation time constant of around 2.5ms followed by a *short-term* adaptation having a time constant of around 50ms as found from experimentation, [52], [132].

It was decided to make the inner hair cell sub-model as self-contained as possible to allow its replacement with other models as they became available and also to allow the use of multiple inner hair cell models. In particular this work focuses on the efforts of Meddis [114], [133], Westerman [134], Davis [12], Allen [11] and Kates [52]. As it stands, the model incorporates two different models of the transduction process.

4.4.3.1 Inner hair cell model 1.

The first hair cell model, based on the work of Meddis and of Westerman, emulates the production, movement and dissipation of transmitter substance in the region of the hair cell auditory-nerve fibre synapse to generate simulated spike trains in response to a stimulus. The original model is an isolated model of the transduction process and does not therefore take into account the rest of the auditory system. It was also designed specifically to take auditory input, not the modelled basilar membrane motion/pressure, thus it requires significant modification to operate in the context of the current complete auditory model. However, the underlying theories behind the model have been proven to be sound and therefore useful in developing a suitable inner hair cell model for the inner ear sub-system.

Detailed information regarding the theory and experimental results of the original model can be found in the literature [133]. The current model being used can be described in conjunction with Figure 4.10.

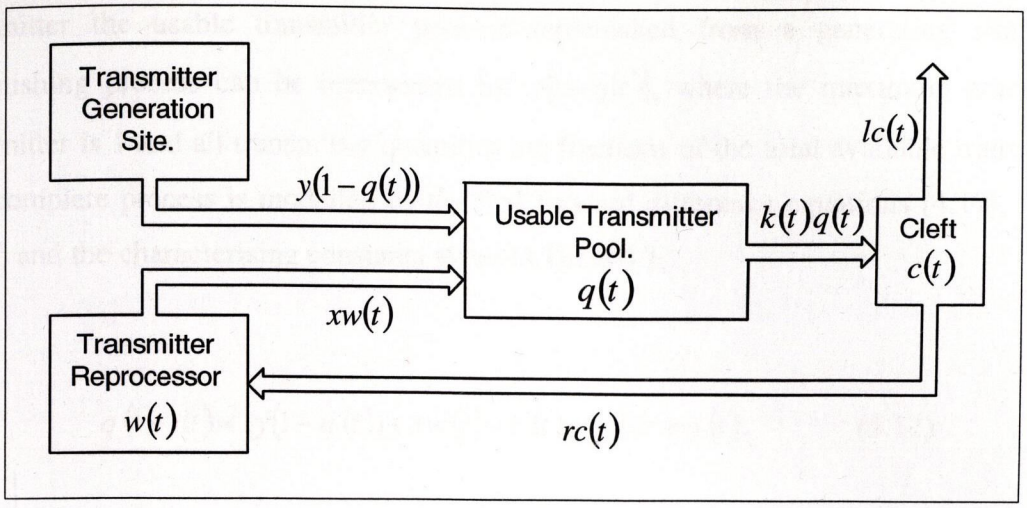


Figure 4.10. Inner hair cell model 1.

The hair cell itself contains a quantity of transmitter $q(t)$ near its cell wall where it synapses with the afferent nerve endings. When the hair cell is stimulated it releases a portion of this transmitter, $k(t)q(t)dt$ across its membrane into the cleft between it and the neuron it synapses with in a short time interval *dt*. The amount of transmitter released depends on the non-linear permeability of the membrane specified by $k(t)$, which depends on the instantaneous amplitude of the stimulating signal.

This dependence of $k(t)$ is given by

$$k(t)=\begin{cases} g\,[S(t)+A]/[S(t)+A+B] & \text{for } [S(t)+A]\rangle 0 \\ 0 & \text{for } [S(t)+A]\leq 0 \end{cases} \qquad (4.16)$$

where $S(t)$ is the instantaneous stimulating signal and A & B are constants. Not all of the transmitter reaching the cleft, $c(t)$, remains there as some of it escapes the cleft and is lost. This portion is represented by $lc(t)dt$. Similarly some of the transmitter leaks back into the hair cell, represented by $rc(t)dt$. The remaining transmitter in the cleft activates the postsynaptic membrane causing a nerve impulse to fire along the neuron. The transmitter that returns to the cell, however, is not immediately available to the cell as it goes through a reprocessing cycle which gradually adds back into the available pool of usable transmitter. The return of transmitter from the reprocessing pool $w(t)$ is represented by $xw(t)dt$. To compensate for the lost transmitter and the delay in the return of reprocessed

transmitter the usable transmitter pool is replenished from a generating site. This replenishing process can be represented by $y[1-q(t)]$, where the maximum amount of transmitter is 1 and all transmitter quantities are fractions of the total available transmitter. The complete process is modelled by the first forward difference equations (4.17), (4.18), (4.19) and the characterising constants given in Table 4.1.

$$q(t+dt)=[y(1-q(t))+xw(t)-k(t)q(t)]dt+q(t) \tag{4.17}$$

$$c(t+dt)=[k(t)q(t)-lc(t)-rc(t)]dt+c(t) \tag{4.18}$$

$$w(t+dt)=[rc(t)-xw(t)]dt+w(t) \tag{4.19}$$

Symbol	Value
A	5
B	300
g	1000
y	6.0
l	1250
r	16667
x	140
dt	0.000025

Table 4.1 Constants require to characterise the inner hair cell model 1.

As each sample of the basilar membrane response enters the hair cell model the future values of $q(t)$, $c(t)$ and $w(t)$ are evaluated. The output of the hair cell model is based on the averaged transmitter remaining in the cleft, $c(t)$, with the neural firing rate being evaluated by appropriately scaling the values of $c(t)$ to match the experimental measurements in [134] for the human auditory system.

4.4.3.2 Inner hair cell model 2.

The second hair cell model, based on the work of Allen and of Kates, develops an electrical circuit representation for the inner hair cell. This model has been designed to take

as its input the modelled basilar membrane pressure and to generate simulated spike trains in response. Originally, Allen designed the basic form of this model as a stand-alone model of inner hair cell transduction. The work of Kates adapted and extended it to take into account the two time constants. This current work re-characterises this model for use with the current modelled basilar membrane motion and expands the model to cater for the fact that all the transmitter chemical released into the cleft is not available to stimulate neural firing. The inner hair cell model uses the circuit in Figure 4.11 to characterise the processes taking place within the hair cell and is described as a variable resistance, dual time constant model.

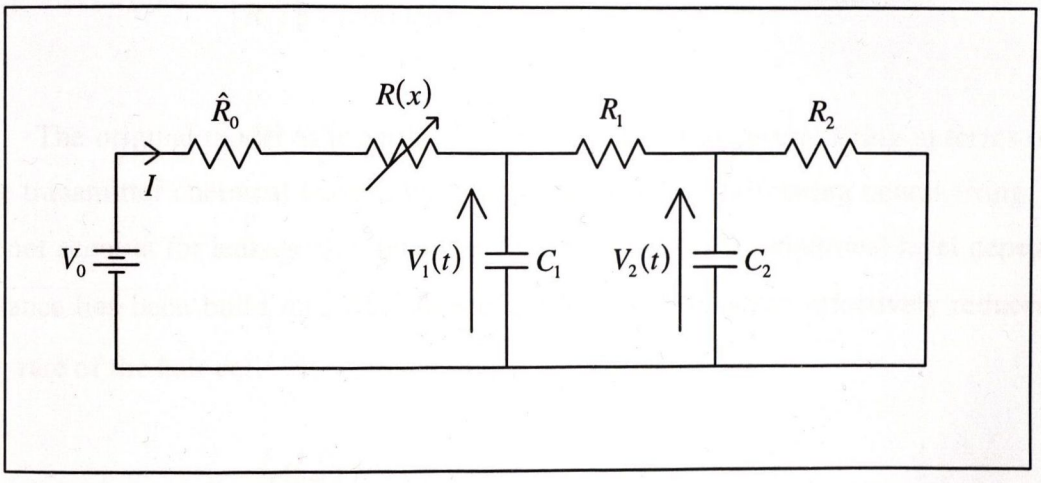


Figure 4.11. Inner hair cell model 2.

The current I flowing in the circuit is considered to represent the instantaneous firing rate taking place on the neural fibres synapsing with the stimulated hair cell. The combination of the two resistors, $R_1 \& R_2$ and the two capacitors, $C_1 \&\ C_2$ provides the dual time constants required to model the rapid and slow adaptation time constants as measured in [134]. The variable resistance $R(x)$ represents the resistance of the stereocilia controlling the flow of ionic current into the hair cell at a particular section (a normalised distance x) along the basilar membrane. It has a level dependence specified by.

$$R(x) = \frac{2r}{1+\exp(\alpha y)} \qquad (4.20)$$

where r is the characteristic resistance of the hair cell, y is the input sample from the basilar membrane model (it is also a function of x) and α is a scaling factor used to control the firing threshold. V_0 can be considered in terms of the potential difference across the cell.

The resistance generated by Equation (4.20) applies for all hair cell models with α being set to control the threshold of the fibre and thus specify the fibre type. The resistance $\hat{R}_0$ is set to a value R_0 which limits the neural firing of each hair cell model to remain within reasonable limits and acts to set the upper firing limit as dictated by the refractory period for nerve fibres. This is suitable for high-SR and medium-SR fibres as they saturate but low-SR fibres do *not* saturate. To cater for the special case of low-SR fibres $\hat{R}_0$ is allowed vary as

$$\hat{R}_o = \begin{cases} R_0 & y \le 0 \\ R_0 / \left[1 + 0.0003(\alpha y)^2\right] & y > 0 \end{cases} \tag{4.21}$$

The original model as proposed by Kates considers the neural firing in terms of the entire transmitter chemical leaving the cell being capable of activating neural firing. This does not account for leakage of transmitter from the cleft; so, an additional level dependent resistance has been build into $R(x)$ to account for this loss which effectively reduces the firing rate of the hair cell

$$R(x) = \frac{2r(1 + LF * y)}{1 + \exp(\alpha y)} \tag{4.22}$$

where LF is the transmitter loss factor from the cell.

The original foundation for this model has some important limitations. Primarily the hair cell model itself has a limited dynamic range going from spontaneous firing to saturated firing over approximately a 15dB range of hair-cell input stimulus [52]. Even with compression one hair cell model alone can only cover a dynamic range of about 37.5dB, which is totally inadequate to cover the auditory dynamic range of 120dB. It is known that multiple afferent neurons synapse with each hair cell ranging from high-SR fibres to low-SR fibres. Kate's solution to the range problem is to use two low-SR models and two high-SR models and average their contributions to generate the resulting neural information being sent from the hair cell to the brain. This solves the problem of range but ignores the medium-SR fibres present. On an important note, a hair cell with different variable-resistive elements does not agree with physiology; one hair cell model should really only have one variable resistance. The use of multiple variable resistances is,

however, a necessary compromise to get the proper population of neural fibres [52]. Analysis of the circuit under various conditions basically provides five equations in six model parameters [52]

Quiescent Current $$I_q = \frac{V_0}{R_0 + r + R_1 + R_2} \tag{4.23a}$$

Quiescent Voltage $$V_{1q} = \frac{V_0(R_1 + R_2)}{R_0 + r + R_1 + R_2} \qquad V_{2q} = \frac{V_0 R_2}{R_0 + r + R_1 + R_2} \tag{4.23b}$$

Saturation Current $$I_s \cong \frac{V_0}{2}\frac{1}{R_0 + R_1} \tag{4.24}$$

Peak Current $$I_p = \frac{V_0}{R_0}\frac{R_0 + r}{R_0 + r + R_1 + R_2} \tag{4.25}$$

Time Constants

$$\tau_1 = \frac{2C_1R_1}{2 + R_1/R_0 + V_1/(V_{1q} - V_1)} \quad \text{where } V_1 = \frac{V_0 R_1}{R_0 + R_1} \tag{4.26}$$

$$\tau_2 = \frac{2C_2R_2}{2 + R_2/(R_0 + R_1) + V_2/(V_{2q} - V_2)} \quad \text{where } V_2 = \frac{V_0 R_2}{R_0 + R_1 + R_2} \tag{4.27}$$

An arbitrary constraint placed on the system gives a final equation

$$R_1 = R_2 \tag{4.28}$$

The circuit operation is governed by the pair of coupled differential equations, Equations (4.29) & (4.30), which when changed into backward difference equations can be solved using the Equations (4.23-4.28).

$$\frac{V_0 - V_1(t)}{R_0 + R(x)} = C_1\frac{dV_1(t)}{dt} + \frac{V_1(t) - V_2(t)}{R_1} \tag{4.29}$$

$$\frac{V_1(t)-V_2(t)}{R_1}=C_2\frac{dV_2(t)}{dt}+\frac{V_2(t)}{R_2} \tag{4.30}$$

The lengthy solution to these equations provides a very elegant result for the instantaneous firing rate. This result is given by

$$I[n]=\frac{V_0-V_1[n]}{R_0+R(x)} \tag{4.31}$$

where $I[n]$ is the instantaneous firing rate on a neural fibre at time sample n. Not included above are the backward difference forms of the coupled differential equations. To produce these equations a sampling period must be introduced, synchronised with the sampling frequency of the system. Continuous time is therefore discretised when the characteristic equations of the circuit are converted to backward difference form. Thus, the solution in Equation (4.31) provides the firing rate at discrete intervals based on the sampling frequency of the model.

The updated model in this work extends the original hair cell model. Apart from including the ability to model the transmitter loss within the hair cell model, a new group of fibres has been added, the medium-SR fibres. Instead of using four hair cell models, the new system uses six hair cell models attached to the section filters and generates the neural contribution of each section to the overall neural map using the *average* neural response of the six outputs. This provides a much more accurate emulation of the transduction process. The updated model in this work uses the recent data of [134] and [131] to characterise the hair cell models. Table 4.2 specifies the particular values used to characterise the different fibres.

Fibre	Type	Quiescent Rate spikes/s	Threshold dBSPL
1	High-SR	75	0
2	High-SR	50	10
3	Med-SR	15	15
4	Med-SR	10	20
5	Low-SR	0.5	40
6	Low-SR	0.3	60

Table 4.2 Values used to characterise neural fibres attached to inner hair cell models.

Both of the models proposed can take as their input the modelled basilar membrane pressure and generate a simulated spike train in response to that input. The auditory model in its current form has the option of using either modelling method for the inner hair cells.

4.4.4 Modelling Feedback within the Inner Ear

When considering active feedback in the auditory system this predominantly deals with a system that manipulates the response of the basilar membrane. As the stimulus level rises the basilar membrane tuning curves broaden and the location of maximum excitation tends to increase. These phenomena are primarily associated with the stimulus level and can be explained in terms of increased displacement of the basilar membrane with increasing stimulus. There is, however, another mechanism at work in the form of outer hair cell activity. The mechanical function of the outer hair cells is believed to alter the micromechanical impedance (local impedance of small sections of the organ of Corti) within the organ of Corti. This follows from ear canal impedance measurements expressed as non-linear power reflectance, defined as the retrograde to incident power ratio [61]. The alteration of the micromechanical impedance is believed to be responsible for the dynamic compression of the auditory range. The Mössbauer measurements of Rhode [119] indicate this compression ratio to be around 2.5:1

In terms of modelling these feedback processes, the feedback mechanism in the model must interact with the section and second filters, as the basilar membrane is the site at which the feedback mechanisms effect their action. Modelling the stimulus level dependence of the tuning curves can be accomplished using adaptive level-dependent reduction of the characterising Q factors in the section and second filters. As the stimulus level increases, the level of the response at the frequency of stimulation will abnormally increase if the gain provided by the section filters is not reduced; the gain provided by the section filters in the region is therefore reduced to generate suitable cochlear tuning curves given the increased level of stimulation. Adaptively reducing the filter gain depending on the stimulus level results in a broadening of the response curve which is required to model the extra displacement of the basilar membrane. The combination of the reduced gain and the broadening of the filters causes a local reduction in the centre frequency of the section filters. This in turn lowers the overall centre frequency of the cochlear filter tuning curves equivalent to a shift in the location of maximum excitation which models the downward trend of the tuning curves with rising stimulus level.

Looking at the contribution of the outer hair cells in the feedback process, it is best considered in conjunction with the 1-dimensional transmission line segments specified earlier. The combination of all the local characteristic impedances of all the segments generates the overall micromechanical impedance of the modelled basilar membrane. The dynamic alteration of the section filter impedance is responsible for the generation of the dynamic micromechanical impedance and hence is responsible for the compression in the cascade. Now considering the way in which the transmission line segments have been modelled using the section filter transfer function, manipulation of their Q factors has the effect of manipulating the local characteristic impedance of the section filters. However, the stimulus level plays a role in altering the Q of the filters to manipulate their gain, bandwidth and centre frequency. The feedback activity of the outer hair cells must therefore operate in conjunction with the stimulus level manipulation of the Q for the section filters. This is a valid approach because the stimulus level on the basilar membrane also governs the activity of the outer hair cells. It is also worth noting that the modelled outer hair cell activity governs the limits of the Q factors as they are ultimately responsible for the overall micromechanical impedance of the basilar membrane. In terms of the auditory model we only consider fully functional outer hair cell involvement; malfunction of the outer hair cells is considered in Chapter 5. The function that specifies the overall adaptation process for the section filters is given by.

$$Q_{pSect} = (0.3 + 0.7 * Adapt) * Q_{SectionMax} \tag{4.32}$$

The overall adaptation function for the second filters is given by.

$$Q_{pSec} = (0.1 + 0.9 * Adapt) * Q_{SecondMax} \tag{4.33}$$

The general form of the adaptation equations is similar to that of [49] as they specify the *characteristics of operation* of the biological human auditory system. However the Q factors required are different because of the new section and second filters. The adaptation factor *Adapt* is also different as it implements the outer hair cell involvement and has to cater for the needs of the new filters. The function for *Adapt* is given by.

$$Adapt = 0.3 * d + 0.75 * d^3 \tag{4.34}$$

$$d = \frac{(BM_{Peak} - y_{PEAK})}{BM_{Range}} \tag{4.35}$$

where y_{PEAK} is the peak mechanical output of the modelled basilar membrane around the current section (in dB), BM_{Peak} is the peak response of the basilar membrane model (typically 120dBSPL) and BM_{Range} is the dynamic range of the basilar membrane response (typically 80dB). The peak detection of the value of y_{PEAK} is generally localised around the section in question with a detector having instantaneous attack and a 100ms release time.

Each filter section and second filter in the cascade requires a feedback system to provide suitable adaptive system control so all the sections together can generate stimulus level dependent cochlear response curves the same as those measured in the auditory system. The adaptive system control also embodies a compressive non-linearity with a compression ratio of approximately 2.5:1 across the full range of auditory stimuli 0dBSPL to 120dBSPL.

A general description of the function of this type of system can be given in conjunction with [17], [10] & [49]. When a stimulus is presented to the cascade all the sections are initially configured to provide maximum gain. If the stimulus is just above threshold only the sections near the place of maximum excitation will see a reduction in their gain by the adaptive feedback system. The outer hair cell activity will also tend to increase to generate compression of the increasing response. Increasing stimulus level brings more sections above threshold and causes their gain to reduce. The stimulation of more sections also results in increased outer hair cell activity in these sections. However, the outer hair cell activity will be affecting more of the basilar membrane thus each section will only see a smaller change to maintain the overall compression in the cascade. Eventually a stage is reached where the stimulus is at its maximum level and is stimulating all the sections possible and all these sections will be operating at their minimum gain.

4.4.5 The Complete Inner Ear Model

The complete inner ear model consists of a cascade of sections with each section incorporating a section filter, second filter, a peak detector, a multi-neuron hair cell model

and a feedback mechanism. All of these sub-models work together to model the inner ear function. Figure 4.12 indicates the layout of one of the 120 sections in the cascade used to model the inner ear.

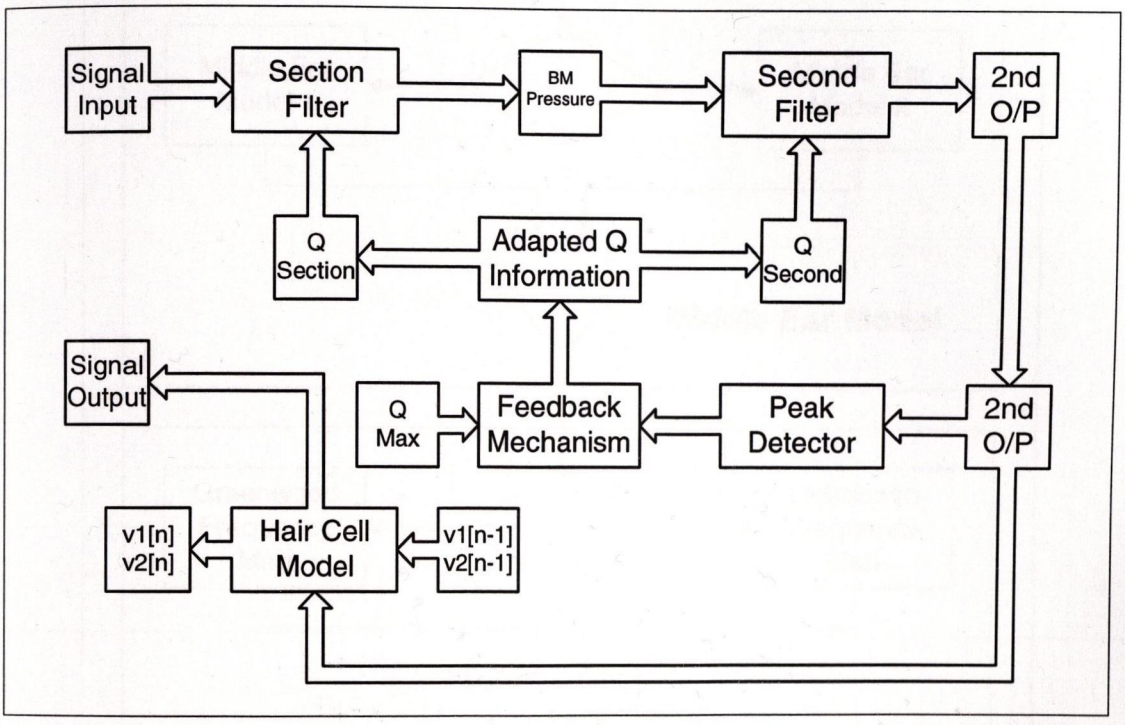


Figure 4.12 Layout of one of the complete sections in the cascade.

4.5 The Complete Auditory Model

The sub-models of the outer, middle and inner ear have been designed to operate as part of a complete model of the auditory system not, however, to the extent that the sub-models are inflexible. The characteristic specification of each component during the integration of each sub-model is such that any of the sub-model's components can be replaced with other models using suitable interfacing elements. This is to facilitate upgrading the model as more information about the auditory system becomes available and also to allow experimentation with other models. The model takes as its input raw sound pressure waves just as the auditory system would receive them and passes the signal through the outer, middle and inner ear models to produce as an output a map of the neural firing patterns sent to the brain. The output is in the form of *joint time-frequency firing rate data* with each channel being the neural output for each point on the basilar membrane. Figure 4.13 indicates the general layout of the complete auditory model.

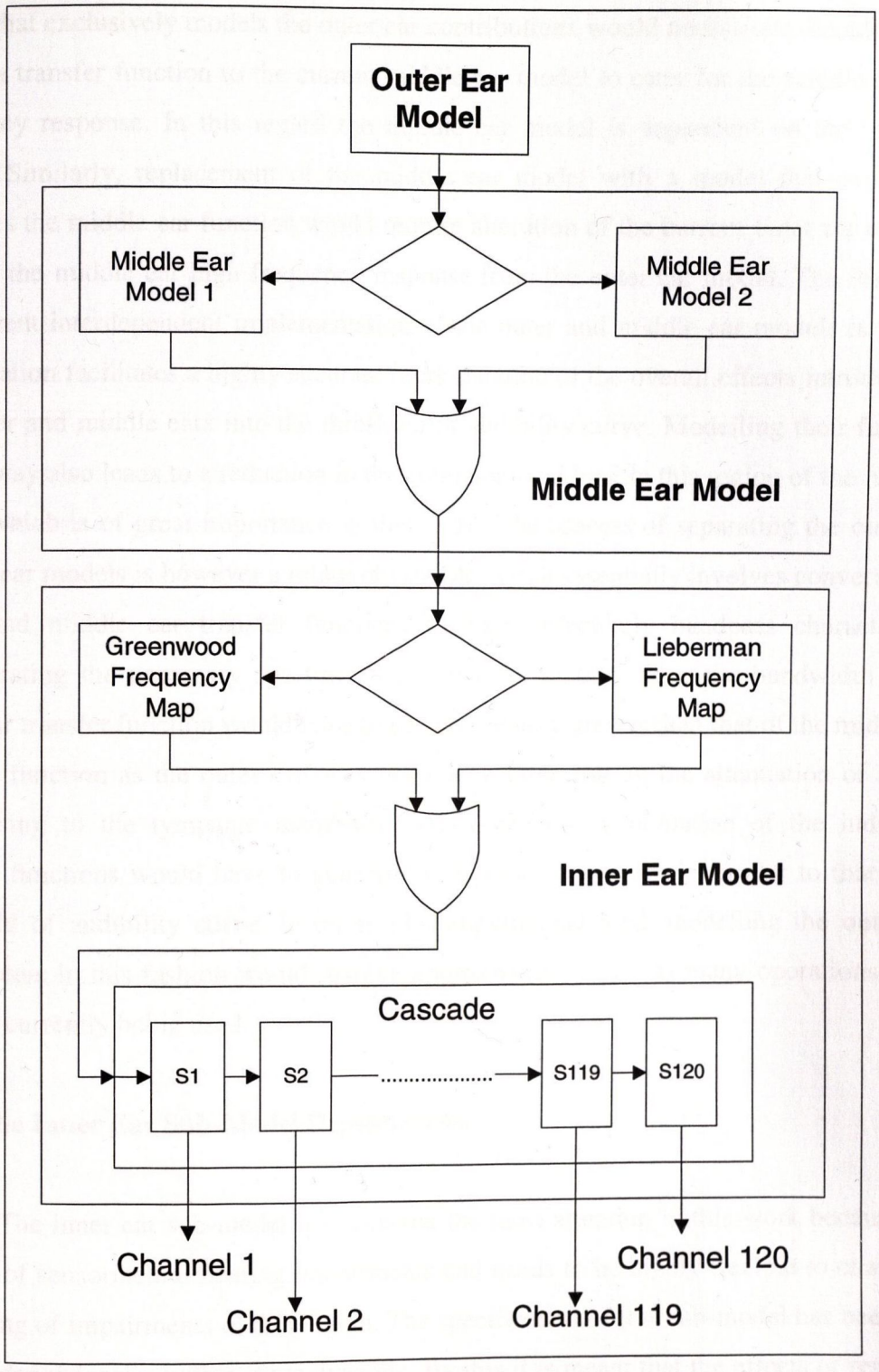


Figure 4.13 Layout of the auditory model.

4.5.1 The Outer and Middle Ear Sub-Model Dependencies

The outer ear model is not particularly dependent on any other section and can easily be replaced with other models that emulate its function. However, as explained earlier, the current arrangement of the outer ear model transfer function incorporates the high frequency response of the middle ear. Replacement of the outer ear model with a

model that exclusively models the outer ear contributions would necessitate the addition of an extra transfer function to the current middle ear model to cater for the middle ear high frequency response. In this regard the middle ear model is dependent on the outer ear model. Similarly, replacement of the middle ear model with a model that exclusively emulates the middle ear function would require alteration of the current outer ear model to remove the middle ear high frequency response from the outer ear model. The reason for the current interdependent implementation of the outer and middle ear models is that the combination facilitates a highly accurate representation of the overall effects introduced by the outer and middle ears into the threshold of audibility curve. Modelling their functions in this way also leads to a reduction in the computational load in this region of the model, a factor which is of great importance in this work. The process of separating the outer and middle ear models is however a relatively simple one. It essentially involves converting the outer and middle ear transfer functions to have effectively bandpass characteristics incorporating the necessary spectral effects of each section. Also the bandwidth of the outer ear transfer function would have to be significantly greater than that of the middle ear transfer function as the outer ear does not play a large role in the attenuation of sounds progressing to the tympanic membrane. Obviously the combination of the individual transfer functions would have to generate a response characteristic similar to that of the threshold of audibility curve. In terms of computational load, modelling the outer and middle ears in this fashion would involve approximately twice as many operations as the method currently being used.

4.5.2 The Inner Ear Sub-Model Dependencies

The inner ear sub-model has received the most attention in this work because it is the site of sensorineural hearing impairments and needs to be highly flexible to enable the modelling of impairments of this region. The specification of this sub-model has been such that all dependencies within it are dynamic. By this it is meant that the effects of replacing any component in the model are incorporated into the calculation of the parameters for the inner ear sub-model. This leads to a delay during the initialisation of the model as the characteristic parameters of the inner ear have to be calculated based on components that are active in the model. However, it makes the model highly flexible. This delay could be avoided by statically specifying the parameters for the inner ear sub-model but would confine the modes of operation of the model and make it difficult to use alternate sub-

models. In this way the inner ear sub-model is dependent on the other components in the auditory model but is not limited by this dependence.

4.5.3 Model Initialisation

Before the model can receive input an initialisation process must take place. Prior to this process taking place, the model has a built-in specification for the outer and middle ear sub-models, default values for all biological values and only a general specification for the inner ear sub-model. It is only at the end of the initialisation process that the parameters for the inner ear model are calculated and the inner ear sub-model is fully specified. When this process has taken place the cascade and the feedback mechanisms are in a *no input* state, only after a few iterations of the complete model cycle does it reach its spontaneous mode of operation and is ready to generate a response to a particular input. During the initialisation process the user can specify a time over which the model is allowed preamble and reach its quiescent state. Output generated during this time is not stored. Apart from this, the initialisation process allows the user specify different values for the various biological variables and also allocates the required local and global memory for inputs, outputs and intermediate data vectors.

4.5.4 Complete Model Implementation and Operation

The entire model is a time domain model, therefore all z-domain transfer functions are realised in their difference equation form. In particular, they are realised as *recursive* difference equations with *sufficient auxiliary conditions* to fully parameterise them. The auxiliary conditions required by these time domain discrete equations are not known beforehand as these auxiliary conditions characterise the model in its spontaneous state and depend on the active components in the system. The model must complete a few full cycles before sufficient auxiliary conditions can be generated to characterise the spontaneous mode of operation of the model. Specifying a preamble time during the model initialisation will allow this process to take place. Alternatively if the response of the model to the first few samples of the stimulus being presented is of little importance, the model can operate straight from the *no input* state with the auxiliary conditions set to zero. Regardless of the auxiliary states the model always moves to the correct operating state based on the data being processed. The effect of moving straight from the no input state to an above threshold input is an initial output that has a slightly lower firing rate response in the neural

output than would occur if the sample was processed in the model from its fully initialised state.

Once the model is fully ready to receive a stimulus, each sample is presented as input to the outer ear sub-model. The outer ear sub-model generates a response that is passed to the middle ear sub-model, which in turn presents its output to the inner ear sub-model. The cascade in the inner ear sub-model processes the sample and generates the necessary basilar membrane response. The peak response is then calculated around each section and the feedback mechanism and outer hair cell action in each section actively adjusts the cascade response for the next sample. Simultaneously the inner hair cell models in each of the sections respond to the basilar membrane pressure and generate the neural output going to the brain. The combination of all the hair cell outputs generates a map of the neural firing patterns on the nerve fibres going to the brain. This output information is stored for the entire input and plotted as a joint time-frequency firing rate map. Provided the model accurately models the auditory system, the output will be a neural map of the information the brain receives when the ear is presented with a particular stimulus.

4.6 Simulation of the Auditory Model

This section deals with extracting the responses of the auditory model and comparing the results with the responses of other models and with the known responses of the auditory system. It will also deal with simulation of the model using various inputs and examination of the resulting outputs from the model. In particular, it will look at the responses of the outer and middle ears, the tuning curves of the basilar membrane model, the tone burst response of the model, the compression ratio generated in the cascade, the ramped and damped responses of the model and finally the model's response to synthetic speech.

4.6.1 The Outer and Middle Ear Response

The auditory model has access to two different middle ear models so there are two possible responses for this section depending on which model is active. Figure 4.14 shows the frequency response of the outer and middle ear system using the first middle ear model with the operational R-value set to 0.989.

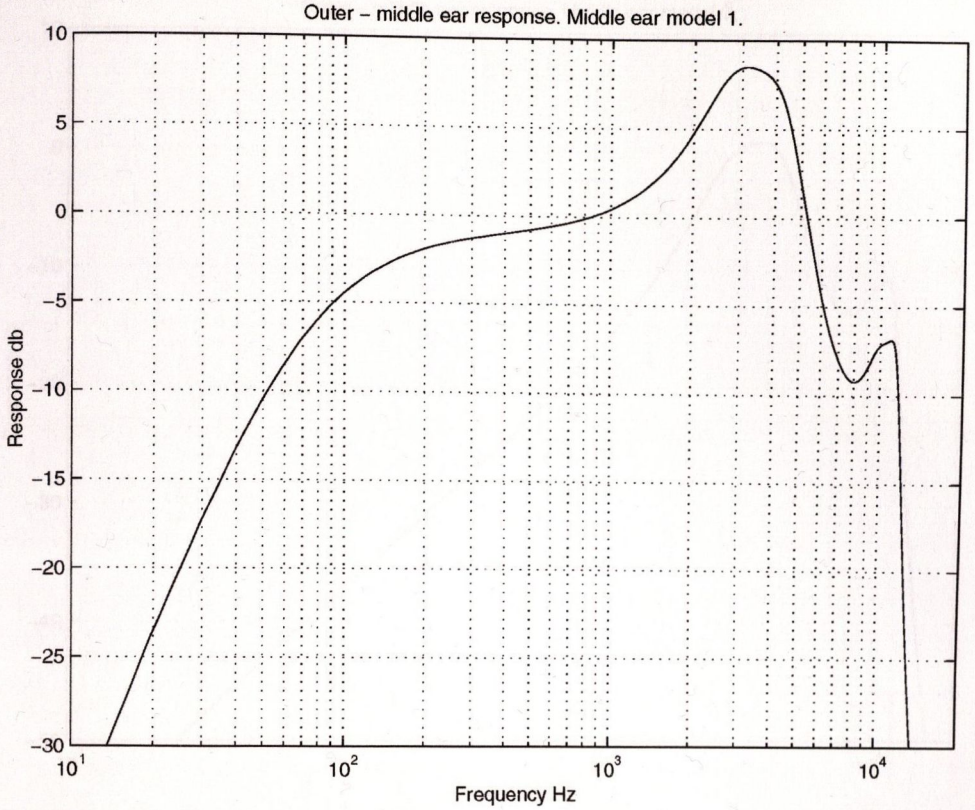


Figure 4.14. Outer – middle ear model response using middle ear model 1.

The response of the combination of the outer ear model and the first middle ear model shows very good correlation with the responses measured audiologically [115]. It demonstrates the required enhancement of the mid-frequencies through resonance effects with a maximum gain between 3-4kHz. The correlation between the response and the threshold of audibility curve can be most easily seen if one considers the response curve inverted. This arrangement focuses on accurate high frequency representation while maintaining good overall frequency representation. An important point to note is that this model assumes that the transmission through the outer and middle ears below 1kHz is reflected in the inverted shape of an equal-loudness contour at a *high loudness level*. This model is based on a modified version of the 100-phon loudness contours specified in ISO 226 [135], [117].

Considering the combination of the outer ear model and the second middle ear model with operational parameters $\omega = 2 * \pi * 4500/40000$ and $Q = 1.4000$, the plot in Figure 4.15 shows the response of this combination.

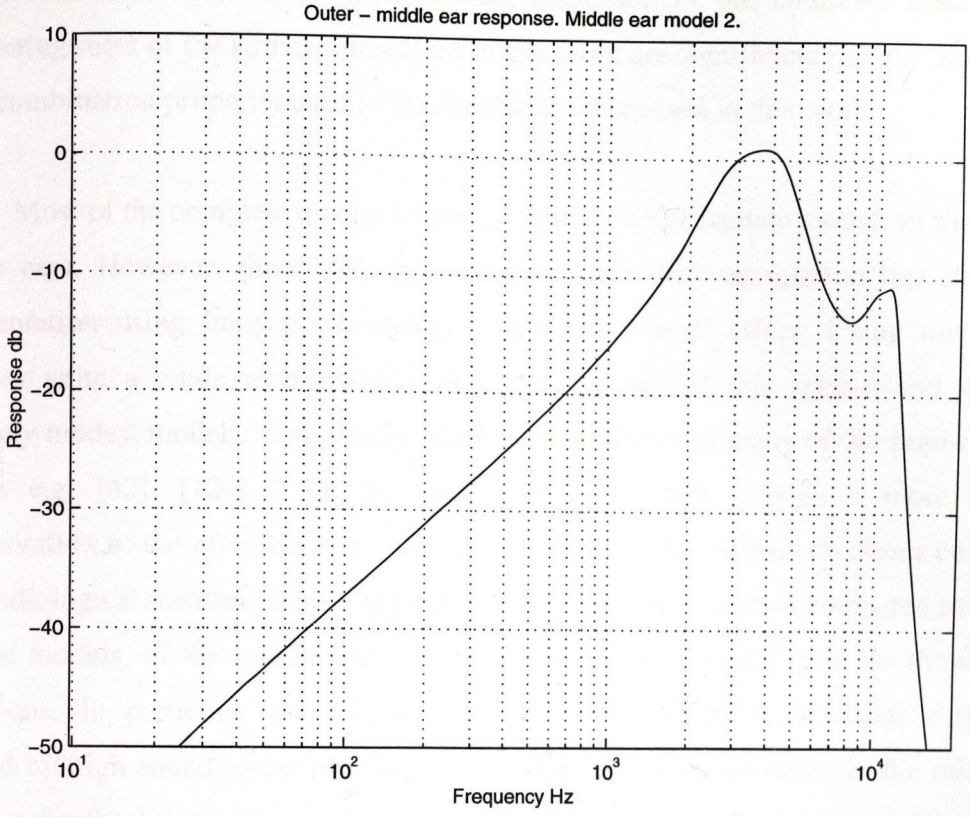


Figure 4.15 Outer – middle ear model response using middle ear model 2.

The response of the second arrangement can be seen to be somewhat different to that of the first. The general characteristics of the gain around the ear canal resonant frequency is present. However, an overall reduction in gain of approximately 8-18dB can be seen across the spectrum compared to the values present using the first middle ear model. The low frequency reduction in gain shows general agreement with the *low-level loudness contours* presented in ISO 226 [135]. In particular this combination reveals good qualitative correlation with the recent work of Pascal, Bourgeade, Lagier and Legros [59]. Quantitatively this arrangement and the model in [59] show the same trends; however, the current model shows a reduced gain of around 18dB over all frequencies in comparison. This can be easily remedied by introducing a gain factor. Although not exhibiting a particularly good match with the audibility threshold curve at low frequencies, this combination shows better correlation with low-level loudness contours in ISO 226 [135].

Both combinations of models provide good overall responses. The low frequency attenuation of the first combination shows the best response for signals well above threshold while the second combination shows the best response for signals around threshold compared to the ISO loudness contours derived experimentally [135]. Of the two

arrangements, the first model shows closest emulation of the complete spectrum and considering most of the stimuli considered in this work are significantly above threshold, it is the combination primarily used in the simulations discussed in this work.

Most of the complete models currently around do not include models of the outer or middle ears. However, those that do, tend to opt for the transmission line method of representation using lumped parameters to give an overall effect. Using this method provides quite a crude representation of the physiology of this system and generates relatively modest models. Compared to the responses shown in many of the semi-complete models e.g. [42], [124], [52], the models proposed here provide a more accurate representation of the effects of the outer and middle ears showing much closer correlation with audiological measurements and recent standards. However when compared to specific *isolated* models, of the middle ear for example, the models being used do show certain limitations. In particular isolated models emulate the complex muscular contraction induced by high sound levels reaching the cochlea. This acoustic reflex in the middle ear acts as a feedback loop to decrease the sound pressure entering the cochlea. The models being used cannot emulate the dynamic effects of acoustic reflex in the middle ear as they have static responses that emulate the average acoustic reflex for normal sound levels. More complete isolated middle ear models such as [59] and [56] can provide for such features as that is their focus and they do not consider the rest of the auditory system. In conjunction with this acoustic reflex, the middle ear also exhibits to a first order approximation a frequency-independent delay [56] something that the models being used do not implement.

Provision for acoustic reflex would require making the model's response dynamic and dependent both on the stimulus duration & frequency, adding significantly to the processing required in this section, with relatively minimal accuracy enhancement. Using an averaged acoustic reflex for normal sound pressure levels shows a very good match with these isolated models; where the greatest discrepancy is found is when very high and very low level signals are presented to the models. In such cases the acoustic reflex is under and over estimated respectively in the models being used but is not a severe limitation when considering the types of signals being dealt with. Considering the frequency-independent delay, it would not be overly complicated to implement but in terms of a complete model of the auditory system a feature such as this is not needed when considering input-output pairs. Overall, the outer and middle ear models being used show a

good correlation with measured data and compare favourably with other models in the literature.

4.6.2 Tuning Curves of the Basilar Membrane Model

Considering the cascade being used to emulate the basilar membrane motion in the inner ear model, the auditory tuning curves that are emulated in its structure are of particular importance. These auditory tuning curves can be extracted for each of the centre frequencies in the cascade using a relatively simple but computationally intensive process as follows.

Initially a reference value must be generated by stimulating the model with a tone at the centre frequency of the section being examined, having intensity just above threshold to give the steady state 0dBSPL level. A test vector of frequencies must then be generated for each of the filter section centre frequencies in question. This vector of frequencies needs to cover the range from around 100Hz to one octave above the centre frequency of the section being analysed. A 47.5ms tone is generated for each frequency in the test vector at the same intensity level as the signal used for the reference. The model is then stimulated with each tone and the input amplitude adjusted until the steady state output of the section is within 0.5% of the reference level. This process uses the method of bisections to converge on the required input amplitude and generally takes between 10 and 30 steps to converge. The generation of an auditory tuning curve for a particular frequency therefore requires a significant number of complete iterations of the model. The application of this process to all the section centre frequencies generates a complex family of 120 auditory tuning curves. There are, however, two possible sets of curves that can be extracted, one based solely on the basilar membrane motion and the other based on the basilar membrane motion and the resonance of motion between the basilar and tectorial membranes. This can be accomplished through selection of the output extraction point, whether it is at the section filter output or the second filter output. It is important to note the feedback mechanism and the compression for both sets of curves depend on the localised peak second filter output regardless of the extraction point. Display of the complete sets of curves is not possible as the graphs become extremely cluttered making it impossible to distinguish the salient features of the curves. A sample of the curves extracted from the cascade around the mid-frequency region of the auditory spectrum is therefore shown in Figures 4.16 and 4.17. These curves can be directly compared to

auditory tuning curves measured experimentally where the lowest points indicate the location of maximal sensitivity.

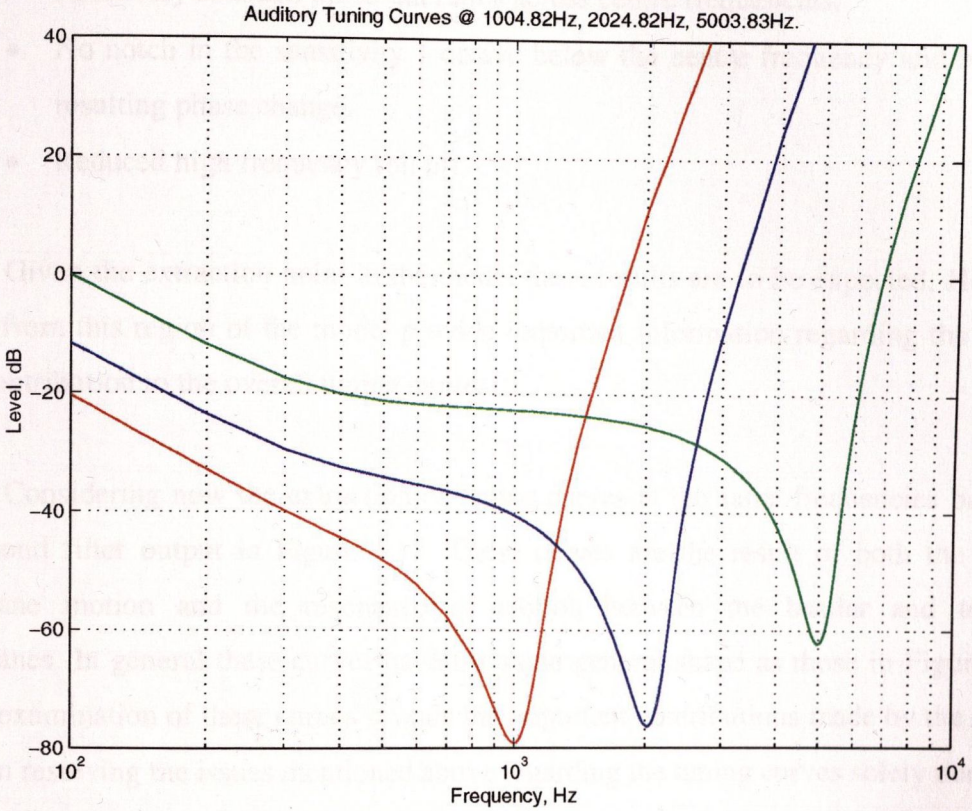


Figure 4.16 Family of auditory tuning curves based on the section filter output alone.

The plot shown in Figure 4.16 shows the auditory tuning curves based on extraction from the filter section output and hence shows the contribution of the cascade of section filters on their own. The three plots depict the auditory tuning curves at 1004.82Hz, 2024.82Hz and 5003.83Hz. Examination of these curves shows that the cochlear filters have been located at the correct centre frequencies as required by the frequency mapping functions. Qualitative analysis of the curves shows that they are relatively broadly tuned, respond highly non-linearly around their centre frequency and respond linearly further away from their centre frequency. They have a much steeper roll off on their high frequency side than their low frequency side with the tuning curves becoming gradually less tuned for lower centre frequencies. Most importantly, the curves exhibit the low frequency extended tails consistent with displacement output. In general, the overall curves have the required shapes showing sharp tuning around the centre frequencies; however, the absence of the contribution of the tectorial membrane results in several important effects:

- Excessive broadening around the centre frequencies.
- An overly rapid decrease in the peak sensitivity across centre frequencies.
- Overly sensitive tails in the tuning curves.
- Relatively constant tip-to-tail ratios across centre frequencies.
- No notch in the sensitivity 1 octave below the centre frequency and hence no resulting phase change.
- Reduced high frequency roll off.

Given the extraction point in the model these results are to be expected. However, curves from this region of the model provide important information regarding the section filter contribution to the overall tuning curves.

Considering now the extraction of tuning curves at the same frequencies based on the second filter output in Figure 4.17. These curves are the result of both the basilar membrane motion and the resonance of motion between the basilar and tectorial membranes. In general these curves have the same general shape as those in Figure 4.16. Closer examination of these curves reveals the important contributions made by the second filters in resolving the issues mentioned above regarding the tuning curves solely due to the section filters output. In particular the following can be seen:

- Reduction of the bandwidth of the tuning curves around the centre frequency; in general their bandwidths are linearly proportional to their centre frequencies above 800Hz; for lower centre frequencies the tuning curves become gradually less tuned.
- Reduction of the rate of decay in the peak sensitivity across mid-range centre frequencies.
- Reduction in the sensitivity of the tuning curves tails.
- Tip-to-tail ratios that increase with increasing centre frequency.
- A notch in the sensitivity 1 octave below the centre frequency and a corresponding phase change at this frequency.
- Enhanced high frequency roll off.

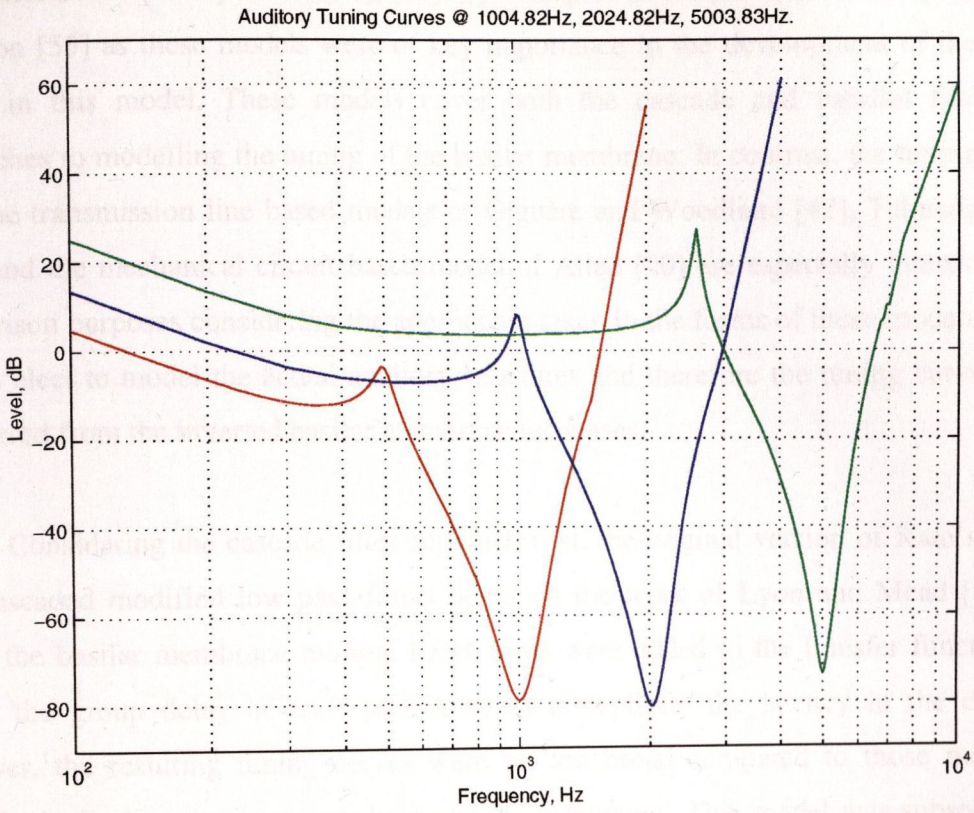


Figure 4.17 Family of auditory tuning curves based on the second filter output.

Quantitative analysis of the two different sets of tuning curves can be seen in Table 4.3, indicating the contribution of the second filters as described above.

Frequency Hz	Peak BM dB	Peak BM&TM dB	Tip-to-tail BM dB	Tip-to-tail BM&TM dB	Roll-off BM dB/octave	Roll-off BM&TM dB/octave
1004.82	-78.97	-78.97	35.89	66.73	89.93	134.47
2024.82	-76.28	-80.68	40.99	73.50	100.81	141.76
5003.83	-62.09	-73.22	38.79	77.21	106.35	133.52

Table 4.3 Analysis of the auditory tuning curves with and without tectorial membrane contributions.

Clearly the incorporation of the second filter in the model greatly enhances the response of the emulated basilar membrane in this system. The auditory tuning curves generated in this model compare very favourable with the tuning curves generated in other recent filter based models, transmission line based models and mechanical circuit based models. In terms of filter based models, of particular interest for comparison purposes are

the models developed by Kates [52], [49], by Pflueger, Hoeldrich and Riedler [54] and by Patterson [55] as these models were of key importance in the development of the tuning curves in this model. These models cover both the cascade and parallel filter bank approaches to modelling the tuning of the basilar membrane. In contrast, the tuning curves from the transmission line based models of Giguère and Woodland [42], Talmadge et al. [124] and the mechanical circuit based model of Allen [20] are especially interesting for comparison purposes considering the approaches taken in the forms of these models. These models elect to model the actual auditory structures and therefore the tuning curves must be inferred from the inverted basilar membrane responses.

Considering the cascade filter approach first, the original version of Kate's model [52] cascaded modified low pass filters based on the work of Lyon and Mead [123] to model the basilar membrane motion. Extra zeros were added to the transfer functions to reduce the group delay of each section so as to replicate the latency in the cochlea. However, the resulting tuning curves were far too broad compared to those measured physiologically in animals or psychophysically in humans. This model was subsequently updated by Kates [49] to use transmission line sections. In the light of the earlier explanation of the problems with the updated model, the effects can be clearly seen in the tuning curves extracted from that model. Kate's models are based on suitably modified cat auditory tuning curves from [128]. However, the tuning curves the model generates fall short in their representation of these curves particularly at mid to high frequencies. The peak sensitivities are too low, the high frequency slopes over the first octave underestimate the human slopes for simultaneous masking [136] and the high frequency slopes decrease with increasing frequency more than in measured tuning curves. Similarly, the change in sensitivity with frequency is not nearly as pronounced as in the measured curves and the bandwidths of the tuning curves are still too broad at low sensitivity levels. These details can be seen comparing the inverted tuning curves from Kates [49] in Figure 4.18 to the measured data in Figure 4.19 Kiang [128] and of Nelson and Fortune in Figure 4.20 [136].

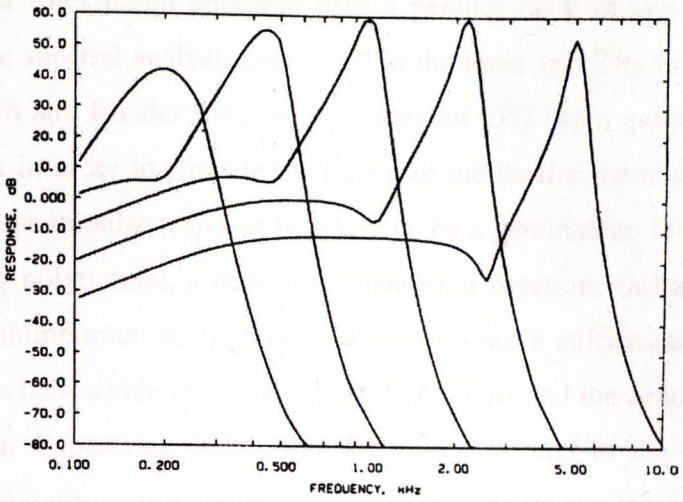


Figure 4.18 *Inverted* tuning curves in Kate's recent model, [49].

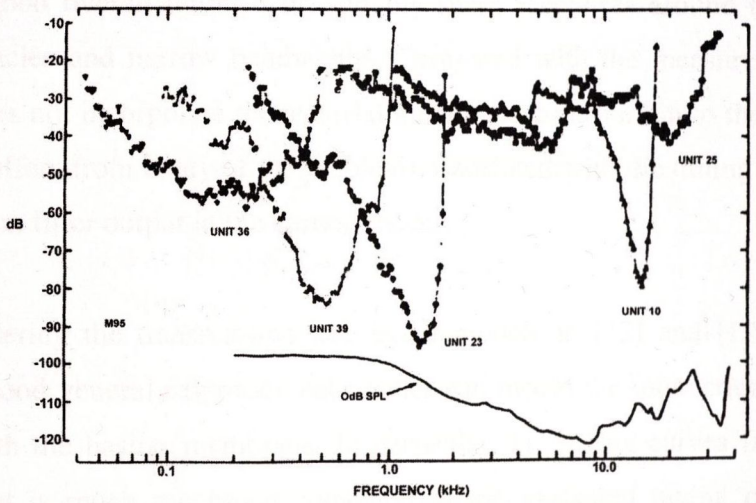


Figure 4.19 Set of measured auditory tuning curves, Kiang [128].

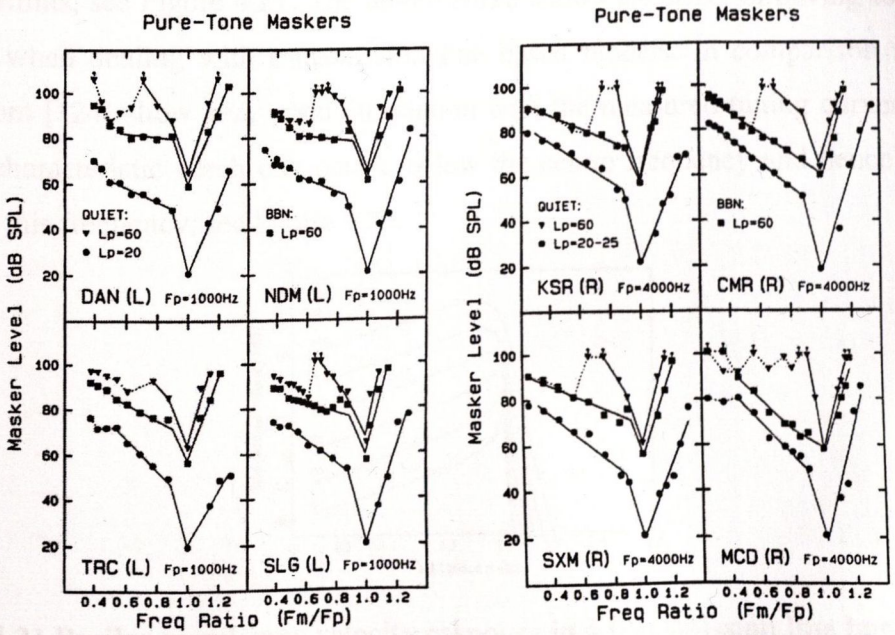


Figure 4.20 Psychophysically measured auditory tuning curves, Nelson et al. [136].

The second filter based approach uses a parallel bank of overlapping band pass filters to model the spectral analysis performed in the inner ear. The models developed by Pflueger, Hoeldrich and Riedler [54] and by Patterson [55] use a gammatone filter in the parallel filter bank in order to emulate the tuning of the basilar membrane. A gammatone filter is a filter whose impulse response is described by a gammatone function, which is the product of a rising polynomial, a decaying exponential function, and a cosine wave. The exact function is unimportant in the present discussion but it suffices to say that it allows specification of the bandwidth, the slopes of the filter skirts and the centre frequency of the filter. Of particular importance is that this function is similar to the impulse responses obtained from cat experiments using the *revcor* technique and can therefore provide a link between physiological and psychoacoustic data [54]. In general the responses of these filters show good overall tuning with suitably sharp responses around the characteristic centre frequencies and narrow bandwidths. Compared with the measured tuning curves, this model does not incorporate the tectorial membrane contribution to the auditory tuning and as such suffers from many of the problems associated with the tuning curves extracted from the section filter output in the current model.

Considering the transmission line based models in [42] and [124], both models exhibit very good general responses but neither can model the interaction of the tectorial membrane with the basilar membrane. In particular, the tuning curves from [42] have a bandwidth that is much too broad compared to the measured tuning curves at low to normal stimulus levels; however, at higher stimulus levels the bandwidths are within tolerable limits, see Figure 4.21. The dotted curve shows the effect of having too coarse a step size when dealing with transmission line based models. In comparison the tuning curves from [124] show very good correlation with the measured tuning curves and only lack the characteristic notch one octave below the centre frequency and hence the phase change at this frequency, see Figure 4.22.

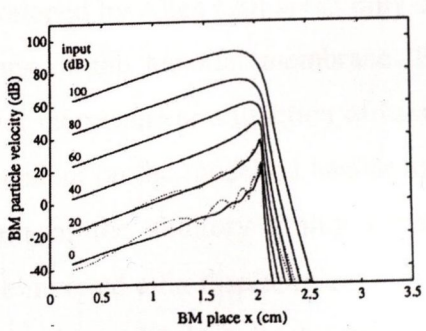


Figure 4.21 Basilar membrane velocity response in a transmission line based model. [42].

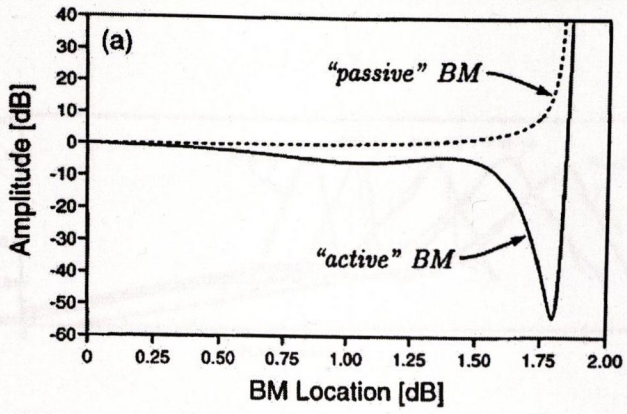


Figure 4.22 Basilar membrane amplitude response in a transmission line based model. [124].

Most transmission line based models exhibit good general auditory tuning curves but suffer from similar problems to those of the model using a gammatone filter bank. In particular, these models do not incorporate the characteristic notch in sensitivity one octave below the centre frequencies and the associated phase change at that frequency. Reduced peak sensitivities in the auditory tuning curves are also found as the resonance of motion between the two membranes is absent. As a result the bandwidths of the tuning curves tend to be underestimated at low stimulus levels or overestimated at high stimulus levels. Similarly the peak sensitivity levels tend to show a good match at low intensity levels but not at high intensity levels. These problems predominantly arise because of the compromises made in the approximation of physiological characteristics with lumped parameter blocks within these transmission line based models. Unfortunately, in many cases the use of lumped parameters in transmission line models cannot be avoided given the physiological data presently available.

Of all the models examined, apart from the models by Kates in [49] and [52], the macromechanical model developed by Allen [20] is the only other model that successfully incorporates the contributions of the tectorial membrane. Figure 4.23 shows the cilia displacement along the basilar membrane as a function of basilar membrane location. This displacement is directly dependent on the modelled basilar membrane response. The cilia stimulation is a direct result of the auditory tuning curves exhibited by the basilar membrane motion; hence the inverted cilia displacement curves are a close representation of the auditory tuning curves in the model. Despite the cluttered nature of the plot in Figure

4.23, the model can be seen to exhibit many of the salient features required for successful tuning curves.

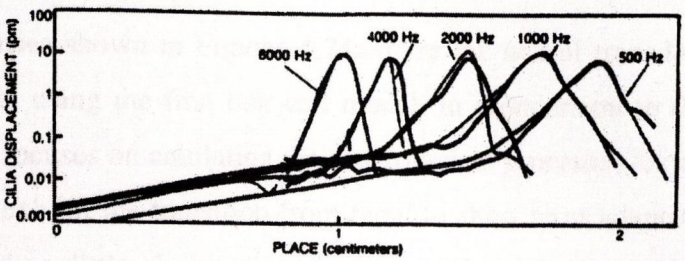


Figure 4.23 Cilia displacement curves in a mechanical circuit based model, [20].

In comparison to Kate's models, the model in the current work shows much better agreement with the peak sensitivities and high frequency roll off of the measured tuning curves. Similarly the decrease in the high frequency slopes with increasing frequency does not seem to be a problem using the current system and the change in sensitivity level with frequency shows better correlation with the measured data. Relative to the gammatone model and the transmission line models, the current model shows superior emulation of the detailed physiologically measured tuning curves in animals and the psychophysically measured tuning curves in humans. Of all the models examined, the tuning curves extracted from Allen's and Kate's models are the only ones that include the tectorial membrane interaction and show many of the finer details found in auditory tuning curves. In general, all the models show good overall agreement with the measured tuning curves; however, the current model exhibits the greatest correlation with the current data in the literature.

4.6.3 Tone Burst Response of the Auditory Model

Stimulation of the auditory model with tone bursts allows extraction of the inner hair cell characteristics. In particular, the tone burst response is used to extract the temporal characteristics of the inner hair cell adaptation and the steady state response. The inner hair cell response should show a transition from rapid to short term adaptation and a reduction in latency as the tone burst level increases. With this in mind the model has been stimulated using a 1kHz tone at four different levels, 30dBSPL, 60dBSPL, 90dBSPL and 120dBSPL. Prior to application of the stimuli, the model is allowed reach its spontaneous mode of operation. As there are two different neural transduction models there are 2

groups of responses. These can be seen in Figures 4.24a-d (Model 1) and Figures 4.25a-d (Model 2).

The responses shown in Figures 4.24a-d are the neural transduction outputs from the auditory model using the first hair cell model. In comparison to the second hair cell model this model focuses on emulating the physiological processes taking place in the hair cells. The results exhibit the transition from rapid to short term adaptation between 10ms and 20ms which is a little shorter than experimental values but within tolerable limits [137]. The firing rates peak at the stimulus onset and decay to their steady state values after the adaptation time. In general the firing rate levels show good correlation with the responses in the literature; however, at low intensity levels the peak responses are a bit low and at high intensity levels the steady state level is a bit low. The reason for this is that this hair cell model is configured to model medium spontaneous rate neural fibres and hence shows excellent correlation with mid range intensity levels. The high and low intensity response levels are, however, still within acceptable limits. The tone burst responses illustrate the model's response for increasing stimulus intensity in steps of 30dB. In each case the latency of the modelled firing rate reduces as the input level rises; this can be seen in the fact that the firing-rate peak occurs earlier as the input level increases showing good correlation with results of [138], [134] and [10].

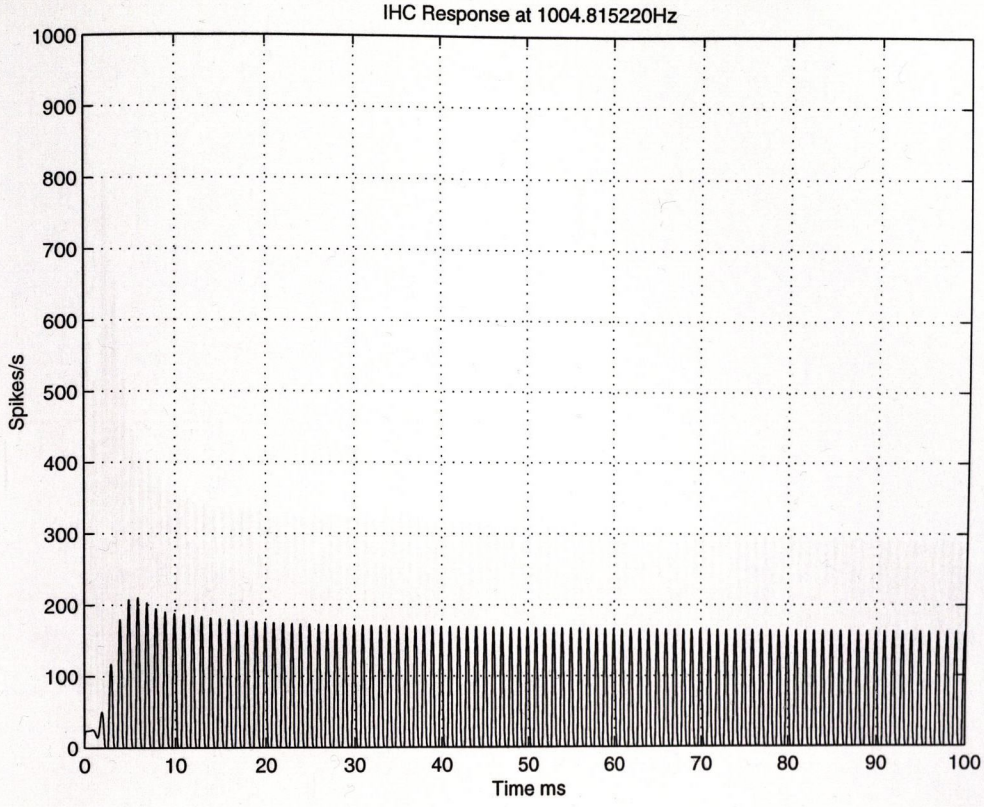


Figure 4.24a Tone burst response of the auditory model at 30dBSPL (Model 1).

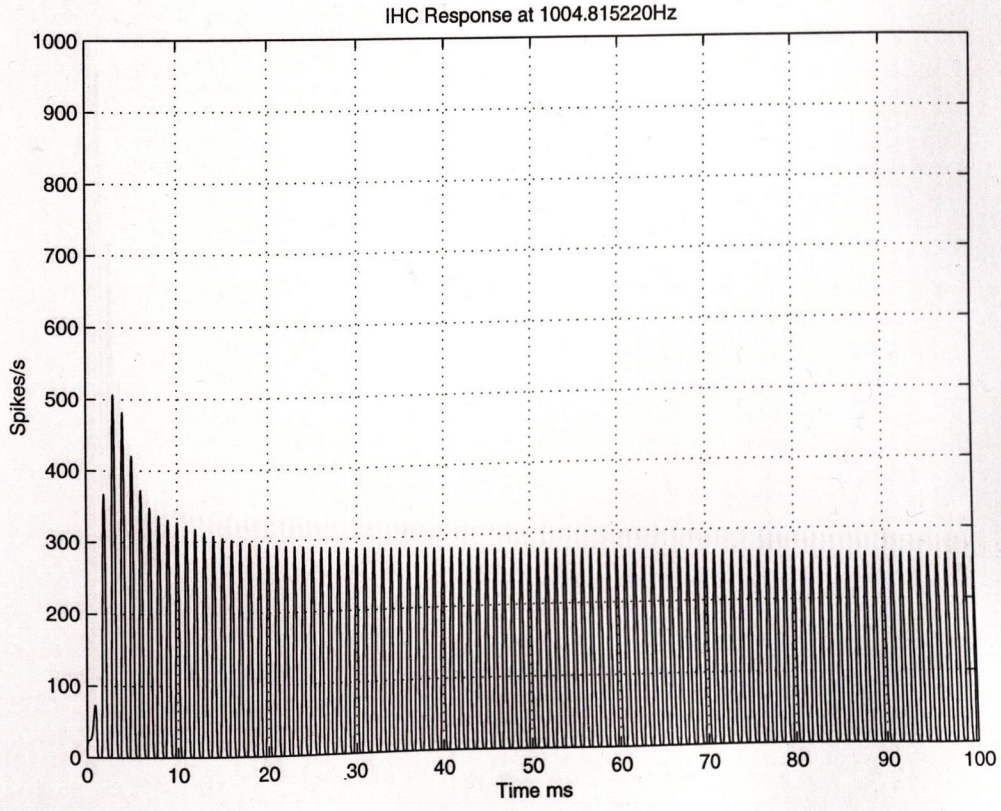


Figure 4.24b Tone burst response of the auditory model at 60dBSPL (Model 1).

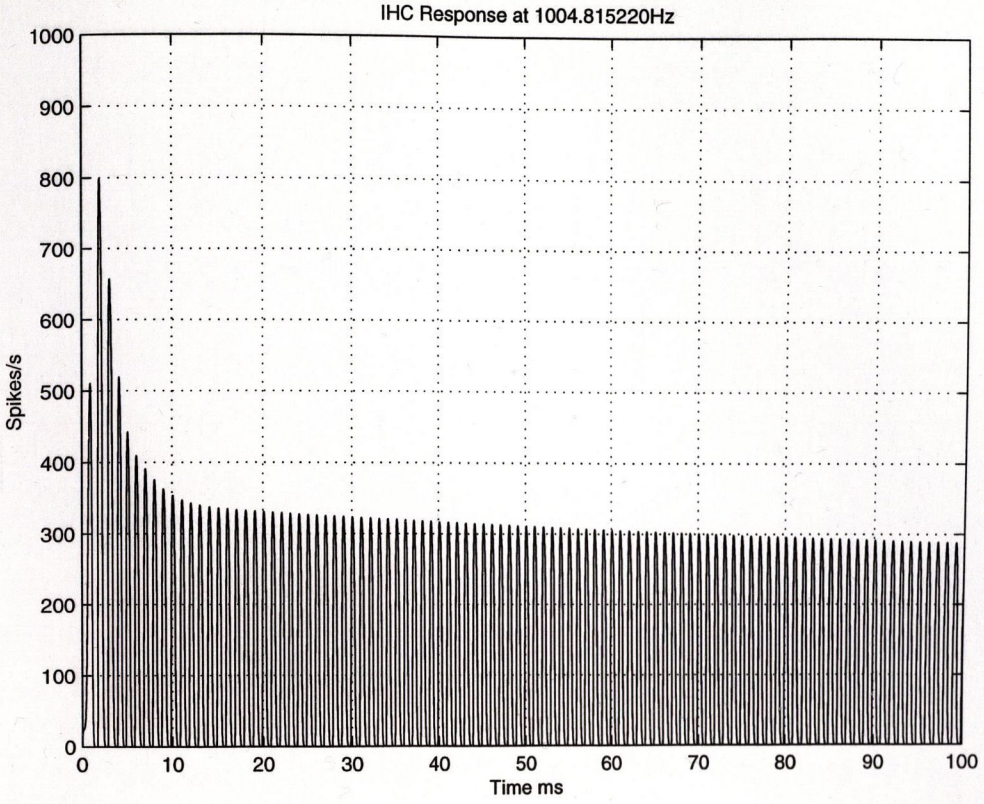


Figure 4.24c Tone burst response of the auditory model at 90dBSPL (Model 1).

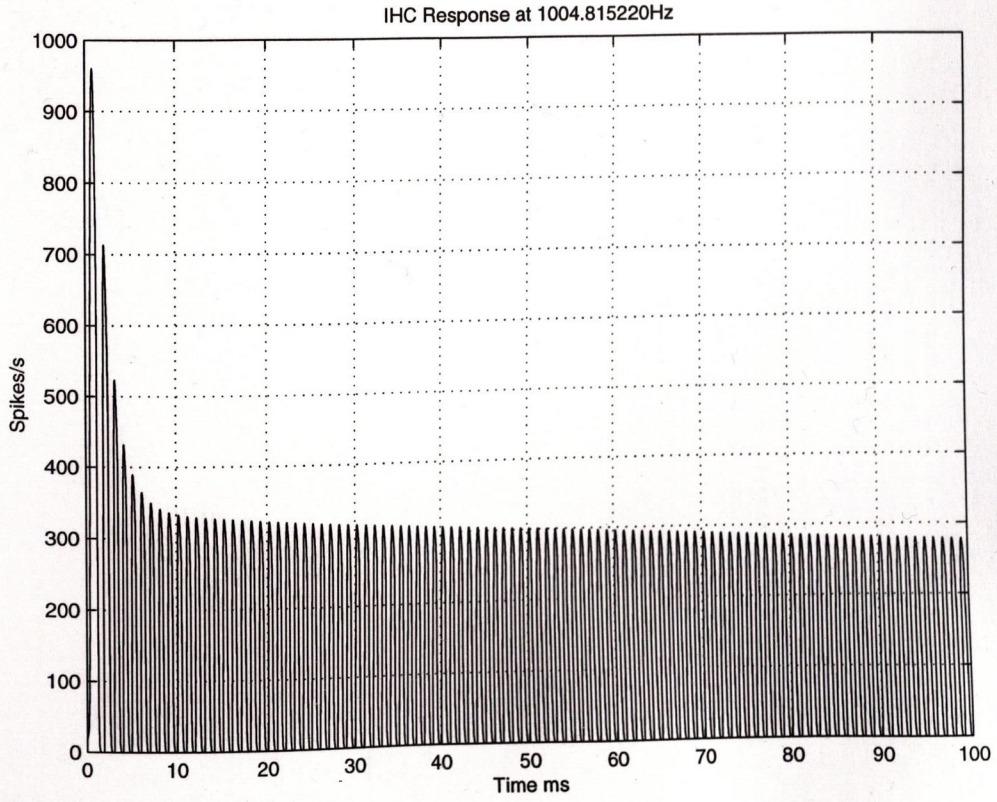


Figure 4.24d Tone burst response of the auditory model at 120dBSPL (Model 1).

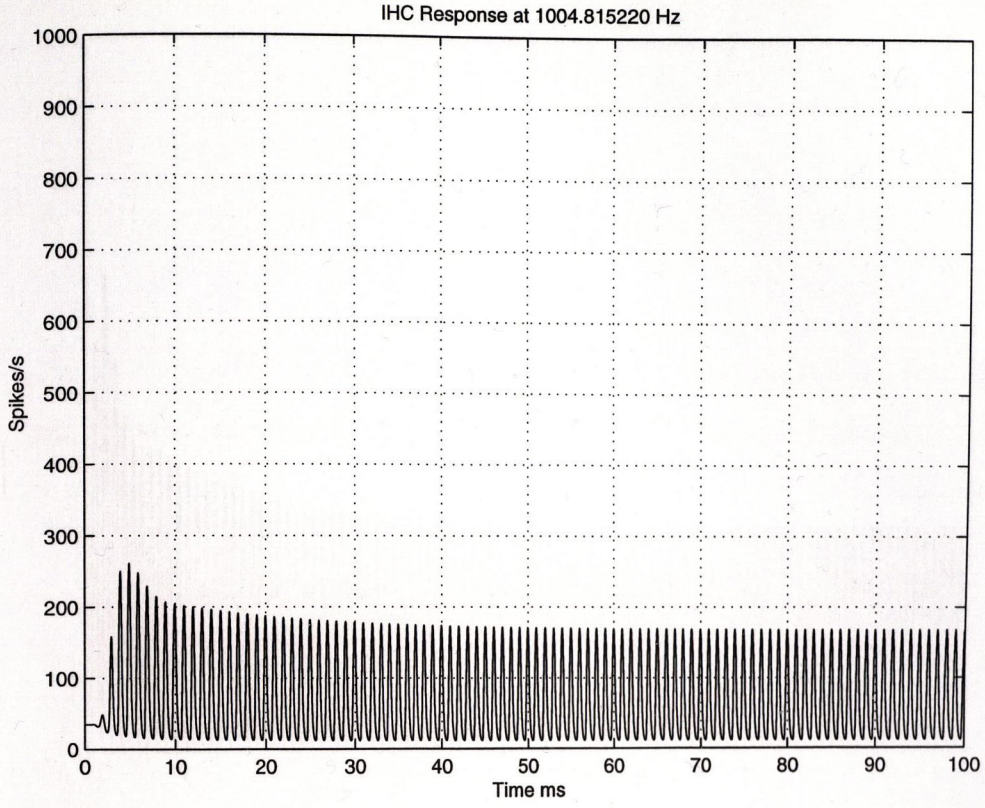


Figure 4.25a Tone burst response of the auditory model at 30dBSPL (Model 2).

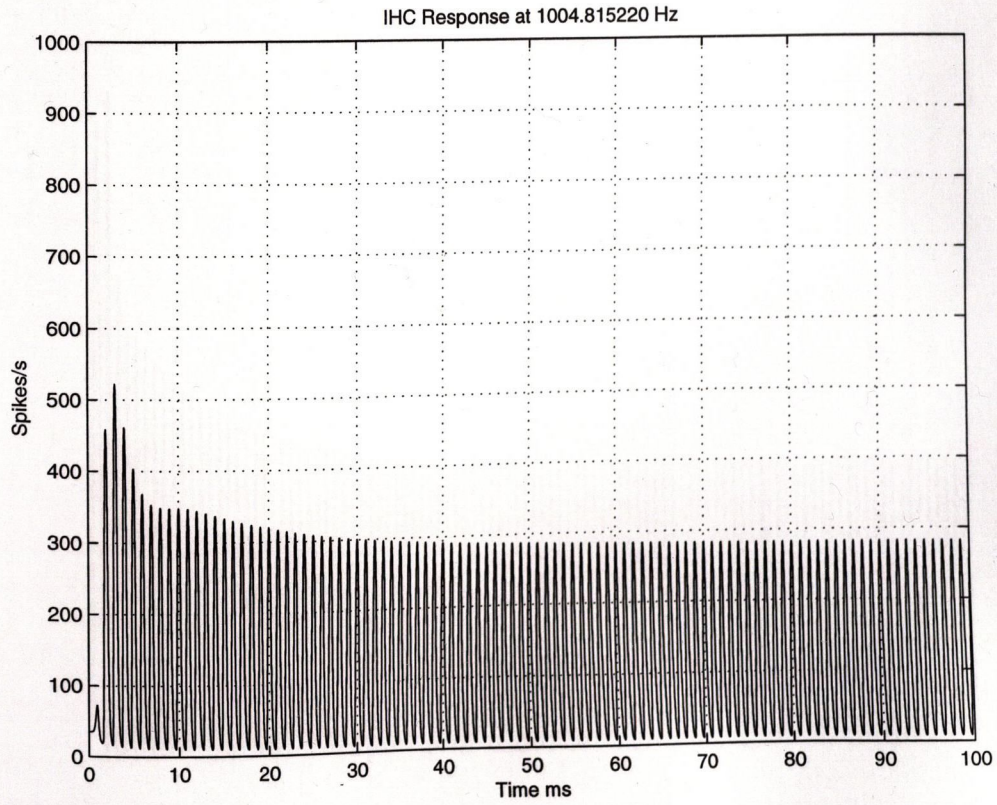


Figure 4.25b Tone burst response of the auditory model at 60dBSPL (Model 2).

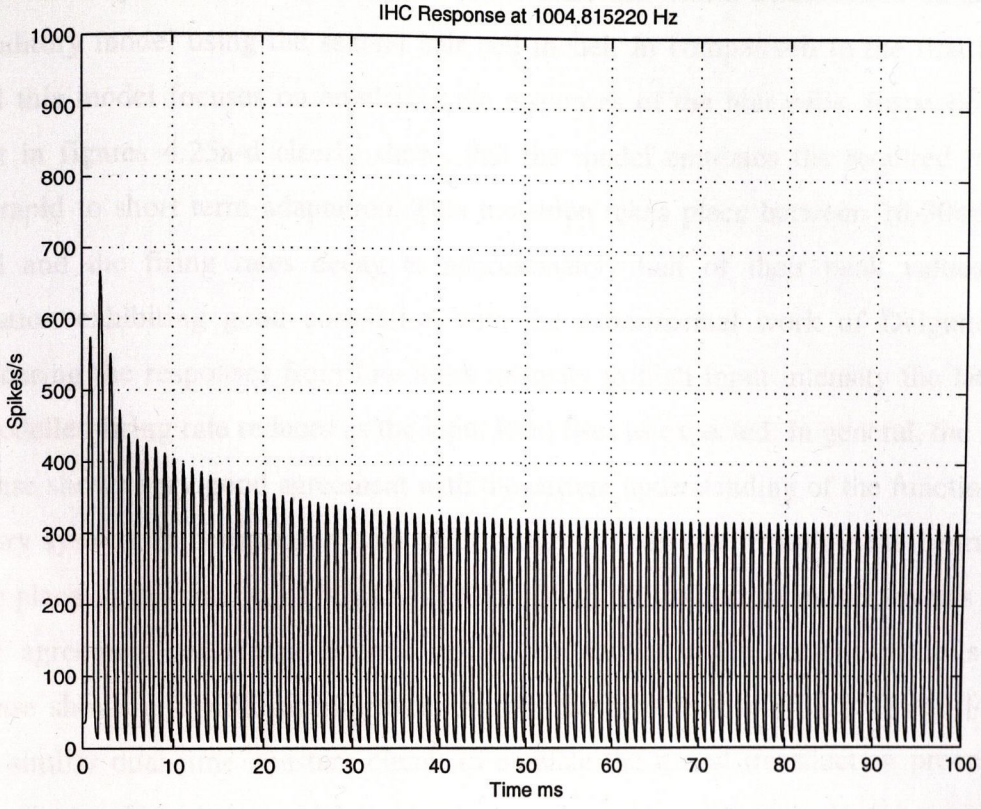


Figure 4.25c Tone burst response of the auditory model at 90dBSPL (Model 2).

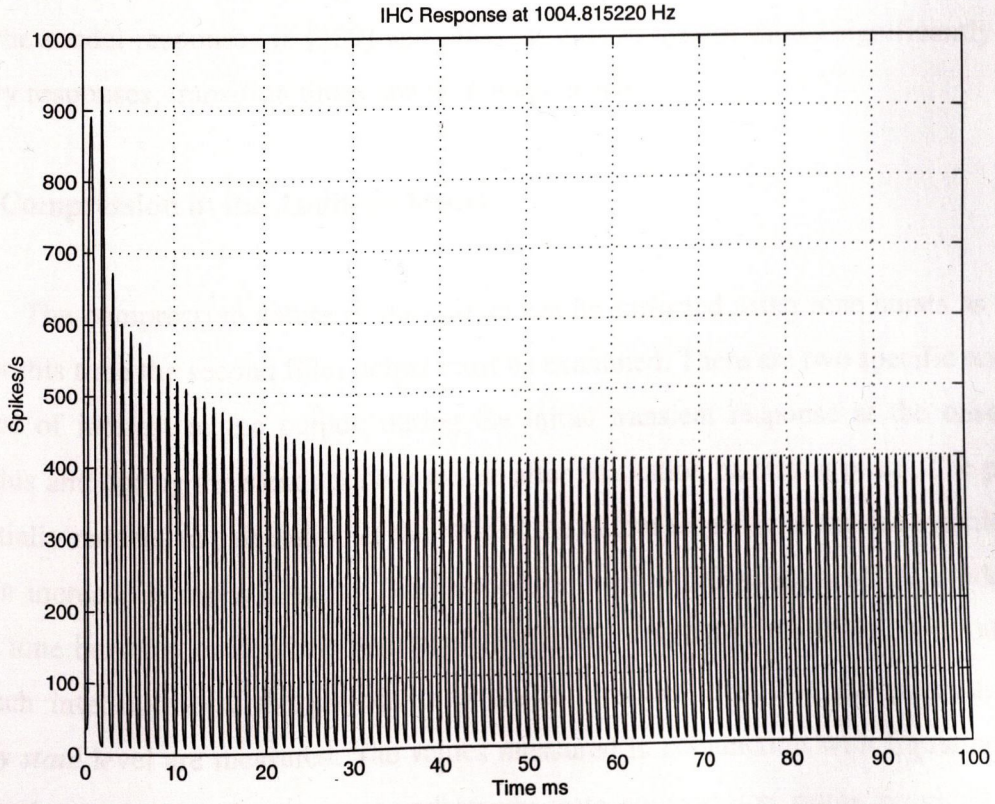


Figure 4.25d Tone burst response of the auditory model at 120dBSPL (Model 2).

The responses shown in Figures 4.25a-d are the neural transduction outputs from the auditory model using the second hair cell model. In comparison to the first hair cell model this model focuses on emulating the responses of the hair cells. Inspection of the output in figures 4.25a-d clearly shows that the model emulates the required transition from rapid to short term adaptation. This transition takes place between 20-30ms in this model and the firing rates decay to approximately half of their peak values during adaptation exhibiting good correlation with the experimental work of Delgutte [137]. Considering the responses from low input intensity to high input intensity the latency of the modelled firing rate reduces as the input level rises as expected. In general, the model's response shows very good agreement with the current understanding of the function of the auditory system. In comparison with other auditory models that emulate the neural firing taking place in the inner ear [78], [52], [49], [10], [139] the current model reveals a much closer agreement with experimental results. In particular the current model's neural response shows a significant improvement over the responses in [78], [52] and [49] that use a similar dual-time constant circuit to emulate the neural transduction process. The peak spike levels in these models are too high and transition times too short; however, they are not at unreasonable levels. This indicates that the modelled basilar membrane motion and the modified dual time constant transduction circuit better emulates the processes taking place within the auditory system than other similar representations. In comparison with the model responses in [139] and [10], the current system shows significantly better latency responses, transition times and peak spike levels.

4.6.4 Compression in the Auditory Model

The compressive nature of the system can be extracted using tone bursts as before except this time the second filter output must be examined. There are two specific temporal regions of interest in the output; during the initial transient response at the onset of a stimulus and during the steady state response after adaptation has taken place. The process essentially reduces to calculation of the ratio of the output response of the second filter to a known increase in input intensity. The procedure involves stimulation of the model with 1kHz tone bursts at each of the four levels 30dBSPL, 60dBSPL, 90dBSPL & 120dBSPL. At each intensity level the pre-adaptation *peak* transient level and the post-adaptation *steady state* level are measured. The values measured in conjunction with Equation (4.36) are used to calculate the transient and steady state compression ratios provided by the system.

$$\text{Compression Ratio} = \frac{\Delta Ip}{Pj} \qquad \Delta Ip = I_2 dB - I_1 dB \qquad Pj = 20\log\left(\frac{P_2}{P_1}\right) \tag{4.36}$$

where ΔI_P is the relative increase in input intensity in dB and P_j is the resulting relative increase in response of the second filter output in dB. Table 4.4 identifies the measured second filter outputs and the calculated compression generated in the model.

Tone Burst Level dBSPL	Transient Response	Transient Compression	Steady State Response	Steady State Compression
30	0.0504		0.0521	
60	0.2197	2.35:1	0.2153	2.43:1
90	0.9793	2.31:1	0.9185	2.38:1
120	5.8376	1.93:1	4.5170	2.17:1

Table 4.4 Measured second filter outputs and resulting compression.

Considering the results in Table 4.4, the compression ratio is calculated for three different jumps in input stimulus level for both the transient and steady state cases. The steady state compression ratio required is 2.5:1 and the results clearly show that for normal intensity levels the model generates this compression to within 4% of the required level. When the intensity level heads towards the threshold of pain the model loses its ability to compress the high level signals as would be expected in the auditory system. The compression within the auditory system during the onset of an input is lower than the steady state compression since the auditory system cannot instantaneously compress the stimuli it receives. It takes as long as the adaptation time for suitable compression to build up within the auditory system. The transient compression results from the model replicating this phenomenon and also indicates a reduction of the instantaneous compression with increased stimulus level showing significant reduction at high levels. In general the compressive behaviour of the model exhibits very good correlation with the known responses of the auditory system [119], [140] and [141]. In particular it shows a reduced compressive behaviour between transient and steady state modes of operation, loss of compression at high intensity levels and reduced transient compressive behaviour with rising input intensity.

Many older models have a tendency to underestimate the importance of the compressive nature of the auditory system, consequently newer models show a greater emphasis on this behaviour [52], [78], [55], [42] and [20]. In general the different basilar membrane representations can be considered to build up their compression through manipulation of their emulated local segment characteristic impedance relative to the impedance of the entire structure. Naturally, transmission line based models can relatively easily incorporate this feature given their approach; however, of greater interest is the compression generated in filter based and wave digital filter based approaches considering the form of the current model. In particular, [52] and its subsequently improved versions [49] and [78] show the current model in a very favourable light.

Level	Kates [52]		O'Reilly [78]		Current Model	
dBSPL	Transient	Steady State	Transient	Steady State	Transient	Steady State
30 => 60	2.1:1	2.3:1	2.12:1	2.45:1	2.35:1	2.43:1
60 => 90	1.9:1	2.3:1	2.02:1	2.30:1	2.31:1	2.38:1

Table 4.5 Compression ratios in various models.

The steady state compression generated in two of these models is shown in Table 4.5 and clearly indicates the improved performance of the current model. It is also worth noting that these models are only specified up to 90dBSPL because they are not capable of operation across the complete auditory range. This is because their compression systems are not capable of dealing with higher intensity levels. The current model is capable of dealing with high level stimuli up to 120dBSPL.

Considering the wave digital filter implementation of [42], it specifies a level dependent compression of 43dB over the full range of inputs, which would indicate a compression ratio in the region of 2.79:1. This is higher than the values reported in [142]; however, this seems to be a compromise required to cover the complete auditory range. This appears to be a general feature of models that fundamentally model the physiological structure of the basilar membrane and arises because of approximations that have to be made in developing these models.

In overall comparison with other auditory models, the current system functions very favourably and in many cases indicates a significant improvement. Modelling of the compressive nature of the auditory system is of particular importance in this work as it is

the loss of this dynamic range compression that causes the non-linear attenuation of sound in the time domain.

4.6.5 Ramped and Damped Responses of the Auditory Model

When considering the ramped and damped response of the system, this deals with stimuli that consist of one main peak whose width broadens as the half-life of their decay function decreases. Of particular importance is the fact that the Fourier energy spectra of damped and ramped sounds with the same half-life are identical. Realistic auditory models should show in their auditory spectra that the width of the spectral peak of a damped sinusoid is narrower than that of a ramped sinusoid for half-lives in the range of 2-16ms and are more like the auditory spectrum of an unmodulated sinusoid [143]. The Fourier energy spectrum is often used as a sufficient model of auditory perception for these types of sounds when comparing the pitch or loudness of various stimuli; however, such models are not sufficient as a representation of sound quality because the Fourier energy spectra of damped and ramped sinusoids are identical. These simplified models provide no basis for distinguishing between pairs of damped and ramped sinusoids or explaining the pronounced difference in sound quality that is heard [143]. When considering sensorineural impairments such capabilities are important particularly when dealing with frequency smearing and recruitment.

In terms of auditory models, time domain models employing linear auditory filter banks to perform the spectral analysis and generating an excitation pattern based on that filtered output will be unable to explain the timbre of sounds with asymmetric temporal envelopes because the excitation produced will be identical for these types of sounds. i.e. the excitation resulting from ramped and damped sinusoids with the same half-lives will be identical. Physiological descriptions of cochlear processing make it clear that spectral analysis performed in the cochlea is accompanied by amplitude compression, temporal adaptation and frequency suppression, features all of which can be found within the current model.

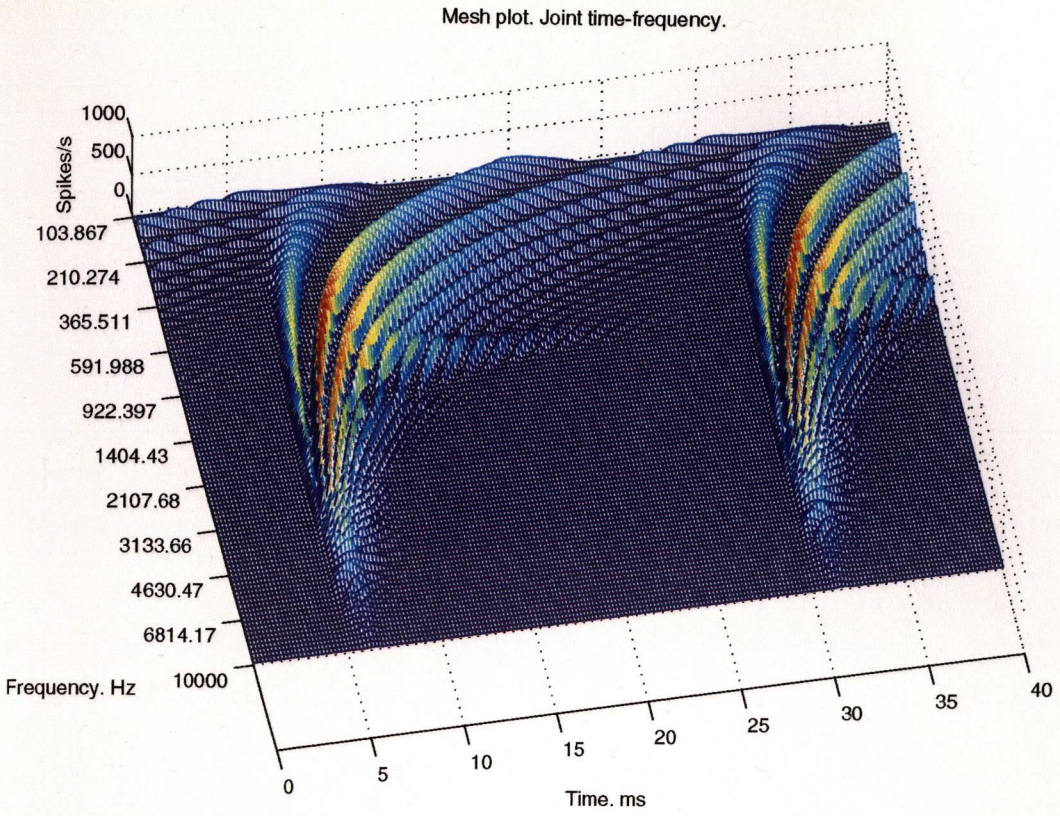
The forms of the input stimuli used are the same as those in [143] to allow direct comparison of the results produced by each model. The general form for one cycle of the damped and ramped stimuli are given by.

$$damp(t) = A * \exp[c * t/hl] * \sin[\omega t] \qquad (0 \le t \le T) \qquad (4.37)$$

$$ramp(t) = A * \exp[c * (T-t)/hl] * \sin[\omega(T-t)] \qquad (0 \le t \le T) \qquad (4.38)$$

where ω is 2π times the carrier frequency (800Hz), A is the starting amplitude, hl is the half-life of the damped and ramped sinusoid, c has a value –0.693147 which reduces the envelope down to $A/2$ in hl milliseconds and T is the repetition period. As the half-life of the stimuli decreases, the energy and the loudness also decrease; so to maintain the stimuli at roughly the same loudness the starting amplitude is increased by the square root of 2 each time the half-life is decreased by a factor of 2 as specified in [143].

The responses of the auditory model to damped and ramped stimuli are presented in the form of joint time-frequency neural firing rate plots, with the firing rate on the z-axis. Figures 4.26a, 4.26b & 4.26c display the response of the model to damped sinusoids with half lives 1ms, 4ms and 16ms, and Figures 4.27a, 4.27b & 4.27c display the response of the model to ramped sinusoids with half lives 1ms, 4ms and 16ms. Each Figure contains two plots to give the clearest possible view of the complex data. The uppermost plots in each figure show the data from an elevated angle whereas the lowermost plots show a top down view of the data. The colour coding goes from the lowest firing level being dark blue to the highest firing level being bright red. A commentary on these results can be found after presentation of all six sets of figures to allow best viewing by the reader.



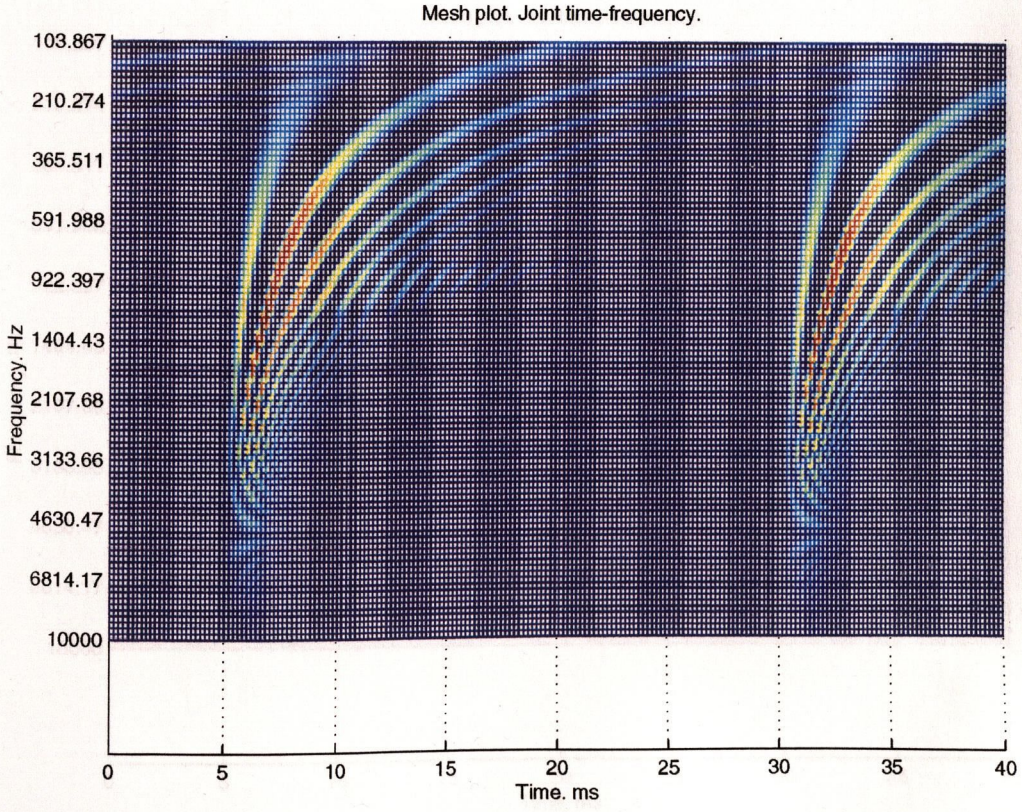


Figure 4.26a Auditory model response for a damped sinusoid with half-life = 1ms.

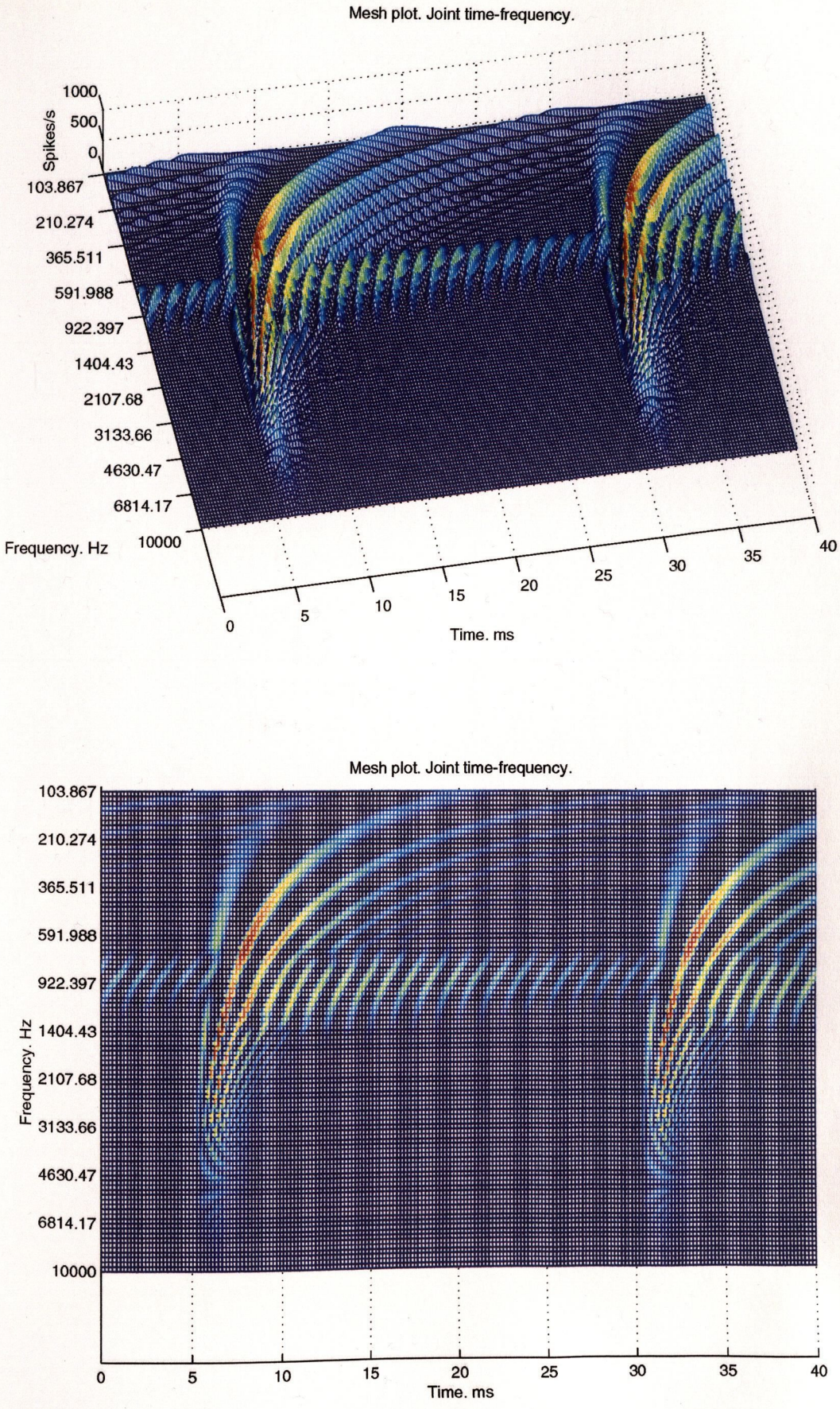
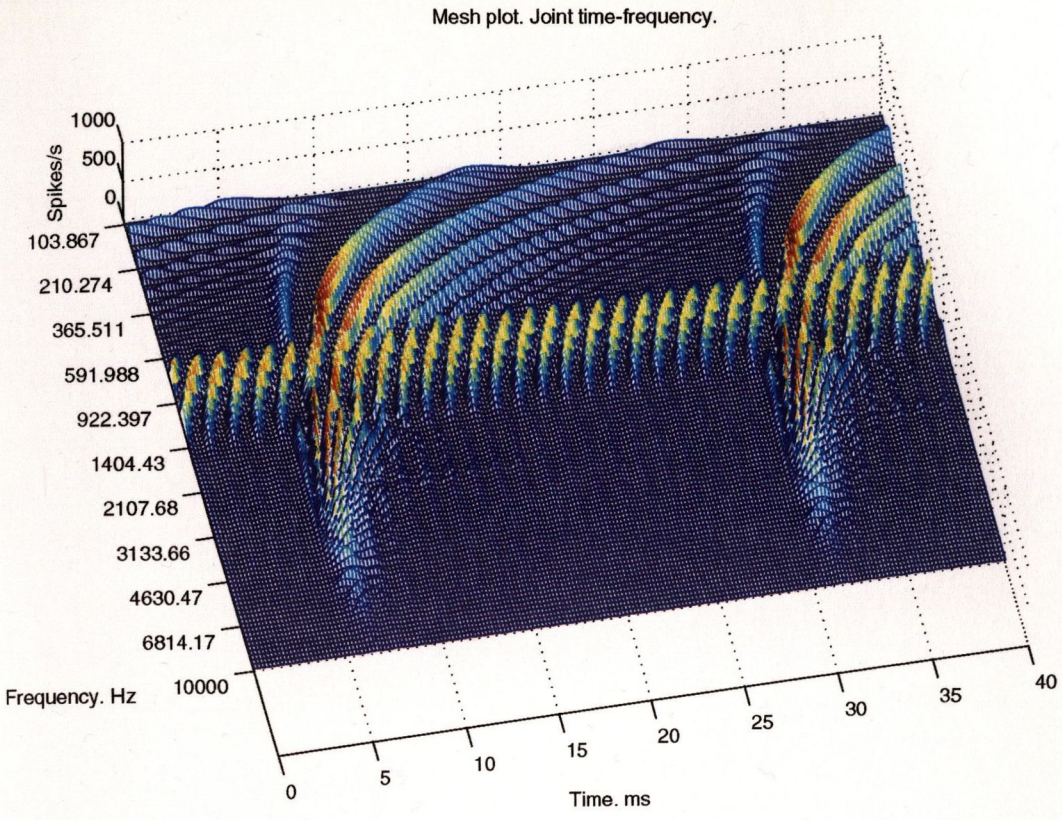


Figure 4.26b Auditory model response for a damped sinusoid with half-life = 4ms.



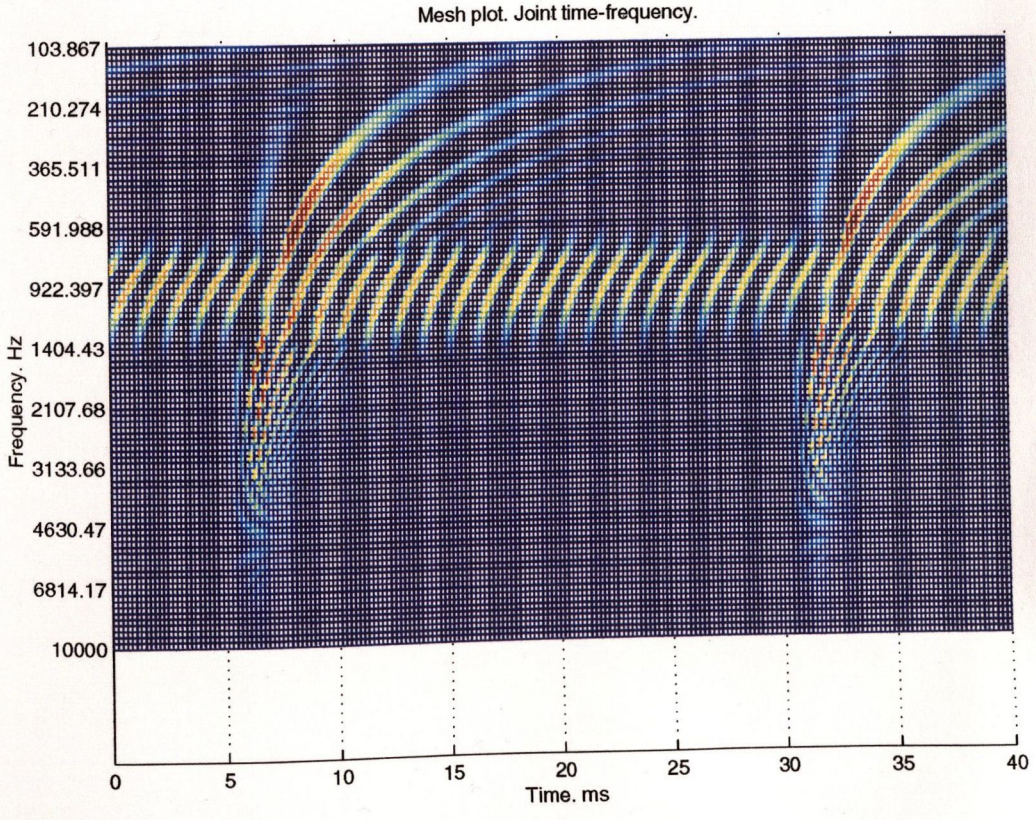


Figure 4.26c Auditory model response for a damped sinusoid with half-life = 16ms.

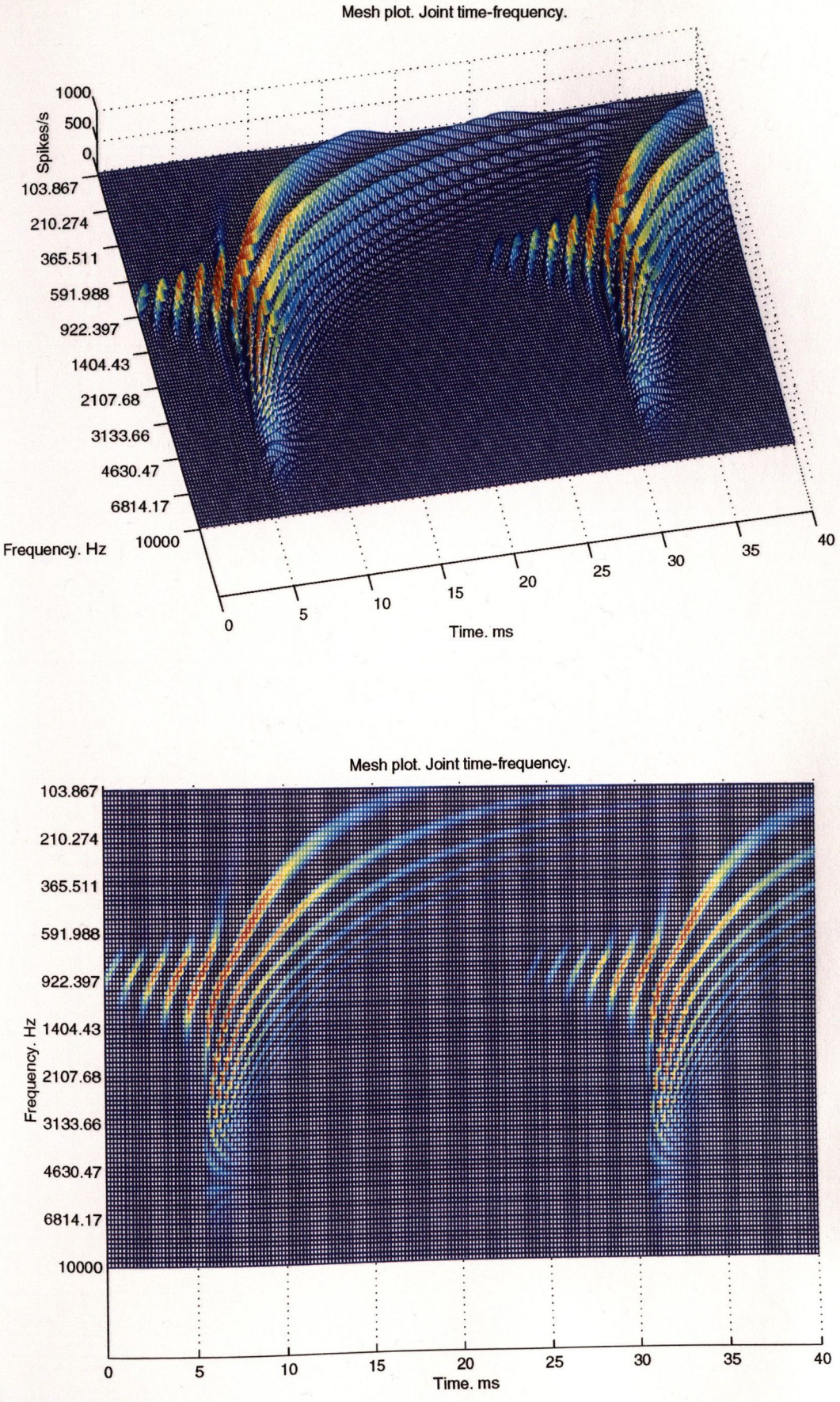
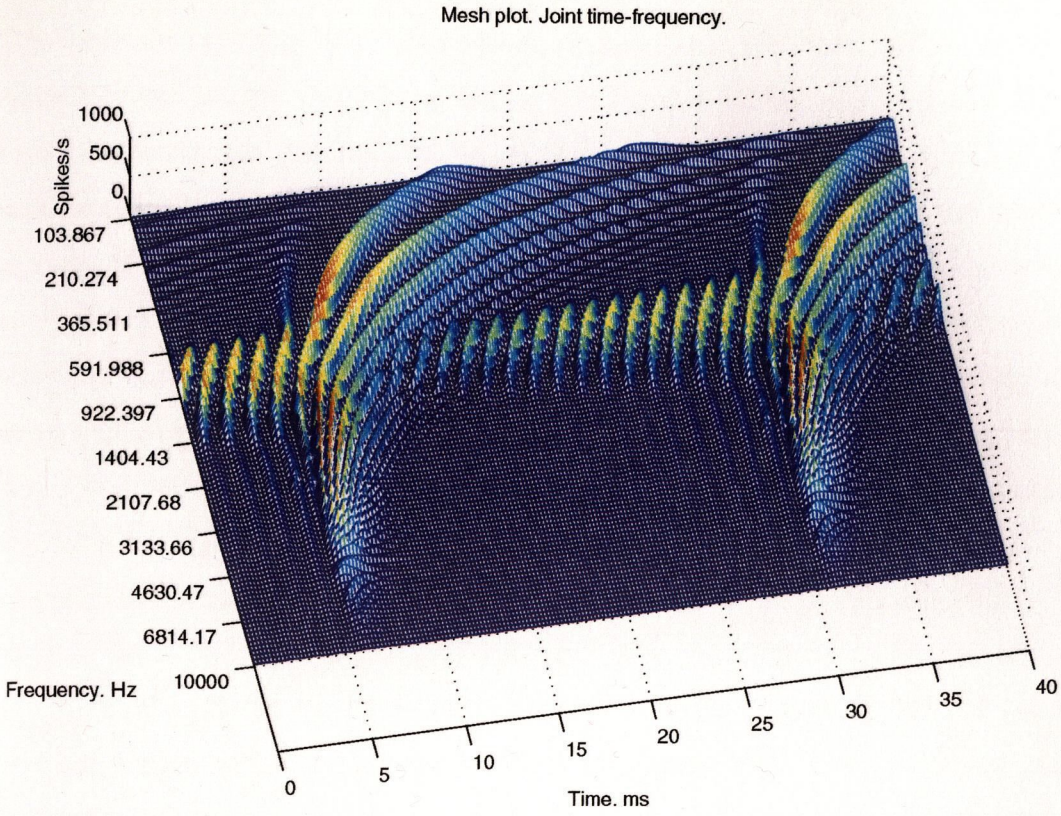


Figure 4.27a Auditory model response for a ramped sinusoid with half-life = 1ms.



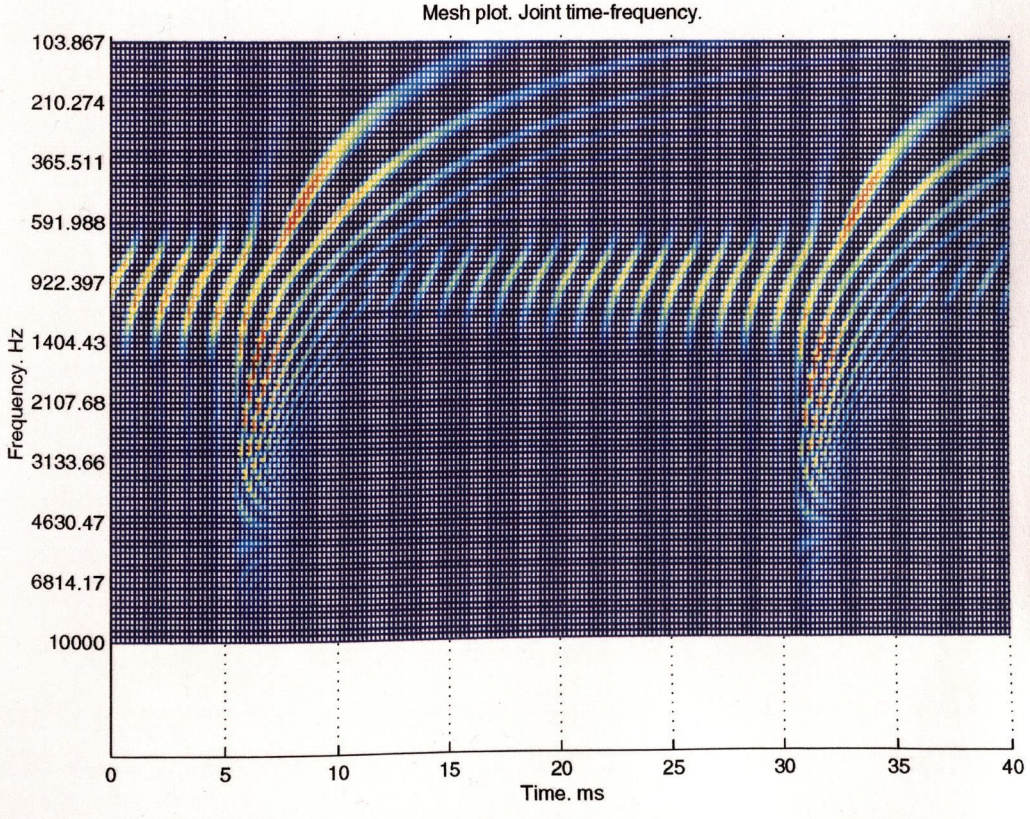


Figure 4.27b Auditory model response for a ramped sinusoid with half-life = 4ms.

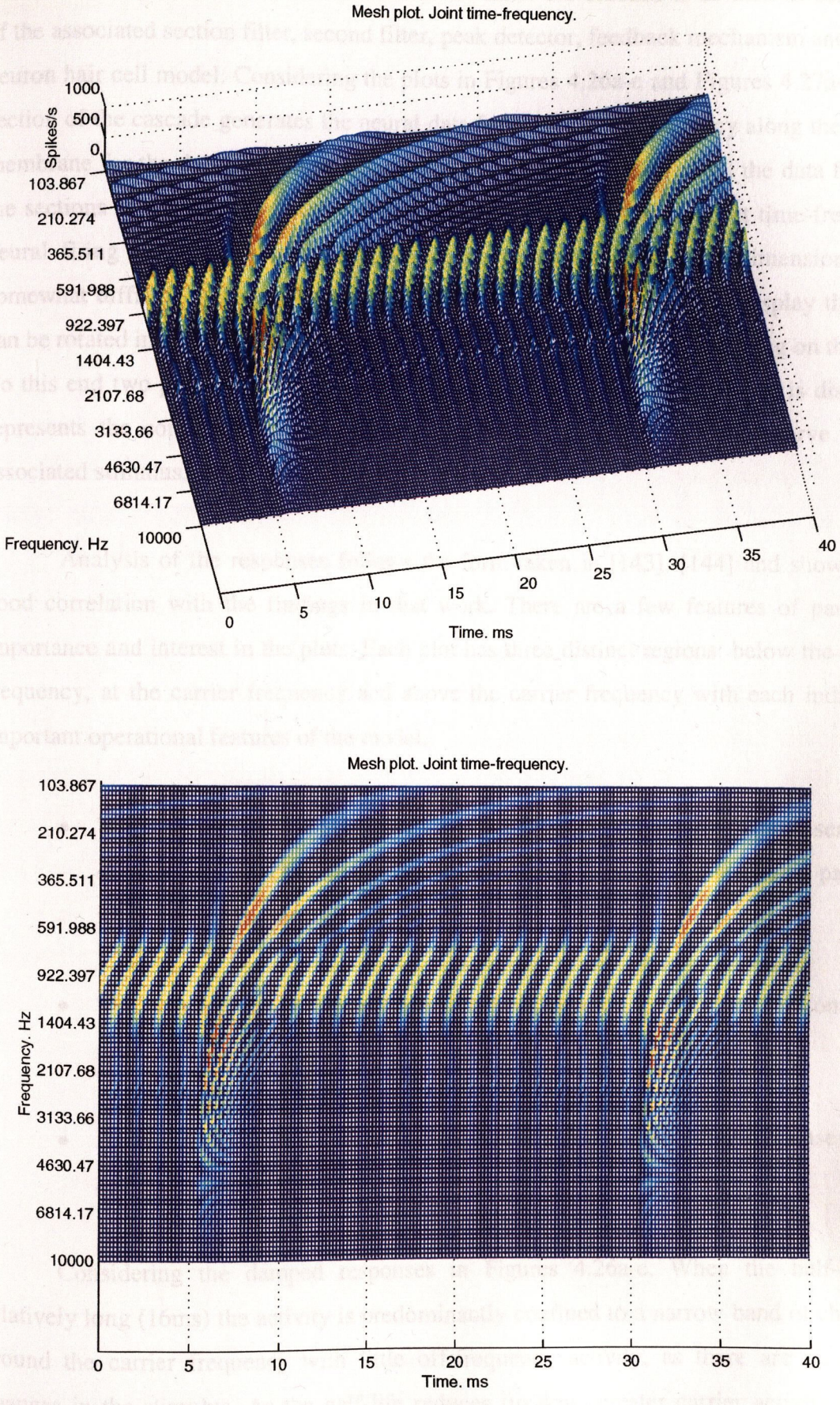


Figure 4.27c Auditory model response for a ramped sinusoid with half-life = 16ms.

To simplify matters, each section of the inner ear cascade is defined as consisting of the associated section filter, second filter, peak detector, feedback mechanism and multi-neuron hair cell model. Considering the plots in Figures 4.26a-c and Figures 4.27a-c, each section of the cascade generates the neural data for each centre frequency along the basilar membrane for the duration of the input stimulus. The combination of all the data from all the sections in the cascade is the source of the data displayed in the joint time-frequency neural firing rate plots shown. The generated results are thus three dimensional and somewhat difficult to interpret when displayed on paper; on a computer display the plots can be rotated in three dimensions to achieve the best viewing angle depending on the data. To this end two plots have been provided for each set of data. The data that is displayed represents the complex pattern of information generated on the auditory nerve by the associated stimulus.

Analysis of the responses follows the form taken in [143], [144] and shows very good correlation with the findings in that work. There are a few features of particular importance and interest in the plots. Each plot has three distinct regions: below the carrier frequency, at the carrier frequency and above the carrier frequency with each indicating important operational features of the model.

- The low frequency region shows that the adaptation sharpens the representation of individual cycles of the basilar membrane motion in the neural patterns, indicating phase locking at low frequency.

- The high frequency region shows that the adaptation reduces the duration of the impulse response relative to that in the basilar membrane motion.

- At the carrier frequency the transduction process enhances the response to the carrier frequency relative to the system impulse response.

Considering the damped responses in Figures 4.26a-c. When the half-life is relatively long (16ms) the activity is predominantly confined to a narrow band of channels around the carrier frequency with little off-frequency activity, as there are no abrupt changes in the stimulus. As the half-life reduces (to 4ms) greater carrier activity can be seen towards the start of the cycle, activity which spreads to higher and lower channels. Towards the end of the cycle this activity reduces in the channels surrounding the carrier as

the stimulus level drops. The resulting abrupt changes in amplitude at the start of the cycle cause stronger off-frequency activity. Further reducing the half-life to 1ms causes the carrier activity to be briefer and broader.

Comparing the ramped responses in Figures 4.27a-c to those of the damped responses, it can be seen that the ramped stimulus activates a broader frequency region around the carrier frequency for all half-lives as required. Examining the off-frequency activation of channels around the carrier frequency it can be shown that the average value over each cycle is greater for ramped sounds than damped sounds when measured as "*the magnitude of the amplitude difference between successive maximum and maxima in the fine structure of the wave*" [143]. The reason for this is that ramped and damped stimuli drive off-frequency filters differently. When a damped stimulus with carrier frequency 800Hz activates a filter of higher or lower centre frequency, it does so with a large pulse of energy resulting in a large response in the filter. However, the filter will try and ring at it's centre frequency. Over several cycles the ringing energy builds up (at the off-frequency centre frequency) while the stimulus level decreases. Since the ringing energy and the stimulus are associated with two different frequencies they eventually drift out of phase, partially cancelling the off-frequency filter's output. In comparison when the ramped stimulus activates off-frequency filters it does so with much less energy. Over several cycles the ringing energy builds up (at the off-frequency centre frequency) however the stimulus level is increased, therefore the ringing energy is small compared to the stimulus level. Even though the response in the off-frequency filter drifts out of phase with the stimulus, the ringing energy only causes negligible cancellation of the off-frequency filter's output. In a wide range of off-frequency channels the ramped sinusoid therefore generates greater activity than damped sinusoids showing very good correlation with the recent work of Patterson in [143] and [144].

The overall result provided by these simulations is that the frequency range occupied by a damped sinusoid is narrower than that occupied by a ramped sinusoid for half-lives in the range of 2-16ms and that the response of the model to such damped sinusoids is more like the response to an unmodulated sinusoid. Strobed temporal integration can be used to convert the neural activity data into a stabilised auditory image that numerically proves this point. This indicates that the auditory model can discriminate between damped and ramped sinusoids just as can be done in the biological system. An important point to note is that experimental results show that listeners presented with

damped and ramped sinusoids *consistently choose the ramped sinusoid as the stimulus with the stronger sinusoidal quality* [143]. This seems to be in contradiction with the model results. However analysis of the stabilised time-interval from the strobed temporal integration of the neural activity data

> "*reveal that, in the portion of the cycle away from the transient, the ramped sinusoids produce proportionately more time intervals at the carrier period than do the damped sinusoids. Damped and ramped sinusoids produce the same number of carrier periods in the carrier channel but, away from the carrier channel, the ramped sinusoid produces more carrier periods per channel. It also produces carrier periods in a wider range of channels.*" [144].

A more complete treatment of these effects can be found in [144]. However, it suffices to show that the response of the current model corresponds very closely to those in [143], [144] and that the model is capable of discriminating between damped and ramped sinusoids.

4.6.6 Response of the Auditory Model to Synthetic Speech

When considering stimulation of the auditory model with synthetic speech we are dealing with stimulation of the model using synthesised syllables more so than actual words or sentences. The reason for looking at syllables is that if the model can generate suitable responses for the fundamental building blocks of speech then it should in theory be able to generate the required responses for words and for connected speech. The sheer complexity of words and speech make it difficult to ascertain the response of the model because of the sheer quantity of data produced, the time required to generate this data and the processing power needed to run the model with such complex inputs. To this end the response of the model to the synthesised syllable **/da/** is presented in this section allowing comparison with the responses in [52]. The following description of the stimulus comes from [52].

"*The first three formants of the vowel portion of the syllable are at 0.7, 1.2 and 2.4kHz, respectively. The consonant portion of the syllable is represented by formant transitions that occur during the first 50ms, with the starting frequencies for the first three formants being 0.5, 1.6 and 2.8kHz, respectively. The formant frequencies are held*

constant during the second 50ms of the stimulus, and the total stimulus is 100ms. The fundamental frequency is 120Hz during the first 50ms of the stimulus and drops to 116Hz during the second 50ms. The average power in the signal is about 65-dBSPL at the onset of the stimulus and increases to about 69-dBSPL at the end." [52].

Figure 4.28 shows the first 25ms of the neural response from [52] to this syllable.

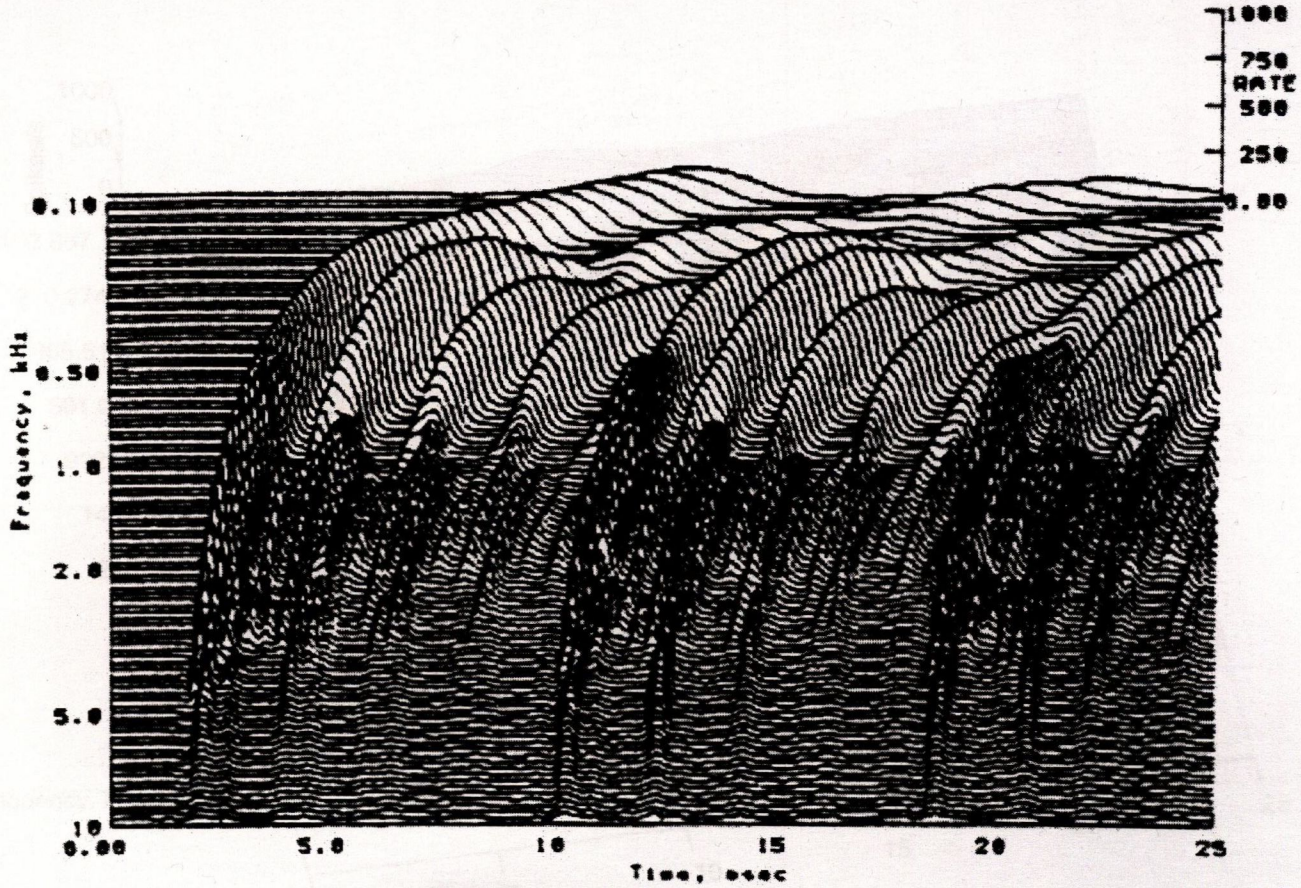


Figure 4.28 Neural response in [52] to synthesised syllable /da/.

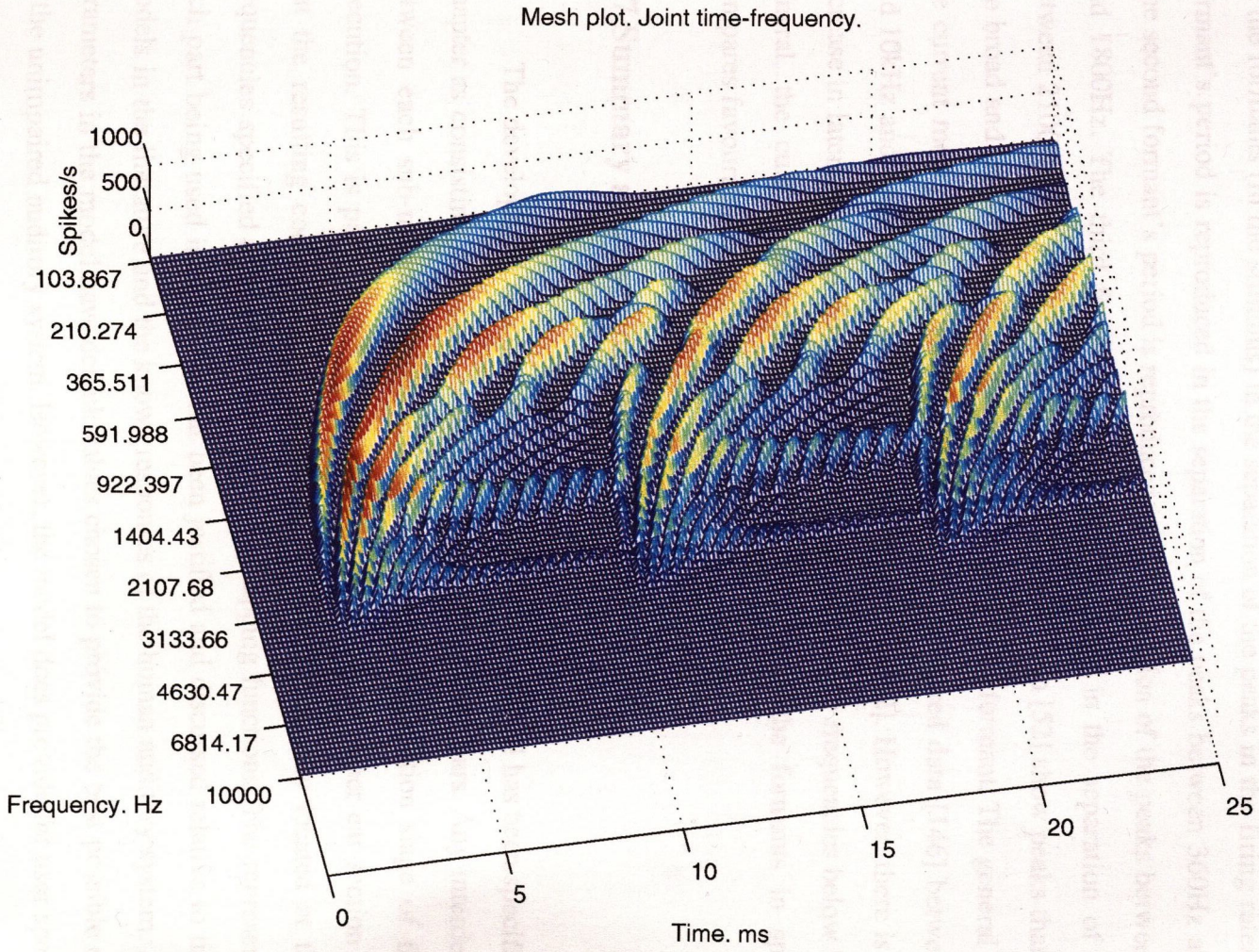


Figure 4.29 Neural response of the auditory model to the synthesised syllable /da/.

Figure 4.29 shows the first 25ms of the neural response of the current auditory model when it is stimulated using the syllable **/da/** conforming to the specification laid out above and generated using a Klatt synthesiser [145] as proposed in [52]. Considering the information presented in Figure 4.29, the response of the current model shows better overall behaviour and phase locking in the firing rates with the data of [146] than the response presented in Figure 4.28. Similarly the current model shows better reproduction of the formants (in the stimulus) in the separation of the peaks in the firing rates. The first formant's period is reproduced in the separation of the peaks between 360Hz and 700Hz. The second formant's period is reproduced in the separation of the peaks between 1200Hz and 1800Hz. The third formant's period is reproduced in the separation of the peaks between 2100Hz and 3100Hz. In comparison, the results in [52] show peaks that are much too broad and show only minimal reproduction of the third formant. The general latency in the current model shows good correlation with the measured data [146] between 360Hz and 10kHz and shows better latency than the response in [52]. However, there is a relative increase in latency due to the section depth in the cascade for frequencies below 360Hz. In general, the current model shows good reproduction of the formants in speech and compares favourably with the work in [52].

4.7 Summary and conclusions

The development of a time domain digital auditory model has been specified in this Chapter as consisting of models of the outer, middle and inner ears. Any interdependence between each sub-model is accounted for during the initialisation stage of the model execution. This is particularly important when specifying the inner ear section filters so that the resulting cochlear responses of the cascade are suitably located at the centre frequencies specified by the place to frequency mapping functions. The representation of each part being used in the model has been justified and discussed relative to the current models in the literature and the known responses of the human auditory system. All of the parameters in the model have default values chosen to provide the best possible emulation of the unimpaired auditory system. However, the model does provide for user specification of these parameters within limits. The user also has a choice of

- Two different outer-middle ear models
- Two different frequency mapping functions.
- Two different inner hair cell multi-neuron models.

The responses of the model have been extracted using various types of stimuli with a view to comparing the model with the known responses of the human auditory system and the current information in the literature. In particular the following responses have been extracted:

- The responses of the outer and middle ear models.
- The tuning curves of the basilar membrane model.
- The tone burst response of the inner hair cell multi-neuron models.
- The compression ratio generated in the cascade.
- The ramped and damped responses of the model.
- The models response to synthetic speech.

The responses of the auditory model show that it is capable of taking as its input raw sound stimuli and generating as its output the neural firing activity on the neurons leading to the brain. In general, the responses above indicate that each section shows good correlation with the known responses of each section of the human auditory system. The complete auditory model has been show to respond very favourable when stimulated with real world sounds such as damped and ramped tones and synthetic speech. In particular it has shown very good correlation with the responses of the recent work in [143] and [144]. The model has also shown to compare very favourably with other complete auditory models and in many cases provides better emulation of the responses generated by the real auditory system.

In general the primary focus of the model is on the emulation of the inner ear as this region is the site of sensorineural impairment which is of particular interest in this work. However, it is important to include models of the outer and middle ear as they have a profound effect on the signals entering the inner ear and are responsible for the enhanced responses of the auditory system to the frequencies that make up human speech. The importance of the emphasis placed on accurately modelling the inner ear will be made apparent in Chapters 5 and 6 which deal with sensorineural impairment of the auditory system. In particular, these Chapters will look at modelling recruitment and frequency smearing.

Given the selection of auditory models detailed in the literature it is important to identify the specific added benefit provided by the auditory model in this work. Primarily

the auditory model is a complete auditory model in that it takes as its input raw sound data and generates as its output the neural activity on the auditory nerve leading to the brain, just as occurs in the real auditory system. In this manner the model encapsulates the entire auditory process providing a very unique method for analysing the effects changes to this system have on input/output data sets. Ultimately, this permits examination of the effects changes to this system have on the neural activity going to the brain. The types of changes made to the model would be those that occur when impairment is present and the auditory model has been specifically designed to provide maximum access to the parameters necessary to introduce impairments into the model. The difference in neural activity generated by the unimpaired and impaired models permit examination of the effects impairment has on the stimuli presented to the brain. Where this becomes particularly useful is in the development and evaluation of compensation systems to counteract the effects of impairment. The development of a compensation system for a particular impairment would aim to minimise the difference between the neural activity in the impaired and unimpaired models. Evaluation of the performance of various compensation systems is facilitated through comparison of the differences achieved. Considering this, one might ask; why not perform such processes using the available auditory models rather than develop a new model? The answer can best be considered by dividing all auditory models into two groups, those that model parts of the auditory system and those that model the complete system and examining the reasons for their unsuitability.

Models of parts of the auditory system do not lend themselves to the type of analysis mentioned above because of the inputs they take and the outputs they generate. Most models of this type take a sound stimulus as their input, typically avoiding the contributions of the outer and middle ears thus the stimuli presented to the modelled inner ear are not realistic of the stimuli that the inner ear typically receives. The output of these models comes in many varieties depending on the detail incorporated in the model; often they take the form of the basilar membrane velocity or pressure, the perceived loudness or the selective basilar membrane activity. Generally the outputs do not represent the real output of the auditory system, which is the neural activity presented to the brain. These types of models therefore are not suitable for the type of analysis mentioned above. Of particular importance is the fact that these models do not incorporate the level of detail necessary for introducing impairment and thus are unsuitable for developing or evaluating compensation systems.

Of the very limited number of complete auditory models, most of these are transmission line models employing lumped parameter sub-models. Although these types of models typically take sound stimuli as their input, model the complete auditory process and generate the neural activity as their output, it is extremely difficult to introduce impairments into these models. The reason for this is that the lumped parameters do not facilitate isolation of the specific parameters required to model particular impairments. These types of models can be used for introducing global effects but are not very suitable for modelling specific impairments (particularly recruitment and frequency smearing) and therefore are of little use in the development and evaluation of compensation systems. Those models that are not transmission line based, although modelling the complete process, are typically quite generalised in their approach to modelling in that they combine many sub-functions of the auditory process together making it difficult to isolate particular parameters. These models are, however, more suitable for the application of the processes mentioned above than transmission line based models but would require significant alteration to facilitate this process.

In general, the primary reason why most of the current models are not suitable for impairment modelling and compensation system development is because they have not been designed with such an objective in mind. It is unreasonable to expect models designed for one purpose to be suitable for another, without significant modification. In the majority of cases the level of modification that would be required easily justifies the development of a whole new model, based on the specific objectives required, thus producing a more optimal solution.

In conclusion the model has been implemented using two layers, an interface layer and a numeric processing layer. The core numeric processing engine has been written in C++ and the interface and simple processing written in Matlab 5.2.1. The calculation of all coefficients and the processing of all data is carried out by the numeric processing layer using 64-bit floating point precision throughout. The human computer interaction, simple initialisation and display of results have been carried out in Matlab with all data being handled as 64-bit precision floating point data. The model was developed in the Microsoft Windows 98 environment operating on an Intel Pentium Pro 200MHz processor. The model is by no means a real time model and takes approximately 200ms to process 1ms of input data at a sampling rate of 40kHz.

Chapter 5 - Recruitment

5.1 Introduction

The previous Chapter has dealt with modelling the unimpaired auditory system, thus providing a reference from which to work when considering modelling an impaired auditory system. One of the main reasons for modelling the unimpaired auditory system is so that impairments can be introduced, thereby allowing us to gain a better understanding of the perceptual consequences of sensorineural hearing loss and provide a framework in which compensatory functions might be developed for these impairments.

This section deals with a particular sensorineural impairment of the auditory system known as recruitment. The root cause of this impairment is the loss of outer hair cell function within the cochlea [61]. The primary function of the outer hair cells is to interact with the basilar membrane to introduce dynamic range compression of the signal excitation stimulating the inner hair cells. When the outer hair cells are damaged, a loss of dynamic range compression occurs and the resulting dynamic range of the excitation of the inner hair cells is greater than they are capable of coding [147], causing recruitment.

Recruitment can be loosely defined as the non-linear attenuation of sound in the time domain or more specifically, the non-linear loss of dynamic range compression, the result of which is increased thresholds of audibility and abnormally rapid growths of loudness level with increasing intensity. Clearly, it can be seen that the outer hair cells play an important role in loudness and speech coding. Loudness is, however, the subjective perception of sound in terms of intensity and a more objective measurement is required for modelling purposes. The work of Javel [131] uses the idea that loudness correlates with the total spike count elicited in the active neural population, a dependence that has seen considerable discussion in the work of Allen [77], the experimental endeavours of Fletcher and Munson [76] and associated works. Comparing the slopes of the responses for neural growth to perceptual loudness growth shows they are different [131] and hence total spike count cannot be used as an *absolute* measure of loudness but can be used as a measure of the change in relative loudness associated with a change in stimulus level. In simple terms the relative change in loudness level can be deemed equal to the relative change in neural activity in the entire active neural population.

Recruitment can be concisely explained in the framework of neural activity and in conjunction with the diagram in Figure 5.1. In Figure 5.1 the solid curve represents the spike growth in an unimpaired auditory system and the dotted line represents the spike growth in an impaired auditory system. Consider an auditory system suffering from a loss of outer hair cell function at a particular frequency f1 and that this loss translates into an elevation of the threshold of audibility for this frequency of $(x_2 - x_1)dB$. Stimulation of either ear at a level above its threshold will cause spike counts to grow above the spontaneous level. If the stimulus level is increased by a specific amount, ΔI, above threshold in both ears, it causes the spike count in the impaired ear to grow more rapidly than in the unimpaired ear. This means that for the same increase in intensity the impaired ear experiences an abnormally loud stimulus compared to the unimpaired ear. As the stimulus level rises further, at some intensity (x_5) the spike count in both ears will be at the same level and recruitment must end, the loudness in both ears at this level returns to normal. However, the impaired ear will have experienced an abnormally rapid growth in loudness level arising from its reduced dynamic range compression.

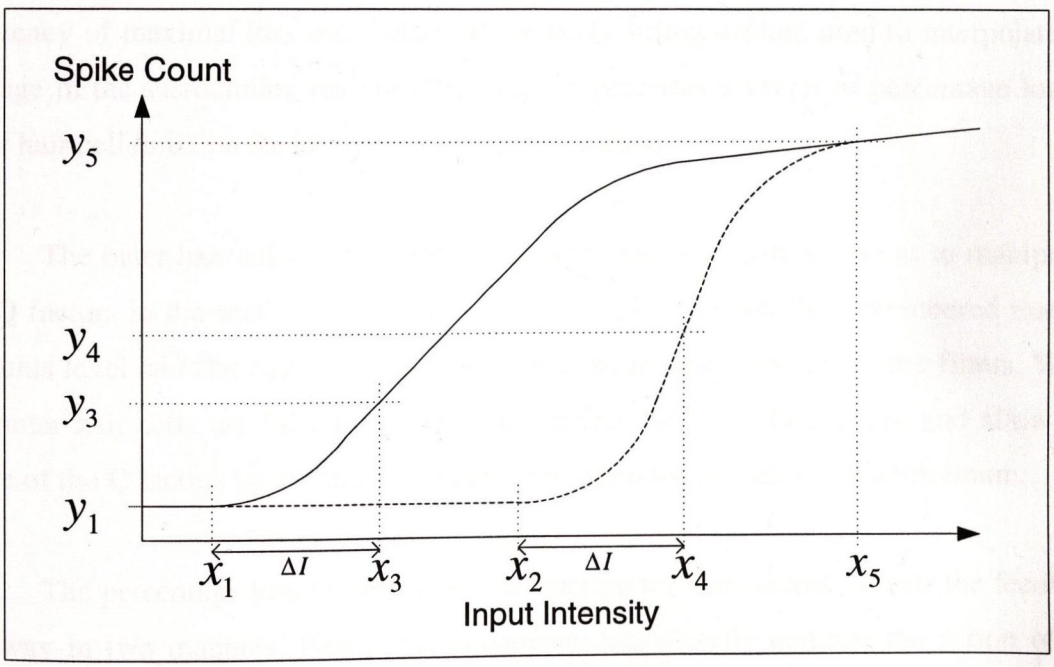


Figure 5.1 Unrecruited and recruited neural activity.

5.2 Modelling Recruitment

The unimpaired model of the auditory system focuses on the processes taking place in the inner ear specifically to facilitate malfunction of the components in the inner ear associated with sensorineural impairments. Incorporation of particular malfunctions in the

model should cause it to manifest the symptoms associated with particular impairments, if the model successfully emulates the functions taking place in the auditory system. In the case of recruitment, malfunction of the outer hair cells is believed to be the cause, therefore outer hair cell damage in the model should cause the effects associated with recruitment.

Considering the sections in the cascade representing locations on the basilar membrane, each section incorporates an outer hair cell emulation that influences the localised response of the basilar membrane at the specific centre frequency of the section. Given this arrangement a percentage loss of outer hair cell function can be specified explicitly for each section and hence for each discrete location on the basilar membrane. This in turn implies that a loss of outer hair cell function can be specified for each of the centre frequencies generated by the frequency mapping functions. It would, however, be unreasonable to specify outer hair cell loss at only the frequency of maximal outer hair cell loss. In general the outer hair cell loss would increase towards the location of maximum loss and decrease away from the location of maximum outer hair cell loss in a relatively continuous manner. With this in mind the percentage outer hair cell loss is specified at the frequency of maximal loss and a cubic spline curve fitting method used to interpolate the damage in the surrounding regions. This process generates a vector of percentage loss of outer hair cell function for all the sections in the cascade.

The outer hair cells effect their action in the feedback pathway so as to manipulate the Q factors in the section and second filters. It must, however, be remembered that the stimulus level *and* the outer hair cells act to manipulate the Q factors of the filters. When the outer hair cells are fully functioning they effect 100% of their action and allow full range of the Q factors up to their maximum level when the stimulus is at a minimum.

The percentage loss of outer hair cell function for each section affects the feedback pathway in two manners. Firstly, the percentage loss directly controls the action of the outer hair cell function by removing the loss percentage from the outer hair cell feedback contribution causing the threshold of audibility in the section to increase by decreasing the Q values. This means that a signal at the normal threshold level will receive reduced gain from the section because of the reduced outer hair cell action. Stimulation of the inner hair cell multi-neuron model will therefore be reduced below the hair cell threshold resulting in only spontaneous firing of the hair cell. Not until the stimulus level reaches a sufficiently high level (above the new threshold), with the reduced gain provided by the section, will

the neural firing rate increase. Secondly, the percentage outer hair cell loss reduces the upper limit on the Q value. This has the effect of reducing the maximum compression the cascade can generate in the defective region and prevents the stimulus level control of the Q value from artificially elevating the section response to compensate for loss of outer hair cell action. At threshold levels of stimulation, the sections provide their maximum gain and the Q control is primarily stimulus level driven. As the stimulus level increases the outer hair cells begin to evoke their compressing action which is reduced due to loss of outer hair cell function. This results in a loss of dynamic range compression in the cascade. Equation (5.1) relates how the level of OHC loss is introduced into the feedback path.

$$Adapt = Adapt * \% \; OHC \; Funtcion \tag{5.1}$$

The main justification for the success of this type of implementation can be considered in terms of Stephen's law, a law that encompasses all power law relationships between the intensity of the physical stimuli and the psychophysical response. In particular, the dependence of loudness on the summation of the cube root of energy [76] is especially important. Given the nature of the gain provided by the cascade the adaptation function turned out to be a *cube* function, see Equation (4.34). This is in agreement with the *cubic root* relationship between loudness and energy [76] because the gain in the sections reduces with increasing intensity and thus the adaptation function in this model should be an inverse cubic root function i.e. a cube function. The actual functions themselves are of little importance as they depend on the models in which they reside; the importance is in the nature of the functions which are both cubic. The model therefore embodies a suitable relationship between physical stimuli and neural responses. Reduction of the outer hair cell contribution in the feedback mechanism combined with the cubic adaptation function will generate a neural output that maps input stimuli to a recruited neural output.

5.3 Simulation of the Auditory Model suffering from Recruitment

The effects of recruitment in the model can be best examined in the responses of the model to various stimuli. In particular, consideration will be given to the tuning curves in the model, the responses of the model to increasing tone levels, to tone bursts and to synthetic speech when the model is suffering from various levels of recruitment. The level of compression in the cascade is also calculated for various levels of recruitment. Each of

these simulations provides important information regarding the effects recruitment has on the auditory system.

5.3.1 Tuning Curves of the Recruited Basilar Membrane Model

Extraction of the tuning curves from the model is performed in the same manner as before, with tuning curves being generated at 1kHz, 2kHz and 5kHz to allow comparison with the unimpaired results. Figure 5.2 shows the tuning curve at 1kHz for 0% and 100% outer hair cell loss overlaid for ease of comparison. Figures 5.3 and 5.4 show the tuning curves for 1kHz, 2kHz and 5kHz for 50% and 100% outer hair cell loss at 1kHz to illustrate the effect outer hair cell loss at one frequency influences the tuning at other frequencies. The numerical data extracted from the unimpaired and impaired tuning curves is summarised in Table 5.1.

Frequency Hz	**Peak BM & TM dB**			**Tip-to-tail BM & TM dB**			**Roll-off BM & TM dB/octave**		
OHC Loss	**0%**	**50%**	**100%**	**0%**	**50%**	**100%**	**0%**	**50%**	**100%**
1004.82	-79.0	-58.9	-38.6	66.7	52.7	30.8	134.5	93.8	74.1
2024.82	-80.7	-74.5	-79.7	73.5	73.3	78.5	141.8	122.6	124.6
5003.83	-73.2	-67.3	-72.3	77.2	77.2	82.1	133.5	133.2	135.7

Table 5.1 Analysis of the tuning curves for 0%, 50% and 100% outer hair cell loss.

The effects of outer hair cell loss can be seen most clearly in Figure 5.2 which overlays the tuning curves at 1kHz for 0% outer hair cell loss (solid curve) and for 100% outer hair cell loss (dotted curve). These tuning curves show the two primary effects outer hair cell loss has on the basilar membrane response. The first is an increase in threshold with increasing loss of outer hair cells. This happens because the interaction the outer hair cells have with the basilar membrane operates to increase the sensitivity of the membrane, thus loss of outer hair cells results in loss of sensitivity. Examination of the measurements demonstrates this fact in the increase of the peak tuning level with increasing outer hair cell loss. It is important to note that increase of the peak tuning level implies a reduction in the sensitivity as the stimuli will require a higher level to reach the peak level and thus be above threshold. The second effect outer hair cell loss has is to increase the bandwidth of the tuning curves. The reason for this is that the outer hair cells function to sharpen the tuning of the basilar membrane, thus outer hair cell loss causes a loss of selectivity/tuning

of the basilar membrane. The low frequency contribution to bandwidth expansion can be seen in the tip-to-tail measurements and the high frequency contribution can be seen in the roll-off measurements. On the low frequency side, the tip-to-tail measurements reduce with increasing outer hair cell loss for two reasons. The first is that the peak tuning level is increasing and the second is that the tail of the tuning curve is becoming more sensitive; the net effect is that the tail and peak are heading towards each other, thus the tip-to-tail ratio reduces and the low frequency bandwidth increases. The high frequency contribution can be seen in the decreasing roll-off. Similarly there are two reasons for this effect. The first is the increase in the peak tuning level and the second is an increase in the sensitivity of the high frequency slope; the result of this is a reduction in the roll-off and an increase of the high frequency bandwidth. The overall effect is an increase in the tuning curve bandwidth and a corresponding decrease in the tuning of the basilar membrane around the region with outer hair cell loss.

The tuning curves at 2kHz and 5kHz are also affected by loss of outer hair cells around 1kHz. In general at up to around 50-60% outer hair cell loss similar effects are seen as occur in the 1kHz tuning curve, however on a *much* reduced level, as to be expected. The effects are particularly noticeable in tuning curves near to the 1kHz tuning curve; as outer hair cell damage will to a lesser extent, occur in these regions too. A different situation arises however as the outer hair cell loss increases towards 100%. The peak tuning levels begin to be decreased and the tip-to-tail and roll-off measurements become elevated. This appears to be in contradiction to the effects induced by outer hair cell loss but it is not. The effects of outer hair cell loss as described occur in conjunction with another effect known as the spread of excitation, which is associated with the reduction in peak tuning level and the increase in tuning curve bandwidth around the 1kHz region of maximal excitation. The apparent contradictory measurements at high outer hair cell loss are caused by this latter effect. This spread of excitation is facilitated by the increased tuning curve bandwidth and associated loss of compression around the 1kHz region. In a sense this can be considered as a form of leakage of the response around the 1kHz region into other frequency regions. The loss of compression facilitates enhanced displacement around the 1kHz region, which spreads to higher frequencies thus enhancing the responses in these regions. The net effect is an apparent increase in sensitivity of the higher frequency tuning curves and is the reason for the elevation in the measurements for the 2kHz and 5kHz tuning curves when the outer hair cell loss is 100% at 1kHz.

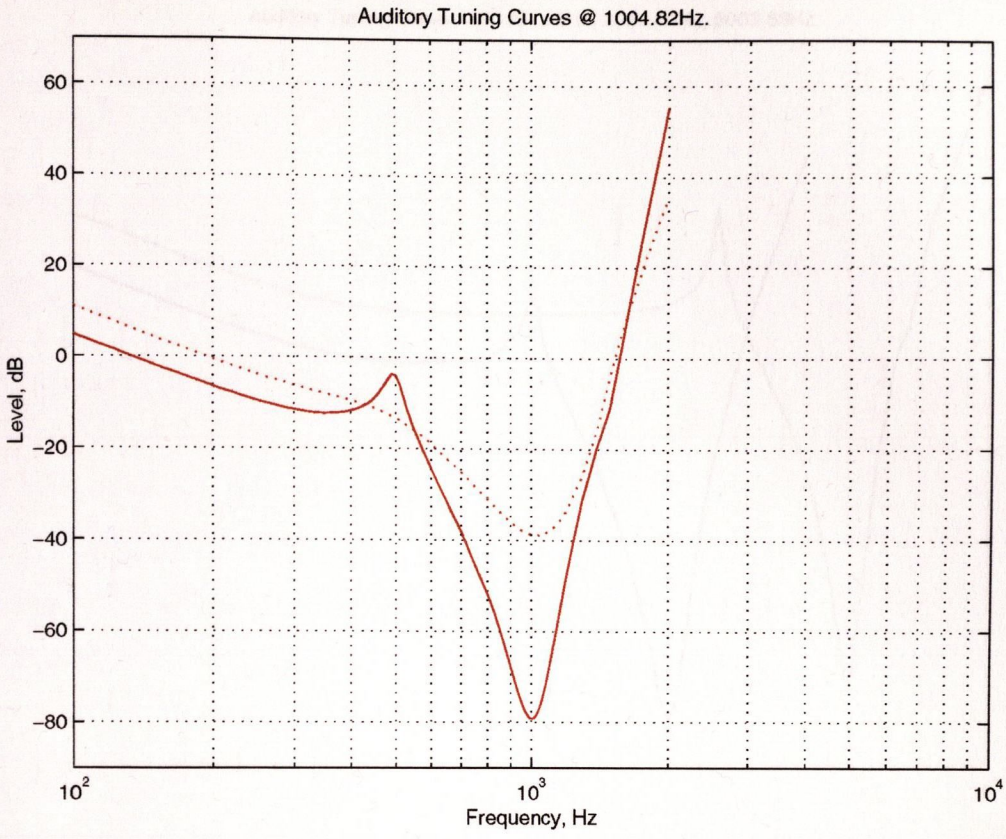


Figure 5.2 Tuning curve at 1kHz for 0% (solid) and 100% (dotted) OHC loss.

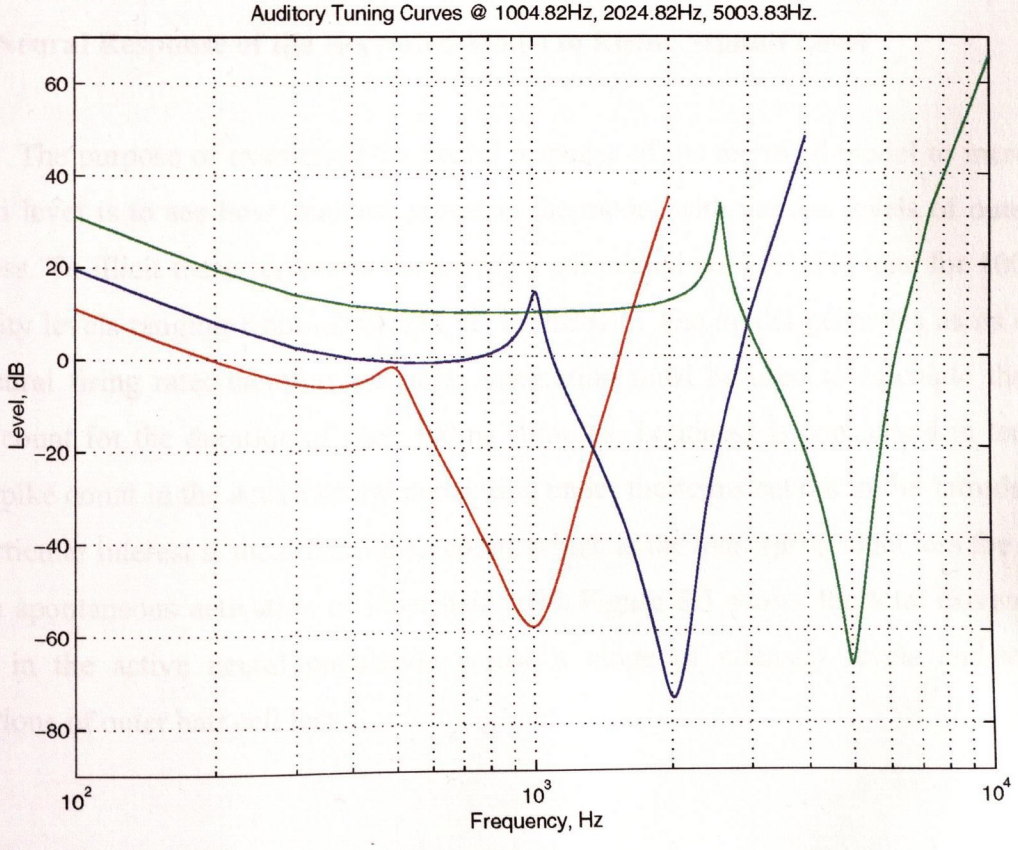


Figure 5.3 Tuning curves for 50% OHC loss at 1kHz.

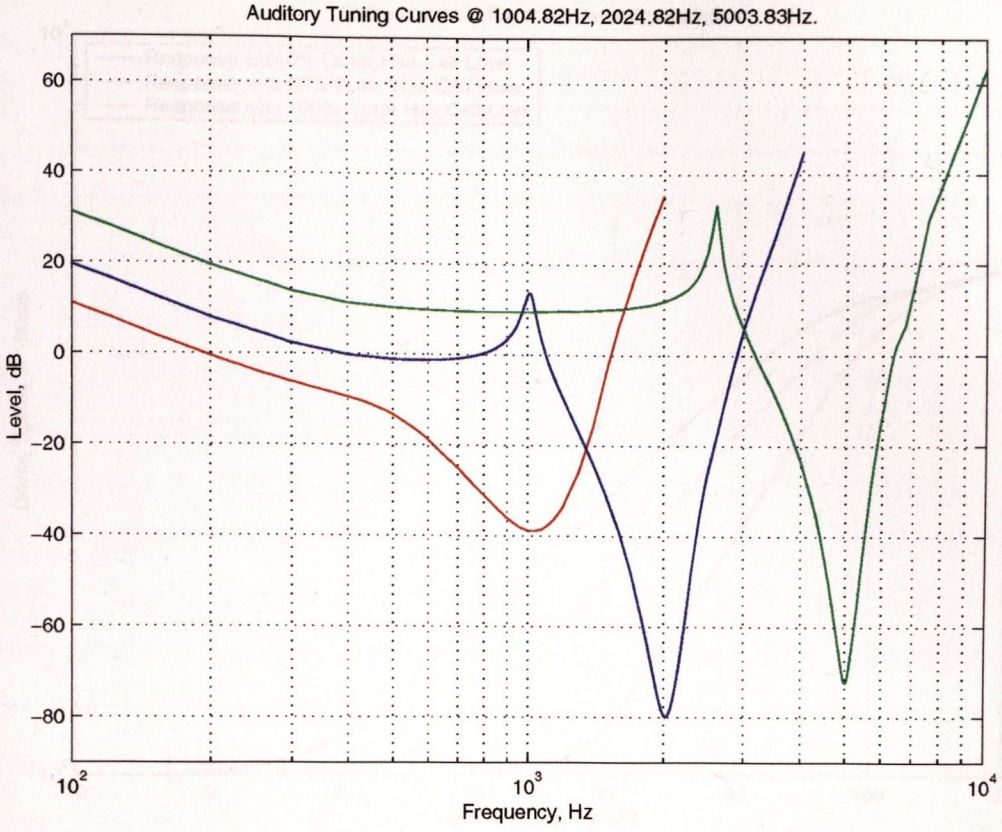


Figure 5.4 Tuning curves for 100% outer hair cell loss at 1kHz.

5.3.2 Neural Response of the Recruited Model to Rising Stimuli Level

The purpose of examining the neural response of the recruited model to increasing stimuli level is to see how loudness grows in the model with various levels of outer hair cell loss. To illicit these responses the model is stimulated with a 1kHz tone for 100ms at intensity levels ranging from –20dBSPL to 120dBSPL. The model generates as its output the neural firing rate; therefore temporal integration must be used to calculate the total spike count for the duration of each 100ms stimulus. Loudness is considered in terms of total spike count in the active neural population under the terms set out in the Introduction. Of particular interest is the driven spike count, which is the total spike count less the spikes due to spontaneous activation of inner hair cells. Figure 5.5 shows the total driven spike count in the active neural population across a range of intensity levels and varying conditions of outer hair cell loss.

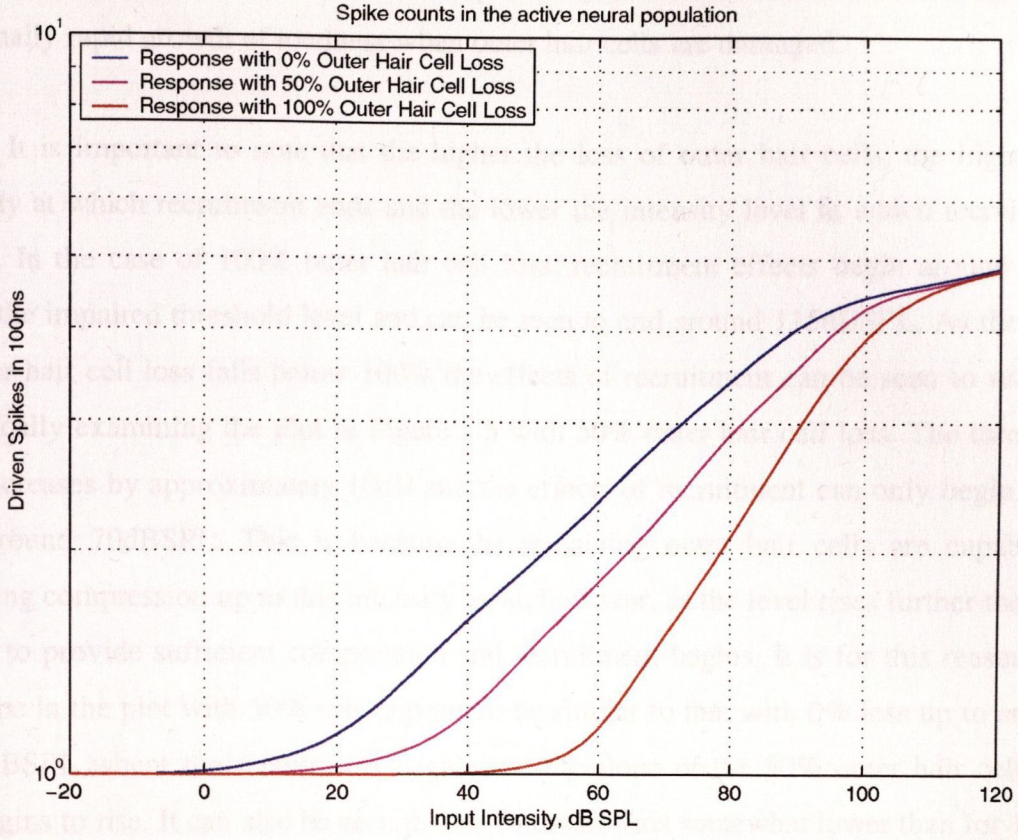


Figure 5.5 Driven spike curves for various levels of outer hair cell loss.

The plots depict the response of the model suffering from 0%, 50% and 100% outer hair cell loss. The outer hair cell loss increases the auditory threshold as seen in the levels at which the spike counts begin to noticeably increase. This progresses from approximately 0dBSPL with no outer hair cell loss to approximately 40dBSPL at 100% outer hair cell loss. Using a similar argument as before, a 100% loss of outer hair cells around the region maximally stimulated by 1kHz signals, causes an elevation of the threshold of audibility by 40dB. Stimulation of the system with a 1kHz stimulus at a level greater than 40dBSPL causes spike counts to increase above the spontaneous level. Considering the spike level at approximately 70dBSPL with no outer hair cell loss. It takes a stimulus approximately 70dB above threshold to stimulate this spike count. However, with 100% outer hair cell loss the same spike count takes a stimulus approximately 50dB above the impaired threshold. The growth in spike count with outer hair cell loss is therefore abnormally rapid and hence the loudness is also growing abnormally rapidly when outer hair cell loss is present. The spike counts however continue growing until they approach the same level around 115dBSPL, therefore the stimuli must be equally loud in both systems at this level. The intensity level at which the spike counts in both systems are approximately equal is the

point at which recruitment ends in the impaired system. The model therefore shows an abnormally rapid growth of loudness when outer hair cells are damaged.

It is important to note that the higher the loss of outer hair cells, the higher the intensity at which recruitment ends and the lower the intensity level at which recruitment begins. In the case of 100% outer hair cell loss, recruitment effects begin around 10dB above the impaired threshold level and can be seen to end around 115dBSPL. As the level of outer hair cell loss falls below 100% the effects of recruitment can be seen to subside. Specifically examining the plot in Figure 5.5 with 50% outer hair cell loss. The threshold only increases by approximately 10dB and the effects of recruitment can only begin to be seen around 70dBSPL. This is because the remaining outer hair cells are capable of providing compression up to this intensity level; however, as the level rises further they are unable to provide sufficient compression and recruitment begins. It is for this reason that the slope in the plot with 50% loss appears to be similar to that with 0% loss up to around 65-70dBSPL where the recruitment begins and the slope of the 50% outer hair cell loss plot begins to rise. It can also be seen that recruitment ends somewhat lower than for 100% outer hair cell loss, around 105dB, where the spike counts converge. Figure 5.6 shows the results of the same tests performed in [131] in which similar effects can be seen.

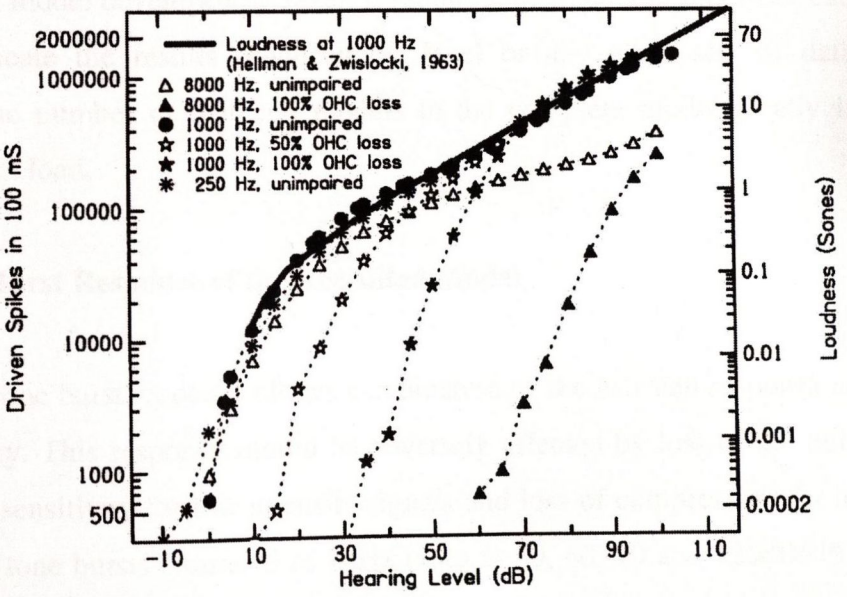


Figure 5.6 Correlates of loudness growth and recruitment [131].

Careful consideration must be taken when examining the plots in Figure 5.6. The loudness of a 1kHz tone measured in humans by Hellman & Zwislocki [141] references the right hand vertical axis. The total driven spike count in Javel's model references the left

hand vertical axis in driven spikes in 100ms. The total spike count in Javel's model is adjusted to match the measured 1kHz loudness curve. What is important to take from this graph is the relative changes between the spike curves with and without recruitment taking place. In particular it is worth noting the elevated threshold and the abnormally rapid growth of spike level present when outer hair cells are damaged in the model.

The curves in Figures 5.5 and 5.6 show good correlation in the trends that arise with increasing levels of outer hair cell loss. The reason for the large difference in driven spike counts between the figures is that Javel's model only involves the cochlea and focuses on representing all the hair cells. The model in this work emulates one hair cell model per centre frequency; however it still generates the required responses on a reduced scale. One of the main effects loss of outer hair cells has is to shift the intensity at which driven spike counts begins to undergo a slope change at high intensity levels where the neural firing begins to saturate. With no outer hair cell damage the slope in the model in this work begins to change at an intensity level some 30dB higher than in Javel's model. This is purely due to setting the maximum Q values in the cascade; these maximum values can be set to locate the turning point at any input intensity level. The changes in the curves due to the outer hair cell loss are always relative to the curve with 0% outer hair cell loss. In general, the response in the model in this work shows very good agreement with the recent neural model developed by Javel. If required, the model in this work can be tuned to exactly replicate the results produced by Javel or any other sets of data. However, increasing the number of hair cell models in the complete model greatly increases the computational load.

5.3.3 Tone Burst Response of the Recruited Model

The tone burst response allows examination of the hair cell response at a particular test frequency. This response should be adversely affected by loss of the outer hair cells, with loss of sensitivity for low intensity signals and loss of compression for high intensity signals. The tone bursts consisted of 1kHz tones at 30, 60, 90 and 120dBSPL. The neural responses of the inner hair cell model centred on 1kHz are displayed in Figure 5.7.

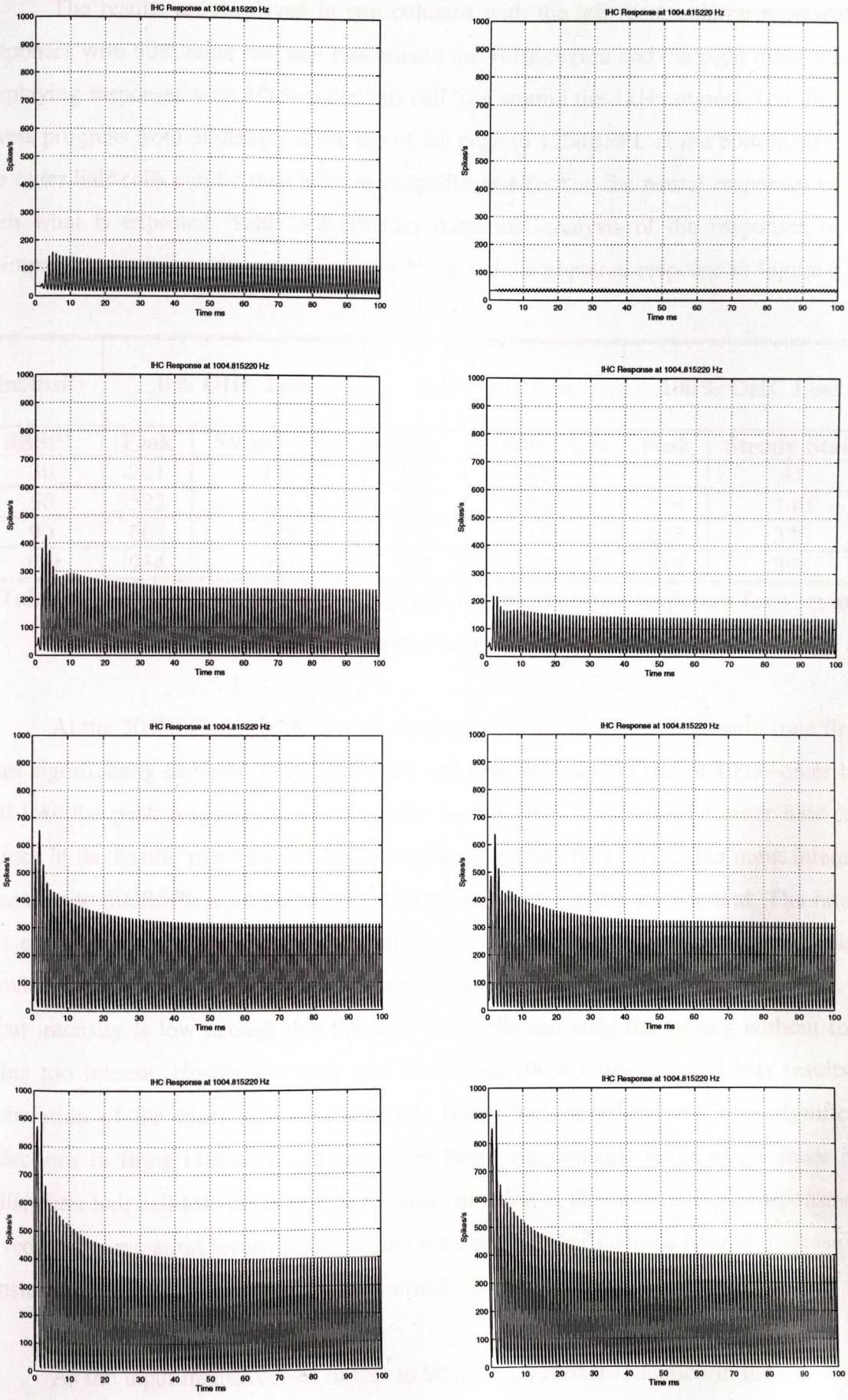


Figure 5.7 1kHz tone burst response at 30, 60, 90 & 120dBSPL top to bottom with outer hair cell loss of 50% and 100% from left to right.

The results are displayed in two columns with the left most column representing responses with 50% outer hair cell loss around the 1kHz region and the right most column displaying responses with 100% outer hair cell loss around the 1kHz region. The intensity levels progress from 30dBSPL at the top of the page to 120dBSPL at the bottom. Overall, the outer hair cells can be seen to have a significant effect on the neural responses in line with what is expected. Table 5.2 provides numerical analysis of the responses of the unimpaired tone bursts shown in Figures 4.25a-d and the impaired response in Figure 5.7.

Intensity	0% OHC Loss		50% OHC Loss		100% OHC Loss	
dBSPL	Peak	Steady-State	Peak	Steady-State	Peak	Steady-State
30	261	169	156	118	-	41
60	522	283	429	240	218	140
90	667	318	659	313	637	320
120	934	401	925	400	919	398

Table 5.2 Peak and steady state firing rates in the tone burst responses for varying levels of outer hair cell loss

At the 30dBSPL intensity level both the peak firing rates and the steady state firing rates significantly decrease as the outer hair cell loss increases. In fact at 100% outer hair cell loss the peak response is lost altogether as the extra sensitivity the outer hair cells induce in the basilar membrane at low intensity completely falls off. As the input intensity increases to 60dBSPL a similar trend occurs, however, not to the same extent. The rate of fall of firing rate is lower as the stimulus level is high enough, without the assistance provided by the outer hair cells, to activate the inner hair cells. Up to about 60dBSPL the input intensity is low enough that the inner hair cells can code the signals without them being too intense. However at very low intensities, 100% outer hair cell loss results in stimulation of the inner hair cells at levels below their threshold and thus significant reductions in firing rate occur. At intensities below the dynamic range of the inner hair cells, outer hair cell loss predominantly causes reduction in the inner hair cell response due to reduced stimulation levels. In this region the outer hair cells can be seen to increase the sensitivity of the auditory system to low/normal stimulation levels.

As the input intensity rises further to 90dBSPL a further reduction in the rate of fall off of firing rates occurs as the stimulus is more than high enough to be above the inner hair cell thresholds. The inner hair cells are actually being over-stimulated as they are

being stimulated with signals above their dynamic range. In fact, the outer hair cells are clearly unable to compress these signals. This can be seen in the rising steady state levels with loss of outer hair cells. Progression up to 120dBSPL shows relatively steady peak firing rates and significant loss of compression in the steady state firing rates. With 100% outer hair cell loss at 120dBSPL the inner hair cells' steady state responses saturate. The reason the peak level is correctly reproduced at high intensity level is because the initial transient response of the inner hair cells are capable of high level response but these cannot be sustained due to the recovery time of the hair cells. If the input intensity rises further the peak levels will similarly saturate. In general, the model shows excellent correlation with the known information about the interdependence of inner hair cell response with outer hair cell action [131], [120], [129] and [132].

5.3.4 Neural Response of the Recruited Model to Synthetic Speech

The neural response of the auditory model is generated for the same syllable /da/ as before to facilitate comparison between the unimpaired response and the impaired response. The model was stimulated in the same fashion and the same 25ms of the response is shown in Figures 5.8 and 5.9 with 50% and 100% outer hair cell loss at 1.6kHz. These two figures are best considered in conjunction with the unimpaired response displayed in Figure 4.29.

The most noticeable effect of the outer hair cell loss is the increase in the threshold around the 1.6kHz region. At 50% outer hair cell loss many of the low-level complex components around 1.6kHz are unable to breach the threshold as can be seen in the reduced neural firing in this region. Even the high intensity components have a reduced response as the enhancement provided by the outer hair cells is reduced. At 100% outer hair cell loss only the highest intensity transients are at a high enough level to cause stimulation of the inner hair cells. Centring the outer hair cell loss on 1.6kHz has the effect of almost eliminating the reproduction of the 1.6kHz formant associated with the consonant portion of the syllable. The outer hair cell loss affects the adjacent regions to a lesser extent showing a more gradual reduction in the responses moving away from this region. As the tuning reduces with increased outer hair cell loss, the distinction between the peeks representing the format transitions around 1.6kHz also reduces which diminishes the system ability to reproduce the formant.

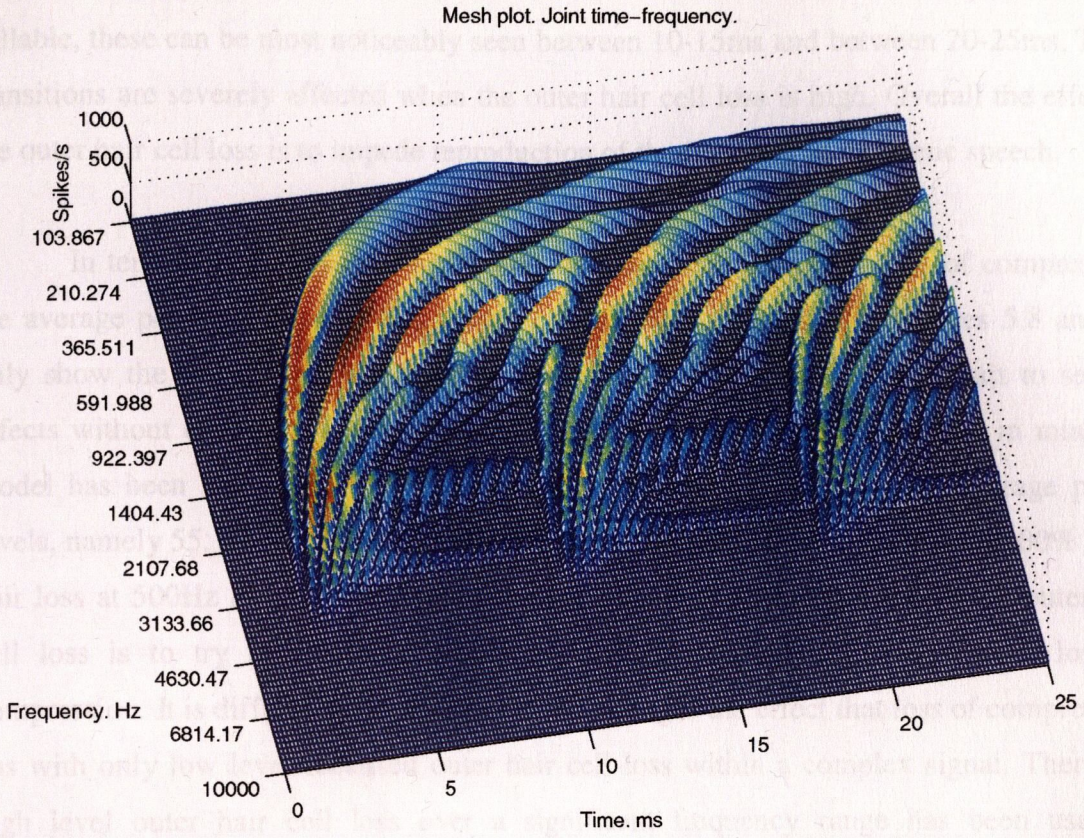


Figure 5.8 Neural response to synthetic speech with 50% OHC loss at 1.6kHz.

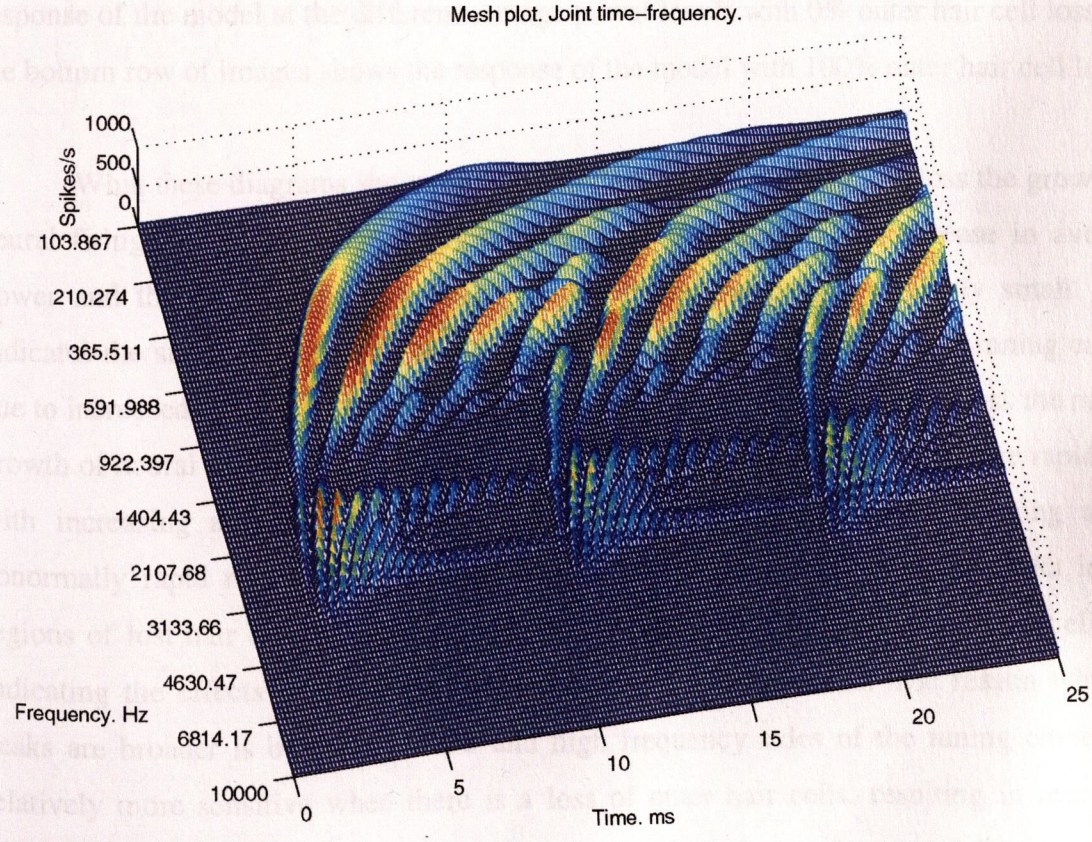


Figure 5.9 Neural response to synthetic speech with 100% OHC loss at 1.6kHz.

Considering the formant transitions associated with the consonant portion of the syllable, these can be most noticeably seen between 10-15ms and between 20-25ms. These transitions are severely affected when the outer hair cell loss is high. Overall the effect of the outer hair cell loss is to impede reproduction of the formants in synthetic speech.

In terms of the abnormal growth in loudness associated with loss of compression, the average power in the stimulus is about 65dBSPL so the plots in Figures 5.8 and 5.9 only show the effect the loss of compression has at this level. It is difficult to see the effects without considering the stimulus at different power levels. With this in mind the model has been stimulated with the same syllable but at three different average power levels, namely 55, 65 and 75dBSPL for the cases of 0% outer hair cell loss and 100% outer hair loss at 500Hz & 1kHz. The reason of implementing such a large range of outer hair cell loss is to try and capture in still images the effects that arise due to loss of compression. It is difficult to demonstrate in still images the effect that loss of compression has with only low level localised outer hair cell loss within a complex signal. Therefore high level outer hair cell loss over a significant frequency range has been used to exaggerate the effects for demonstrative purposes. The effects, however, arise even at low level localised outer hair cell loss. The top row of images in Figure 5.10 show the neural response of the model at the different average power levels with 0% outer hair cell loss and the bottom row of images shows the response of the model with 100% outer hair cell loss.

What these diagrams show is that when there is no outer hair cell loss the growth of neural firing rate in the active population is directly related to the increase in average power and the rate at which the width of the peaks increase is relatively small. This indicates the situation of normal loudness growth and natural broadening of tuning curves due to increased stimulus level. However in the case of 100% outer hair cell loss the rate of growth of neural firing rate in the active population is increasing at a much more rapid rate with increasing average power and thus the loudness must also be growing at an abnormally rapid rate. Similarly the rate of increase in the widths of the peaks in the regions of lost hair cells is significantly larger than in the case of no outer hair cell loss indicating the effects of increased bandwidth of the tuning curves. The reason why the peaks are broader is because the low and high frequency sides of the tuning curves are relatively more sensitive when there is a loss of outer hair cells, resulting in responses being generated over more of the stimulus, thus resulting in broader neural firing patterns. The peaks although broader are still phase locked to the stimulus at low frequency levels.

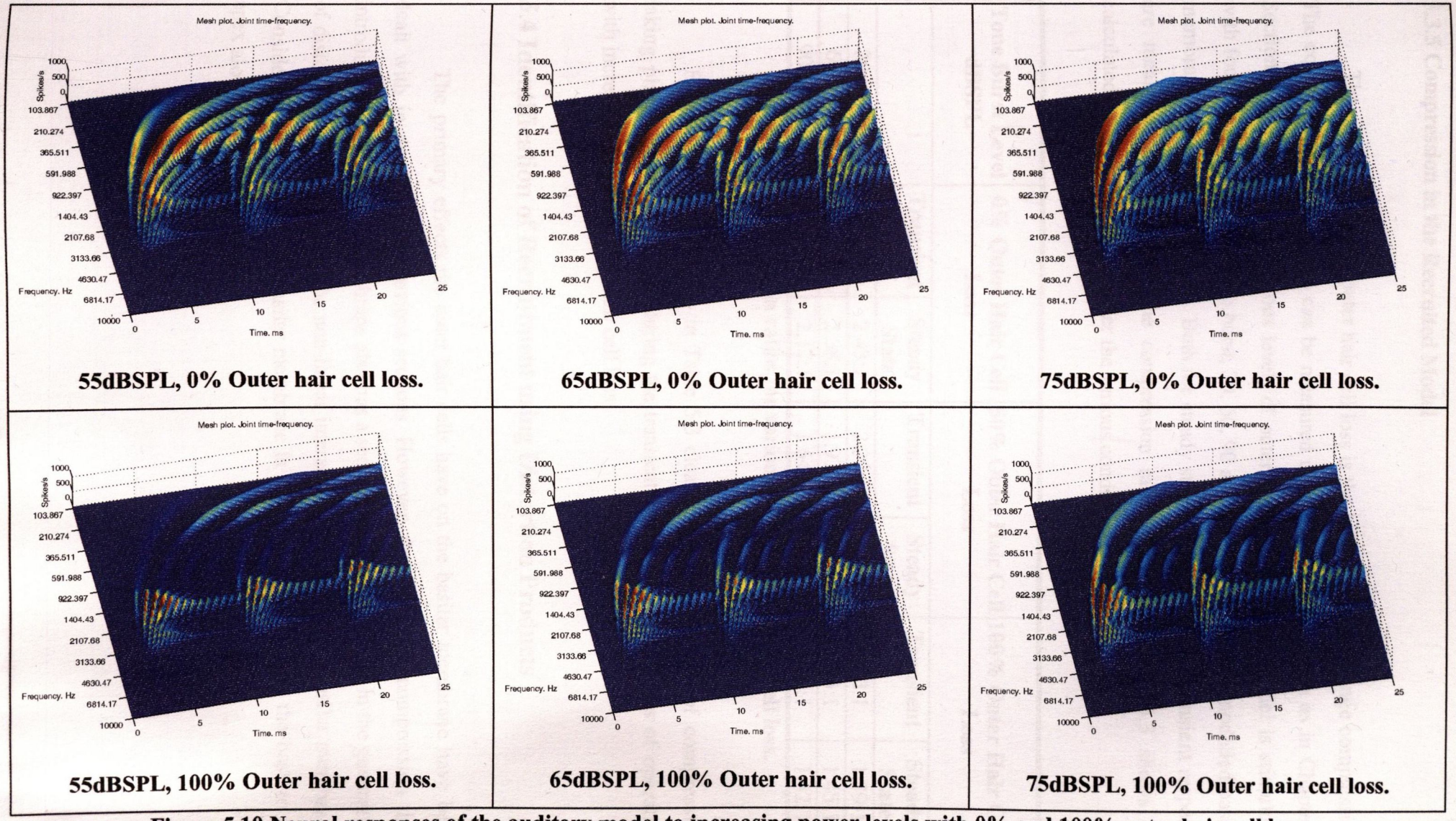


Figure 5.10 Neural responses of the auditory model to increasing power levels with 0% and 100% outer hair cell loss.

5.3.5 Compression in the Recruited Model

The primary effect of outer hair cell loss is the loss of dynamic range compression. The compression in the model can be measured using the same approach as in Chapter 4 Section 4.6.4, except with various levels of outer hair cell loss. The model is stimulated with four different levels of tone burst, 30, 60, 90 and 120dBSPL and the modelled basilar membrane response is measured. Both the steady state response and the transient response are measured as before and the compression ratio calculated. Table 5.3 shows the calculated compression ratios under the various conditions.

Tone Burst Level dBSPL	0% Outer Hair Cell Loss		50% Outer Hair Cell Loss		100% Outer Hair Cell Loss	
	Transient	Steady State	Transient	Steady State	Transient	Steady State
30 => 60	2.35:1	2.43:1	2.09:1	2.14:1	1.43:1	1.39:1
60 => 90	2.31:1	2.38:1	1.97:1	2.08:1	1.43:1	1.45:1
90 => 120	1.93:1	2.17:1	1.64:1	1.80:1	1.31:1	1.42:1

Table 5.3 Compression ratios for various levels of outer hair cell loss.

Examination of the data in Table 5.3 clearly indicates the loss of compression taking place in the model both during the transient and steady state modes of operation with increasing levels of outer hair cell loss.

5.4 Identification of Recruitment using Distortion Products

The primary effects the outer hair cells have on the basilar membrane have been dealt with in detail in the previous sections. However, the non-linear compression they introduce into the basilar membrane also has a secondary effect. It results in the generation of distortion products due to the non-linear interaction of tones on the basilar membrane. Considering Figure 5.11 the basilar membrane is represented going from the base to the apex along the positive x-axis.

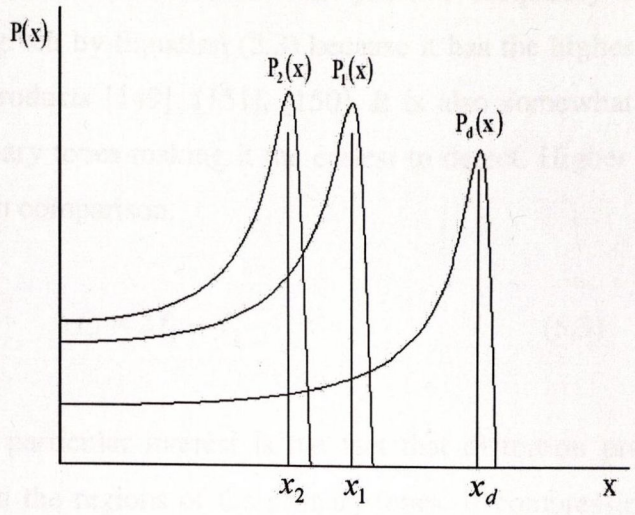


Figure 5.11 Generation of distortion products on the basilar membrane.

In order that distortion products be produced two tones of different frequency f_1 and f_2 must be presented to the model where $f_1 < f_2$. The locations of the points of maximum displacement on the basilar membrane caused by these tones are given by x_1 and x_2. The frequency of the distortion product produced by the non-linear interactions of the primary tones is f_d and is located at x_d. The general expression for the frequencies of the odd order distortion products is given by [148].

$$f_d = f_1 + n(f_2 - f_1) \quad n \in Z, \; -\infty \leq n \leq \infty \tag{5.2}$$

When n is 0 or 1 the distortion product equates to one of the primary tones; for all other integer values of n the equation gives the frequency for the particular distortion product order. E.g. if n is –1 or +2 the equation generates the third order distortion product frequencies. Negative order distortion products have frequencies below the primary tones and positive order distortion products have frequencies above the primary tones, however, all the frequencies generated by Equation (5.2) are taken as absolute frequencies but referred to as negative or positive frequencies to identify whether they are negative or positive as generated by Equation (5.2). The distortion products are believed to be generated near the location x_2 of the primary tone f_2 as this is where the two primaries must have their maximum interaction [149], [150]. It is important to remember that signals of higher frequency peak near the base whereas lower frequencies pass further down the membrane and peak near the apex.

Particular interest is shown in the *positive* frequency third (negative) order distortion product given by Equation (5.3) because it has the highest intensity response of all the distortion products [149], [151], [150]. It is also somewhat isolated in frequency relative to the primary tones making it the easiest to detect. Higher order products can be difficult to isolate in comparison.

$$f_d = 2f_1 - f_2 \qquad (5.3)$$

What is of particular interest is the fact that distortion products depend on the compression around the regions of the primary tones. If compression around the primary tones is reduced then this is reflected in the intensity of the distortion products. Considering these facts the presence or absence of distortion products could be used as an indication of the presence or absence of recruitment. In particular the intensity of the distortion products could be used to quantify the severity of the outer hair cell loss around the test tones and hence the level of recruitment present (see Chapter 8). The model can be shown to successfully generate distortion products by stimulating the model with suitable primary tones. Figure 5.12 displays the neural response of the model when it is stimulated with tones of 2kHz and 3.5kHz. The neural response is displayed instead of the basilar membrane response because it highlights the distortion products occurring in the basilar membrane response and makes it easier to identify them. The third (negative) order (positive) frequency and fifth (negative) order (negative) frequency distortion products can be seen around 500Hz and 1000Hz. The positive third order and positive fifth order distortion products and higher order distortion products cannot be detected as they are at such a low level.

In contrast, when outer hair cell loss is activated the third and fifth order distortion products are significantly reduced. There are three possible combinations of 100% outer hair cell loss, two with loss at only one of the primary tones and one with loss at both tones. Only the plot for 100% outer hair cell loss at both primary tones is shown as this is the most dramatic case. Figure 5.13 shows the neural response for the case of 100% outer hair cell loss at both of the primary tone frequencies.

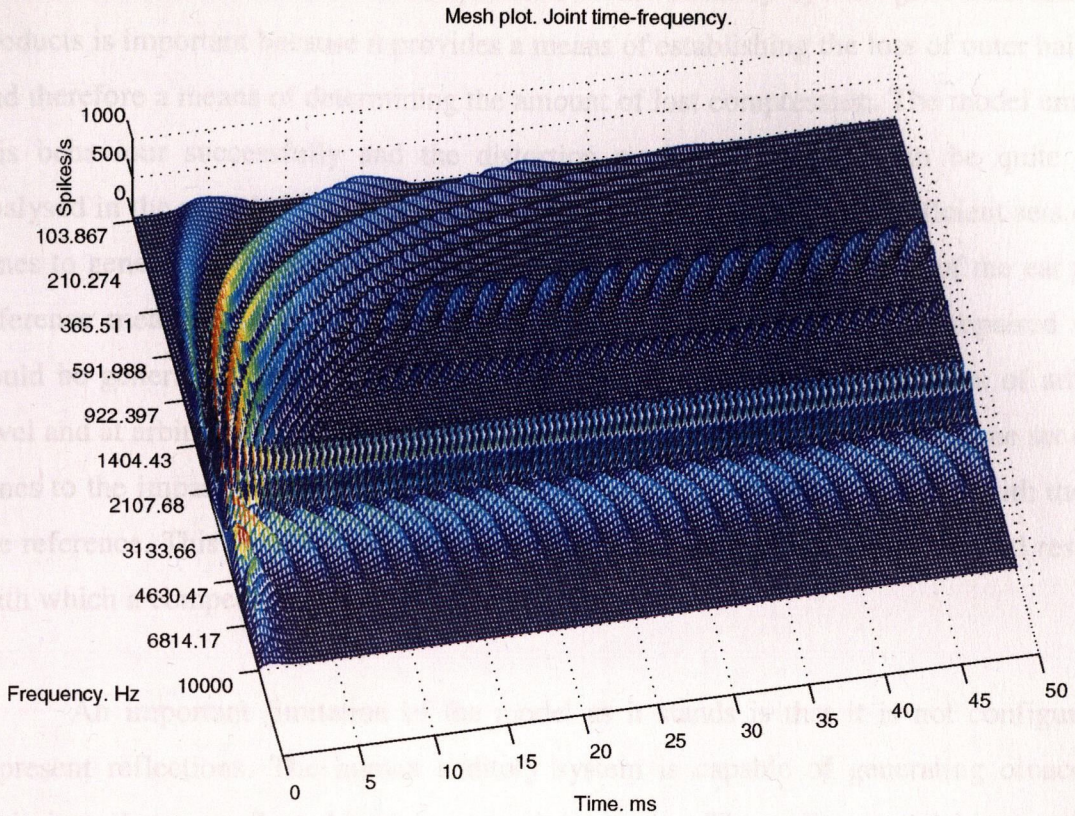


Figure 5.12 Third and fifth (negative) order distortion product generation.

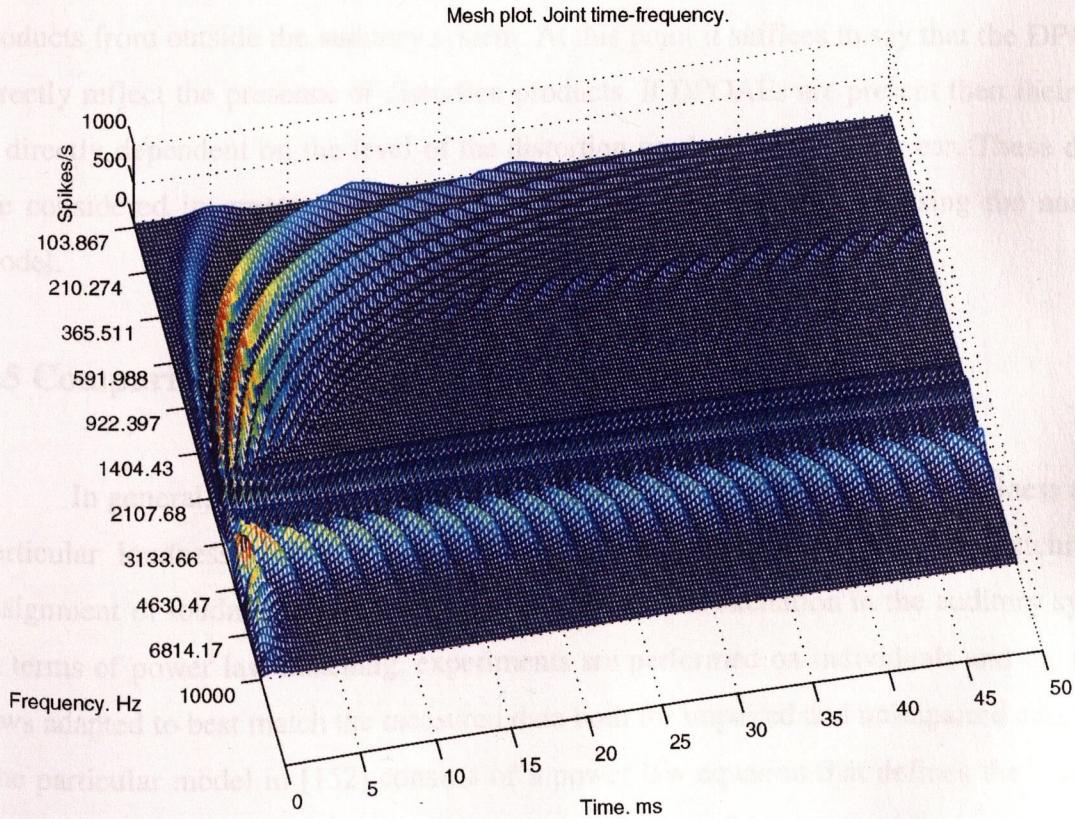


Figure 5.13 Reduced generation of distortion products with 100% outer hair cell loss.

The fact that non-linear compression in the auditory system generates distortion products is important because it provides a means of establishing the loss of outer hair cells and therefore a means of determining the amount of lost compression. The model emulates this behaviour successfully and the distortion products generated can be quite easily analysed in the model. If the unimpaired model were stimulated with sufficient sets of test tones to generate distortion products covering the entire frequency range of the ear then a reference measure of the intensity of the distortion products for the unimpaired model could be generated. Using this reference, introduction of outer hair cell loss of arbitrary level and at arbitrary frequency locations could be detected by applying the same set of test tones to the impaired model and comparing the resulting distortion products with those of the reference. This would provide both an unimpaired reference and an impaired response with which a compensatory function could be pursued.

An important limitation of the model as it stands is that it is not configured to represent reflections. The human auditory system is capable of generating otoacoustic emissions that are reflected back out into the outer ear. The auditory model is not able to channel these reflections out of the model. One of the most important facts about the distortion products is that they generate distortion product otoacoustic emissions, DPOAEs. These emissions can be measured and can be used to quantify the distortion products from outside the auditory system. At this point it suffices to say that the DPOAEs directly reflect the presence of distortion products. If DPOAEs are present then their level is directly dependent on the level of the distortion products in the inner ear. These details are considered in greater detail in Chapter 8 with regard to characterising the auditory model.

5.5 Comparison with other Recruitment Models

In general, most of the models of the auditory system that deal with loudness and in particular loudness recruitment are based on some form of power law matching or assignment of loudness as a post-process analysis of the excitation in the auditory system. In terms of power law matching, experiments are performed on individuals and the power laws adapted to best match the measured data both for impaired and unimpaired ears [152]. The particular model in [152] consists of a power law equation that defines the loudness-intensity relation. As such this is more of a loudness model than an auditory model but is useful when considering the complete auditory model in terms of its loudness response.

The assignment of loudness as a post-process analysis of excitation in the auditory system is often performed on the output of auditory models to account for loudness perception [153], [70].

Considering the experimental measurements from [152] in which three power laws are adjusted to best match measured data, Figure 5.14 displays the measured results with black dots and the adjusted power laws using dashed lines.

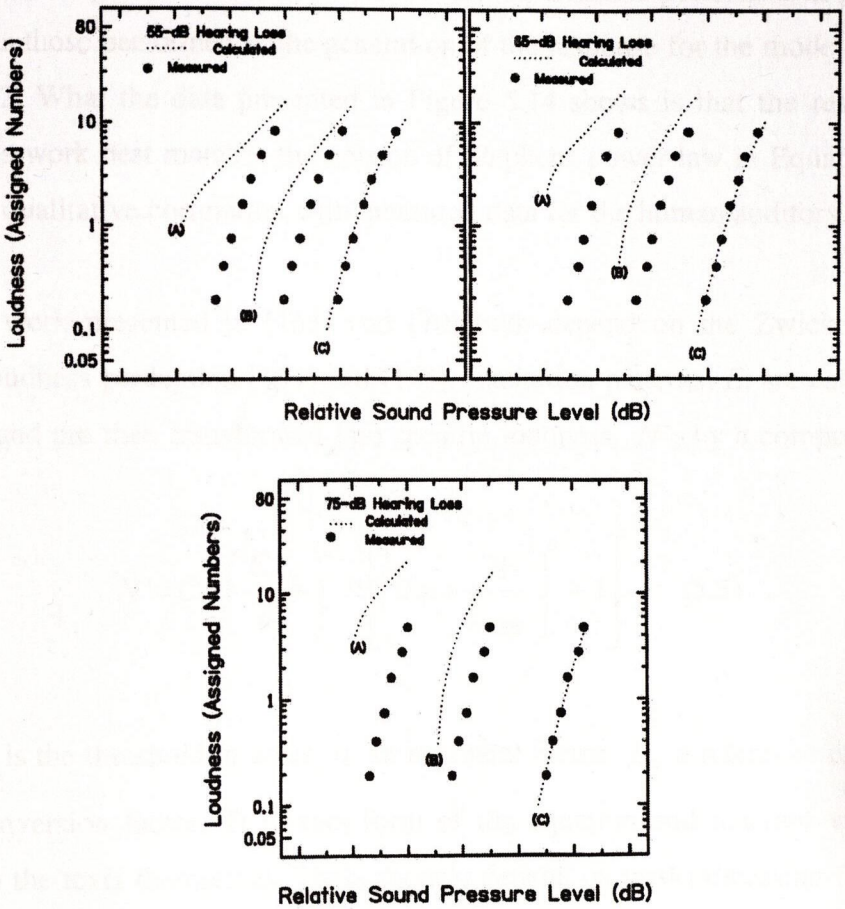


Figure 5.14 Measured and calculated loudness functions for three cochlear impaired groups [152].

Generally only the third adapted equation, which takes the form in Equation (5.4), successfully approximates the shapes of the loudness curves. It is the third curve on all the plots.

$$L = k\left[\left(X^{2} + cX_{t}^{2}\right)^{n} - \left(cX_{t}^{2}\right)^{n}\right] \qquad (5.4)$$

X & X_t are the stimulus intensity and threshold; k, c & n are constants; L is the loudness.

The loudness curves shown for the model developed in this work (Figure 5.5) show similar loudness growth trends as those generated by the third power law in [152] for similar levels of impairment. Care must however be taken when interpreting the graphs in Figure 5.14 as all the graphs represent the same data but are offset for clarity; also the units being used must be considered when making a comparison. The loudness is given in *Assigned Numbers* which are not real units and the intensity is in *relative sound pressure level* due to the presentation of tone bursts through headphones. Each tick on the x-axis represents 10dB. Apart from these differences, the experiments performed in [152] take the same form as those performed in the generation of the response for the model in this work, Section 5.3.2. What the data presented in Figure 5.14 shows is that the response of the model in this work best matches the version of Stephens power law in Equation (5.4) and shows good qualitative correlation with measured data for the human auditory system.

The work presented in [153] and [70] both depend on the Zwicker's loudness model for loudness perception [154] and [155]. Excitation patterns, E, are calculated from the models and are then transformed into specific loudness, N', by a compressive power law.

$$N' = C \times \left(\frac{E_{ThQ}}{E_0} \right)^{\alpha} \times \left[\left(0.5 + \frac{E}{E_{ThQ}} \right)^{\alpha} - 1 \right] \qquad (5.5)$$

where E_{ThQ} is the threshold in quiet, α an exponent factor, E_0 a reference excitation and C a unit conversion factor. The exact form of the equation and relevant values can be examined in the texts themselves. These models depend on post-processing for emulation of the compressive non-linearity of the auditory system. It suffices to say, however, that these models depend on a power law. In a similar manner the power law used is adjusted to best match experimentally measured data. Both models sum the specific loudness across the critical-band-rate scale to generate an overall loudness; N. Figure 5.15 shows the results of [153] plotting the overall loudness in *Categorical Units* (Equation (5.6)) relative to intensity with various noise bandwidths, (13 Bark = 2kHz).

$$CU = 17.6 \times \log(2.5 \times N) \qquad (5.6)$$

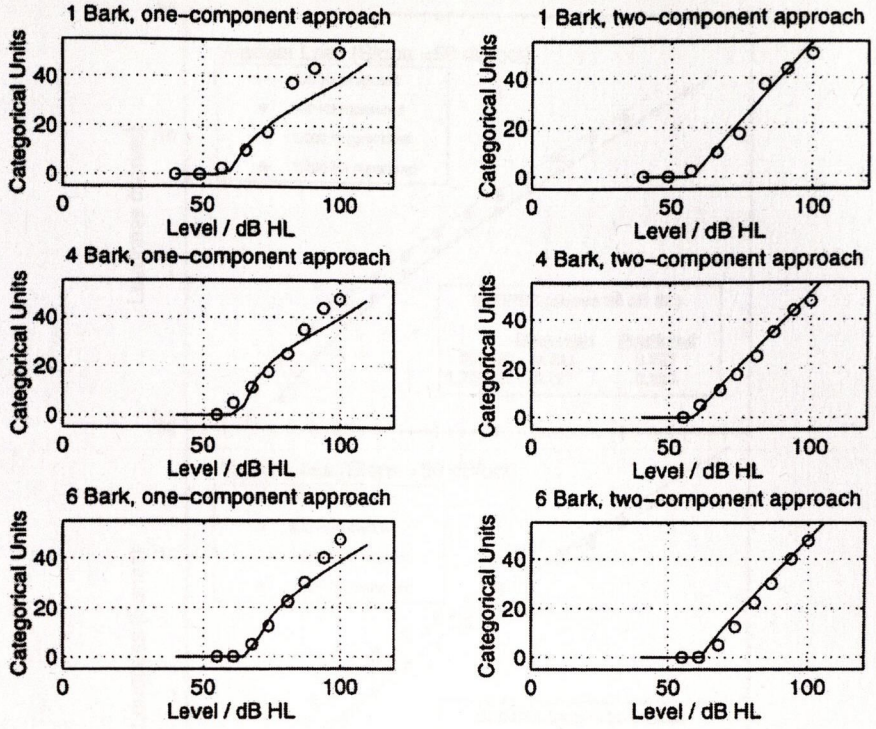


Figure 5.15 Measured and calculated loudness data for band-filtered noise [153].

The left column in Figure 5.15 represents the results varying one component of the power law, the right represents the results varying two components. In general, the two component approach is far superior to the one component approach. Specifically, the one component approach fails at medium to high intensity levels where the slopes of the measured and calculated loudness functions are significantly different. The responses here are similar to those in [152] and correlate very nicely over the ranges displayed with the responses from the model in this work.

Figure 5.16 shows the results of [70] plotting overall loudness in *sones* at different intensity levels for two different degrees of hearing loss. Although not specifically modelling outer hair cell loss the nature of the loudness growth is consistent with total neural activity measurements in the model presented in the current work under impairment conditions consistent with outer hair cell loss accompanied with some inner hair cell loss.

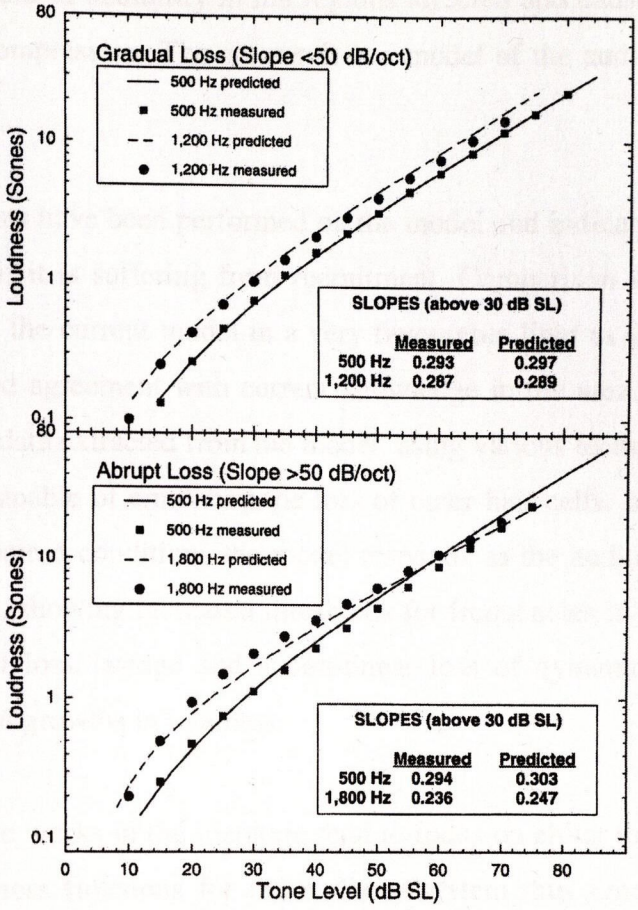


Figure 5.16 Model predictions compared to loudness measurements [70].

In general, the models mentioned above purely focus on the loudness response of the auditory system to the exclusion of the other effects the non-linearity of the auditory system generates. In themselves these models are far from complete representations but they do provide valuable in-depth insights for the development of complete models in that they analyse the different traits of loudness growth within the auditory system. These traits must be replicated by the complete model if it is to successfully generate the required growth of loudness both in the impaired and unimpaired states.

5.6 Summary and conclusions

In summary the unimpaired auditory model can be made to emulate an auditory system suffering from recruitment by implementing damage to the outer hair cells in the auditory model just as it would occur in the real system. Specifically the outer hair cells effect their action in the feedback pathway. Outer hair cell loss can be modelled by reducing the effect of their action and limiting the full range of Q factors below their maximum level in the auditory model. In particular, the outer hair cell damage/loss

increases the threshold of audibility in the regions affected and causes the non-linear loss of dynamic range compression. The net result is a model of the auditory system suffering from recruitment.

Extensive tests have been performed on the model and indicate that the model does indeed function as if it is suffering from recruitment. Comparison with the responses of other models shows the current model in a very favourable light as it generates responses that are in very good agreement with current knowledge in the area. Consideration of all the different sets of data extracted from the model, using various techniques, illustrates that the model is fully capable of emulating the loss of outer hair cells. In particular, it shows that under such impaired conditions the model responds as the auditory system would be expected to respond, showing increased thresholds for frequencies in the regions suffering from outer hair cell loss/damage and a non-linear loss of dynamic range compression resulting in abnormal growths in loudness.

In general the works in the literature tend to focus on either the distortion products or developing loudness functions for the auditory system thus ignoring the connection between the two, namely that they both depend on the outer hair cell function. In particular loudness levels can be associated with the intensity of distortion products. If the loudness is growing abnormally then the distortion products must also be changing abnormally as they both depend on the non-linearity in the auditory system arising from the outer hair cell action. If the outer hair cells are damaged both of these processes will be affected. Seeing as loudness is subjective, its measurement cannot be precise. However, the indirect measurement of distortion products is objective and directly related to loudness and thus can be used as a method for ascertaining loudness level. Although, the distortion products must be measured indirectly through DPOAEs as the basilar membrane and the auditory neurons are not directly accessible. This is considered in greater detail in Chapter 8.

Chapter 6 - Frequency Smearing

6.1 Introduction

Just as for the development of a model of the impaired auditory system suffering from recruitment, the unimpaired model is used as the starting point for the development of a model of a sensorineurally impaired auditory system suffering from a phenomenon known as *frequency smearing*. Frequency smearing is a process whereby the impaired auditory system fails to appropriately respond to the frequency domain resonances in speech.

The cause of this impairment is best explained in terms of the fully functional auditory processes taking place in the inner ear. Once a stimulus has passed through the middle ear and is presented to the inner ear, the travelling waves on the basilar membrane stimulate the inner hair cells. Examination of the tuning curves of the basilar membrane, as presented earlier, shows that they are finely tuned at the location of maximum displacement and it is at these locations that the inner hair cells are maximally stimulated. When the model is unimpaired the level of inner hair cell activity at these locations is significantly greater than the activity around these locations. However, if frequency smearing takes place, the level of inner hair cell activity at these locations becomes less distinguishable from the activity around these locations. There are three possible scenarios in which frequency smearing arises.

1. Damage to the inner hair cells at the maximal response location causes a reduction in the neural firing at this location such that it eventually becomes indistinguishable from the unimpaired neural activity around this region. The auditory system then exhibits a response as if it were stimulated by a stimulus having a broad frequency content and generating a neural output that the brain cannot distinguish as having a primary frequency component.

2. The tuning curves at the maximal response location become broader thus reducing the stimulation of the inner hair cells at the maximal response location and exhibit enhanced stimulation of the inner hair cells around this region. This will decrease the inner hair cell activity at the centre frequency location and increase the activity of the

inner hair cells around this location. Eventually the location of maximal inner hair cell activity becomes indistinguishable from the neural activity around this region.

3. The final possibility is that both of the situations in 1 & 2 can occur together.

These three scenarios can be differentiated from one another when compared to the unimpaired response. Figure 6.1 shows diagrammatic representations of the unimpaired situation and the three possible situations that can arise.

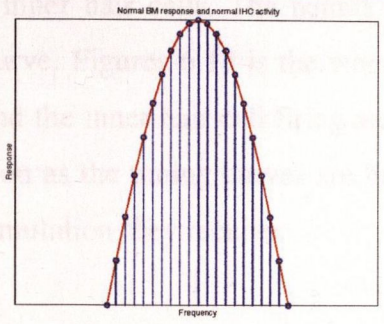


Figure 6.1a Unimpaired System

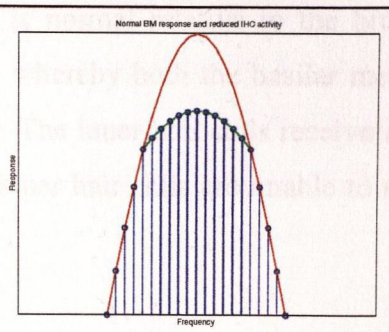


Figure 6.1b Impaired IHC.

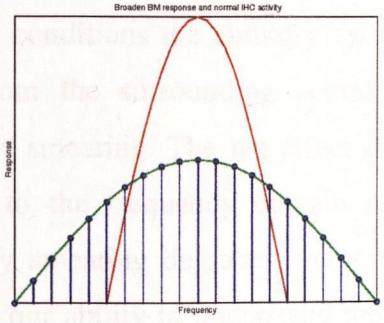


Figure 6.1c Broadened BM tuning.

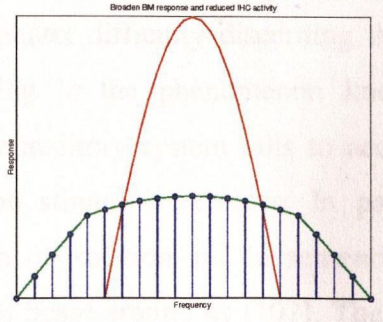


Figure 6.1d Broadened BM tuning and Impaired IHC.

Figure 6.1 Diagrammatic representations of the situations under which frequency smearing can occur.

In all the diagrams in Figure 6.1, the normal basilar membrane tuning and the normal inner hair cell firing is represented in red as a *reference*, the actual inner hair cell firing in blue and the net neural firing in green. Note: These diagrams are purely for illustrative purposes only; therefore no units have been explicitly shown.

In Figure 6.1a neither the basilar membrane tuning nor the inner hair cell firing are defective and the peak neural firing coincides with the peak tuning of the basilar membrane. In this situation the peak firing is clearly distinguishable from the neural firing

taking place around the peak. The most significant firing is that which occurs within the bandwidth of the normal tuning curve.

Figure 6.1b shows the situation where the tuning curves are normal but the inner hair cells at the location of maximal tuning are impaired, resulting in a reduction of the peak inner hair cell firing. In this situation the hair cells are being properly stimulated but they are unable to respond. Figure 6.1c shows the situation of broadened tuning curves and normal inner hair cell firing. In this case the impaired tuning curve matches the green curve and the inner hair cells are receiving reduced stimulation as it is spread out over a wider band of inner hair cells. The neural firing however is normal relative to the broadened tuning curve. Figures 6.1d is the worst case scenario whereby both the basilar membrane tuning and the inner hair cell firing are both impaired. The inner hair cells receive reduced stimulation as the tuning curves are broader and the inner hair cells are unable to respond to the stimulation they receive.

These diagrams show the situations under which the relative difference between the peak neural firing and the surrounding neural activity is greatly reduced. Under these impaired conditions the auditory system has much greater difficulty discerning the peak firing from the surrounding neural activity resulting in the phenomenon known as frequency smearing. The net effect of this is that the auditory system fails to accurately respond to the frequency domain resonances in the stimuli it receives. In particular frequency smearing degrades word recognition, both in isolation and in sentences thus reducing our ability to understand the acoustic stimuli heard around us [102]. The effects of frequency smearing most noticeably degrade the reproduction of formants in the neural output of the auditory system.

Another feature of frequency smearing not represented in the diagrams in Figure 6.1 is that the *apparent* frequency position is not fixed. (The reason for the apparent frequency position will become clear). Consider the scenario in which the auditory system is stimulated at a particular frequency; however, there is hair cell damage at the location on the basilar membrane that is maximally stimulated by this frequency. If the hair cells at the precise location of maximal stimulation are badly damaged but the surrounding hair cells are in relatively good condition then the neural activity at the normal maximum location will be very low but the activity generated by the surrounding hair cells may be quite high; in fact the surrounding activity may even be higher than at the precise location of the

normal maximum location. In this case the neural response generated by the auditory system will be as if its primary stimulation were at frequencies either above, below or both above and below the actual frequency of stimulation (depending on the condition of the surrounding hair cells) thus generating false maxima. In this scenario, the sufferer is presented with a stimulus at a particular frequency but perceives a stimulus at other surrounding frequencies. This type of condition typically occurs where there is severe hair cell damage occurring in a discontinuous fashion, i.e. when there is not a smooth transition in hair cell damage between regions of severe damage and very minor damage. In general though, the more usual case is for a smooth transition of hair cell damage between regions, which leads to the most interesting form of this scenario. If the neural activity generated *around* the location of maximal stimulation is more or less at the same level as the activity *at* the location of maximal stimulation, then the overall pattern of neural activity generated in the region can have a maximum that appears to move around the damaged region as fluctuations in the surrounding faulty hair cell firing generate instantaneous peaks while on average maintaining the same level of activity as that at the location of maximal stimulation. In this situation the sufferer is presented with a stimulus at a particular frequency but perceives a stimulus with an apparently random varying primary frequency content.

6.2 Modelling Frequency Smearing

The unimpaired auditory model can be made to emulate the auditory system suffering from frequency smearing by incorporating into the model the particular malfunctions that cause frequency smearing to take place in the real system. The description provided in the Introduction to this chapter identifies three possible scenarios under which frequency smearing can take place. However, there are effectively two primary causes, inner hair cell malfunction and broadening of the cochlear tuning curves.

When the inner hair cells are damaged a reduction in neural firing ability occurs in the region of the damaged inner hair cells causing reduced neural activity around the centre frequency for that site. The net effect is a reduction in the inner hair cell sensitivity at the particular centre frequency causing the system to be unable to accurately respond at the centre frequency. The broadening of the cochlear tuning curves causes a reduction in the stimulation of the inner hair cells at the centre frequency and an increase in the stimulation of the inner hair cells in the regions around the centre frequency. The net effect is that the

system has a reduced response at the centre frequency. When considering broadening of the cochlear tuning curves it is known that the outer hair cells are responsible for the sharp tuning of the basilar membrane, therefore loss of outer hair cell function, as before, will cause broadening of the tuning curves and a reduction in the peak tuning.

It becomes apparent that the inner and outer hair cells are very important for accurate generation of the neural response to acoustic stimuli. Inner hair cell malfunction causes frequency smearing, outer hair cell malfunction causes frequency smearing to a lesser extent; however, both inner and outer hair cell malfunction cause the most severe frequency smearing. It must also be noted that outer hair cell malfunction causes recruitment. Chapter 5 has dealt with recruitment in detail. Recruitment is therefore *ignored* in this chapter to facilitate the study of frequency smearing. The most realistic scenario of hair cell damage in the auditory system is one where both the inner and outer hair cells at a particular site on the basilar membrane are damaged causing both frequency smearing and recruitment to take place.

How the outer hair cells effect their action and how they can be impaired in the model has already been considered in Chapter 5. In particular Chapter 5 shows how the tuning curve bandwidth increases with increasing outer hair cell loss and how the sensitivity of the tuning curves decreases with increasing outer hair cell loss. This Chapter considers how outer hair cell damage alters the model's neural response in terms of spectral smearing. The method of effecting outer hair cell loss/malfunction is the same as in Chapter 5.

Considering the two different models of the inner hair cell, they both operate in slightly different manners yet produce the same results. Alteration of the inner hair cell firing in the first model requires manipulation of the non-linear permeability of the membrane that controls the release of transmitter into the cleft. The more difficult it is for transmitter to get into the cleft the lower the firing rate, thus reduction of the permeability will reduce the firing ability of the inner hair cells. To alter the firing in the second model the resistance of the stereocilia controlling the flow of ionic current into the hair cell must increase. The more difficult it is for the ions to flow into the cell the more difficult it will be for the cell to reach the ionic charge level required for it to fire, thus reducing the firing ability of the inner hair cells. Regardless of the inner hair cell model used, the auditory model has been set up to take as a parameter the specification of the inner hair cell

percentage operation. This alters the necessary parameters to get the inner hair cell model to emulate an inner hair cell firing at this reduced percentage firing rate. The lower the percentage operation of the inner hair cells the lower the firing ability of the inner hair cells. At 0% operation the hair cells are totally inoperative and are unable to fire at all. In this situation the inner hair cells are effectively destroyed and should not even exhibit spontaneous firing. Equation (6.1) relates how IHC loss is introduced into the IHC models.

$$R(x) = R(x) * \%\ IHC\ Function \qquad k(t) = k(t) * \%\ IHC\ Function \qquad (6.1)$$

6.3 Simulation of the Auditory Model suffering from Frequency Smearing

The effects of frequency smearing on the neural response of the model can be seen most effectively by considering the response of the model to tone burst stimuli and to synthetic speech stimuli. The three possible cases of frequency smearing will be considered for both types of stimulus. The tone burst response shows how the neural activity in the region with impaired hair cells suffers under the various impaired conditions and the response to synthetic speech shows the effects the impairments have on the neural representation of a synthetic speech stimulus. Each type of stimulus will be considered with varying degrees of IHC loss, OHC loss and both IHC & OHC loss.

6.3.1 Tone Burst Response of the Frequency Smeared Model

The tone burst response is considered at a mid-range intensity level with varying degrees of each type of hair cell impairment that causes frequency smearing. High intensity levels are not used as they excessively expand the cochlear tuning curves. The tone burst used consists of a 1kHz tone at 60dBSPL. In this case, however, it is not as useful to only consider the neural response at the frequency of stimulation because we are particularly interested in examining the relative difference between the peak response at the stimulated frequency and the response in the surrounding regions. It is for this reason that the tone burst response in the region of the centre frequency is shown for each type of hair cell loss.

6.3.1.1 Tone Burst Response with Inner Hair Cell Loss

The tone burst results are all shown in a similar manner with the left hand columns

showing an isometric type view of the response for varying degrees of impairment and the right hand columns showing a top down view of the responses for the same degrees of impairment. The first set of responses are for 0%, 50% and 100% inner hair cell loss at 1kHz. The tone burst response under these conditions can be seen in Figure 6.2.

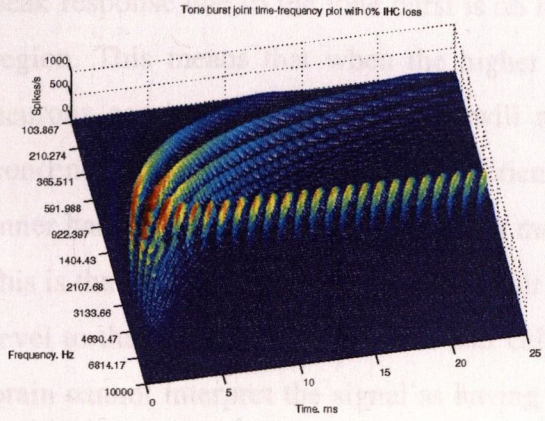


Figure 6.2a 0% IHC loss

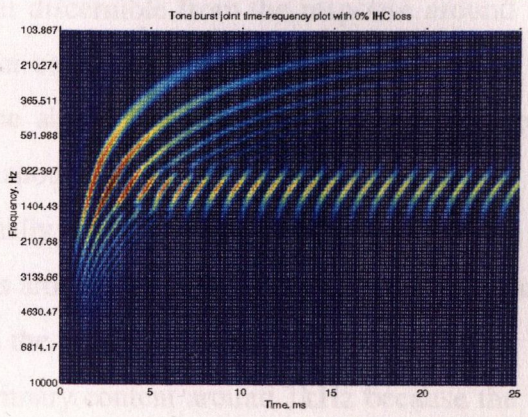


Figure 6.2b 0% IHC loss

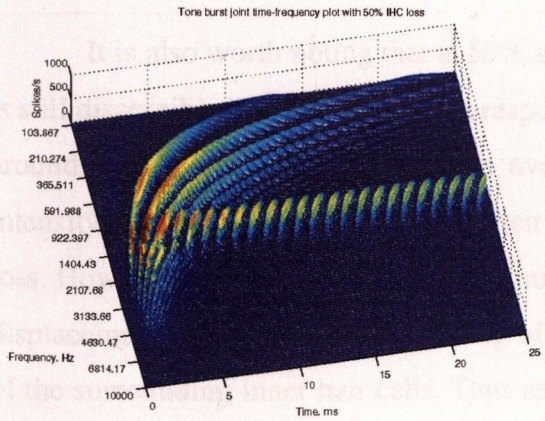


Figure 6.2c 50% IHC loss

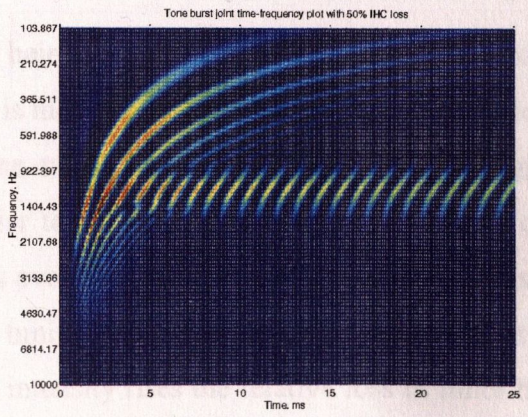


Figure 6.2d 50% IHC loss

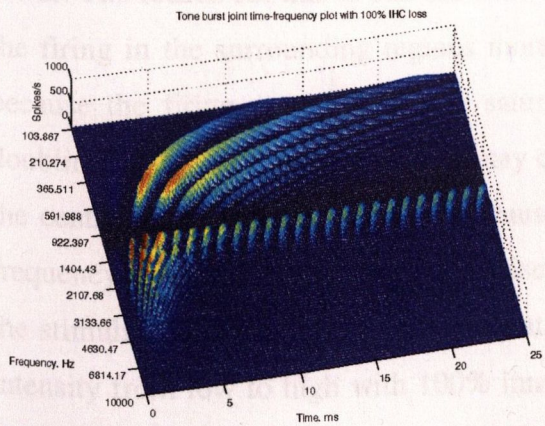


Figure 6.2e 100% IHC loss

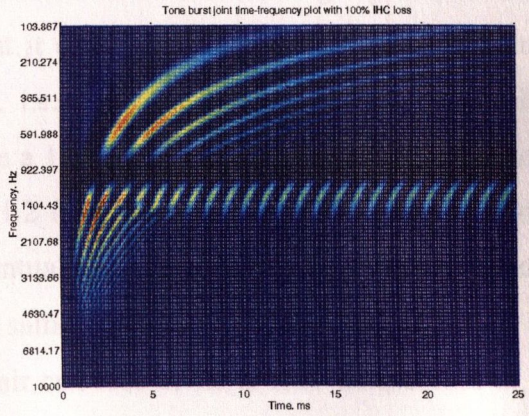


Figure 6.2f 100% IHC loss

Figure 6.2 Tone burst response with 0%, 50% and 100% IHC loss at 1kHz.

Clearly examination of the diagrams in Figure 6.2 shows that for increasing degrees of inner hair cell damage the neural activity decreases. It is particularly important to note that when 100% inner hair cell loss occurs, even the spontaneous firing around 1kHz stops. The reason for this is that if the inner hair cells at 1kHz are totally destroyed then there will be no firing of the neurons connected to these hair cells at all. Comparing the top diagrams and the middle diagrams in Figure 6.2 reveals that, at around 50% inner hair cell loss, the peak response due to the tone burst is no longer discernible from the response around this region. This means that when the higher brain centres try to analyse the spikes on the neurons coming from the ear they will not be able to identify that the highest energy concentration in the signal being experienced should be on the neurons connected to the inner hair cells at the region responding maximally to a frequency of 1kHz. The reason for this is that the response from the inner hair cells around this centre frequency is at a similar level to that of the damaged inner hair cells at the centre frequency. The result is that the brain cannot interpret the signal as having a primary content around 1kHz because the ear is unable to accurately respond to the highest concentration of energy at 1kHz.

It is also worth noting that at 50% inner hair cell loss the transient response at 1kHz is still discernible but the steady-state response is hidden among the response of the regions around 1kHz; however at 100% loss even the transient response is absent. At higher intensity levels the transient will be even easier to identify with medium inner hair cell loss. However, the steady-state response suffers even more at higher intensities as the extra displacement causes further broadening of the tuning curves and even greater stimulation of the surrounding inner hair cells. Thus as the intensity rises the relative loss in inner hair cell function needed to reduce the response at 1kHz, so that it is indiscernible, will be lower. The reason for this is that the extra stimulation of the surrounding regions increases the firing in the surrounding regions more that it does the firing at the centre frequency because the firing rate non-linearly saturates. i.e. if the intensity is at a high level, a doubling of the stimulation intensity may cause a 1.5 fold increase in neural firing *around* the centre frequency but it may only cause a 1.2 fold increase in the firing *at* the centre frequency. If the intensity continues to rise eventually all the inner hair cells at and around the stimulating frequency will be firing at the same level (or higher). Progression of the intensity from low to high with 100% inner hair cell loss at 1kHz sees an increase in the amount of firing below and above 1kHz due to the expanding tuning curves. Even at 60dBSPL the firing rate above 1kHz is quite significant.

6.3.1.2 Tone Burst Response with Outer Hair Cell Loss

In this case the same tone burst stimulus is used except with varying degrees of outer hair cell loss, namely 0%, 50% and 100% outer hair cell loss. The results for stimulation under these conditions are shown in Figure 6.3.

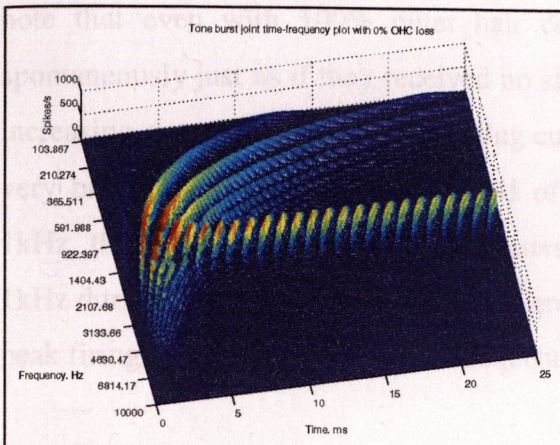


Figure 6.3a 0% OHC loss

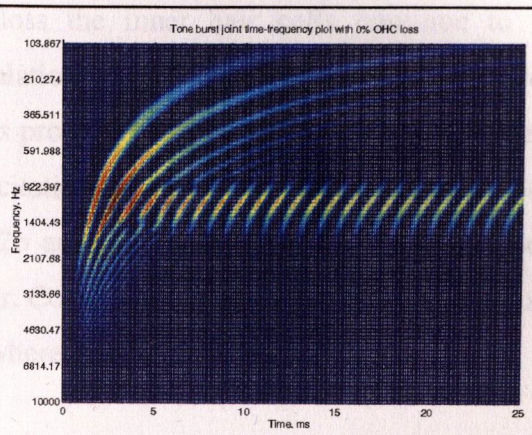


Figure 6.3b 0% OHC loss

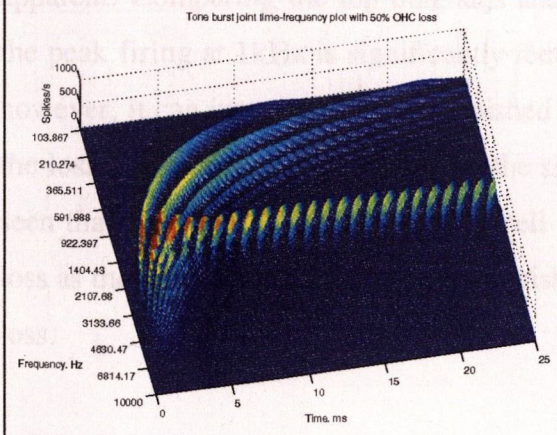


Figure 6.3c 50% OHC loss

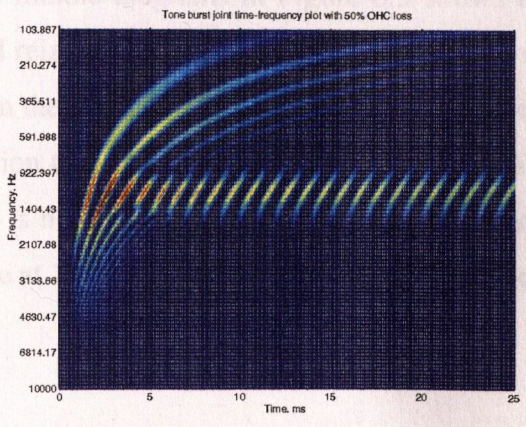


Figure 6.3d 50% OHC loss

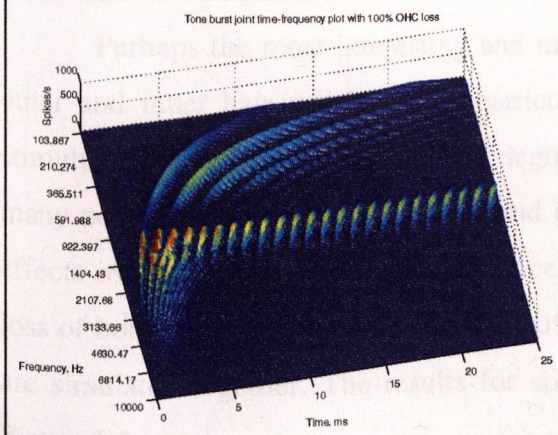


Figure 6.3e 100% OHC loss

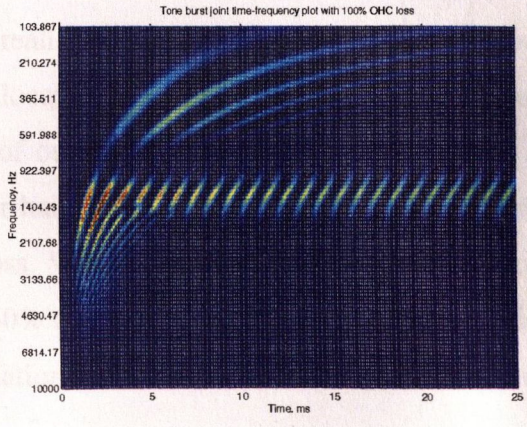


Figure 6.3f 100% OHC loss

Figure 6.3 Tone burst response with 0%, 50% and 100% OHC loss at 1kHz.

Examination of the diagrams in Figure 6.3 shows a progressive loss of peak firing as the outer hair cell loss increases. As before the peak firing is almost totally lost with 100% outer hair cell loss. However, it is not due to malfunction of the inner hair cells, it is because of the significant increases in threshold that arises with major outer hair cell damage. This happens because the gain provided by the outer hair cells is no longer present; a more complete description of this has been given in Chapter 5. It is important to note that even with 100% outer hair cell loss the inner hair cells continue to fire spontaneously just as if they received no stimulation with no outer hair cell damage. With increasing outer hair cell loss the tuning curves progress from being highly tuned to being very broad, causing an upward spread of excitation resulting in increased firing above 1kHz, this can be seen comparing Figures 6.2b and 6.3f. The firing taking place above 1kHz during 100% *outer* hair cell loss is greater. Complete loss of outer hair cells results in peak firing taking place at a higher frequency where there is no damage.

At medium levels of outer hair cell loss, frequency smearing becomes quite apparent. Comparing the top diagrams and the middle diagrams in Figure 6.3 shows that the peak firing at 1kHz is significantly reduced relative to the surrounding neural activity; however, it can just about be distinguished from the surrounding firing. Further increase in the loss of outer hair cells will cause the situation to deteriorate even more. It can also be seen that the effect of 50% outer hair cell loss is not as significant as 50% inner hair cell loss as the peak is still just about distinguishable at 1kHz for the case of 50% outer hair cell loss.

6.3.1.3 Tone Burst Response with Inner and Outer Hair Cell Loss

Perhaps the most interesting and most realistic situation arises when there is both outer and inner hair cell loss at a particular location. In this case the same tone burst stimulus is used except with varying degrees of outer and inner hair cell loss. There are many possible combinations of outer and inner hair cell loss that emphasise the different effects introduced by each type of hair cell loss. We are primarily interested in uniform loss of both inner and outer hair cells so 0%, 50% and 100% inner and outer hair cell loss are simulated together. The results for stimulation under these conditions are shown in Figure 6.4.

Considering the most extreme case of outer and inner hair cell loss at 1kHz we see a broader region of reduced neural firing. The outer hair cell loss is responsible for the broadened neural firing and contributes to the reduction in firing level. The inner hair cell loss is responsible for the significant reduction in neural firing.

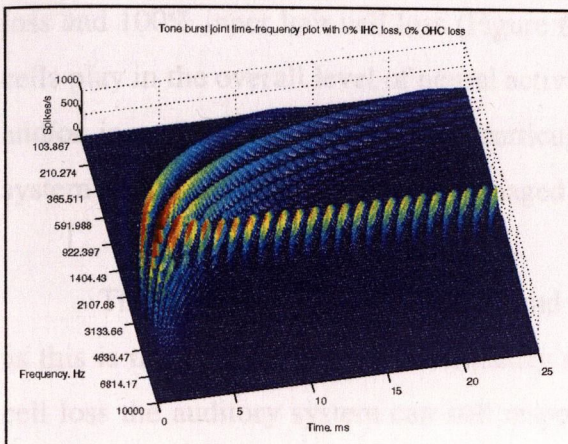


Figure 6.4a 0% IHC loss, 0% OHC loss

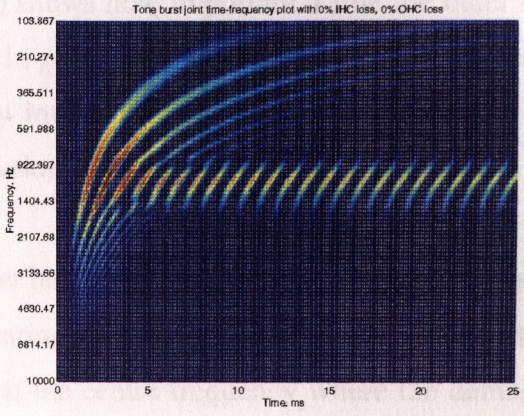


Figure 6.4b 0% IHC loss, 0% OHC loss

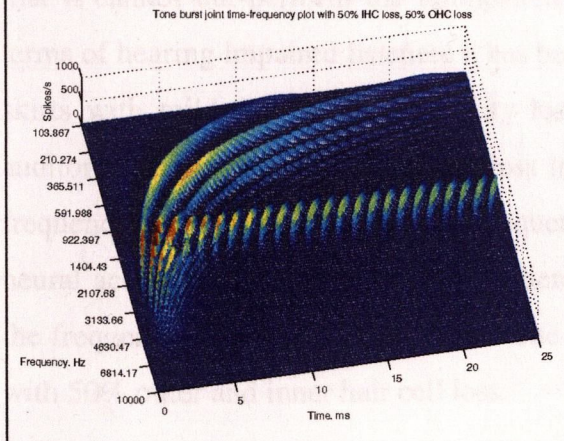


Figure 6.4c 50% IHC loss, 50% OHC loss

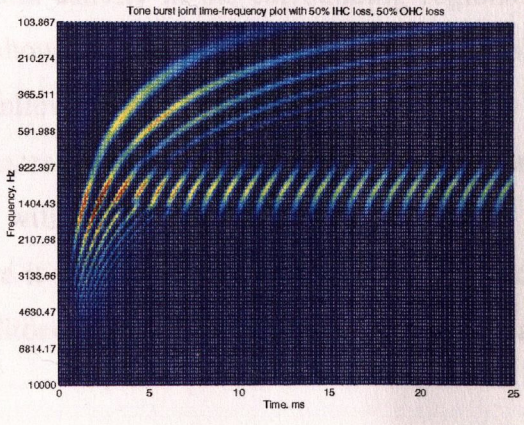


Figure 6.4d 50% IHC loss, 50% OHC loss

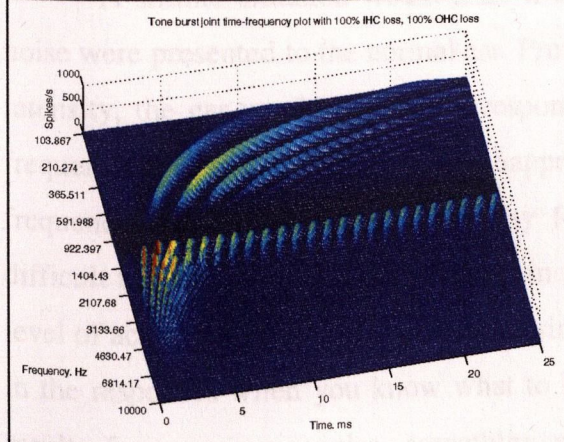


Figure 6.4d 100% IHC loss, 100% OHC loss

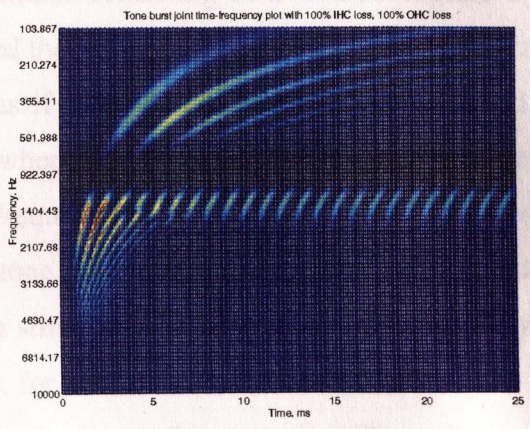


Figure 6.4e 100% IHC loss, 100% OHC loss

Figure 6.4 Tone burst response with 0%, 50% and 100% IHC & OHC loss at 1kHz.

In particular it can be seen that as before with inner hair cell damage the spontaneous firing is absent at 100% inner hair cell damage. Similarly the neural firing associated with upward spread of excitation is reduced below the levels experienced in Figure 6.2f and even in Figure 6.3f. Comparing the neural firing with 100% outer hair cell loss and 100% inner hair cell loss (Figure 6.4f) to the neural firing with 0% outer hair cell loss and 100% inner hair cell loss (Figure 6.2f) shows how important a role the outer hair cells play in the overall level of neural activity. In general, however, the cases of total outer and/or inner hair cell loss are not particularly interesting as they generally lead to the system not responding at all at the damaged location.

The case of medium level outer and inner hair cell loss is of much more importance as this is the situation in which frequency smearing mainly occurs. With this type of hair cell loss the auditory system can still respond at the centre frequency where the damaged hair cells reside but its ability to generate a neural response at the centre frequency is such that it cannot out-perform the unimpaired neural activity in the surrounding regions. In terms of hearing impaired listeners it has been shown that they have elevated cochlear filter skirts with mild degrees of sensitivity loss, followed by a progressive widening of the auditory filter bandwidth as hearing loss increases [115]. In such a case, a stimulus with frequency content at the impaired frequency will not be successfully generated in the neural activity of the auditory system, therefore it will be unable to accurately respond to the frequency content of the stimulus. The auditory model replicates this type of scenario with 50% outer and inner hair cell loss.

A similar situation would arise if a stimulus consisting of a tone multiplied by a noise were presented to the normal ear. Provided the noise was of sufficient bandwidth and intensity, the ear would generate a response as if it were being stimulated at a band of frequencies. This is effectively what happens when the auditory system is suffering from frequency smearing; it loses its ability for frequency selectivity. It can be somewhat difficult to identify frequency smearing in the tone burst responses, as there is quite a low level of activity taking place with such a simple stimulus. The effects can be best identified in the responses when you know what to look for and are able to view and compare the results from various angles, something which is impossible in print. The effects can however be quite easily seen when considering synthetic speech in which formant transitions occur as frequency smearing has a very noticeable effect on formants and their transitions.

6.3.2 Response of the Frequency Smeared Model to Synthetic Speech

The response of the auditory model to synthetic speech is generated by stimulating the auditory model with the synthetic syllable /da/ as specified in Chapter 4. The response of the auditory model is generated for varying levels of inner, outer and inner & outer hair cell damage. Specifically the responses are generated for 0%, 50%, 75% and 100% of each type of hair cell loss focused at 1.6kHz. The formant at 1.6kHz is specifically targeted to clearly illustrate the effects the impairment has on the spectral content of the neural firing.

6.3.2.1 Response of the Model with Inner Hair Cell Loss to Synthetic Speech

The first case to consider is that of inner hair cell loss at 1.6kHz. The response is initially generated without any inner hair cell loss to provide a reference with which to compare the response for 50%, 75% and 100% inner hair cell loss. Figures 6.5a-d show the neural responses of the auditory model. The viewpoint is somewhat unusual as it looks back along the time axis; this is purely to assist identification of the relevant features.

Considering Figure 6.5a and Figure 6.5b the affect of the inner hair cell loss flattens out the peak response at 1.6kHz to approximately the same level as the activity above and below the peak response, especially in the steady state regions. The peak transient response is still however just at a high enough level to be distinguishable from the surrounding neural firing. As the level of inner hair cell loss progresses from 50% to 75% in Figure 6.5c, clearly both the steady state and transient peak firing are at the same levels as the surrounding firing rates. The reduction in the firing rate due to 75% loss of the inner hair cells at 1.6kHz has the effect that the formant at this frequency is no longer being reproduced in the neural activity of the model. Where initially the formant was being reproduced in a highly tuned band of firing at 1.6kHz, after 75% impairment it is being reproduced in a broad band of neural firing. In this broad band of firing the peak that should represent the formant is at the same level of firing as the surrounding activity thus the system is unable to accurately respond to the spectral content at 1.6kHz. The auditory model is therefore suffering from frequency smearing. Further progression onto 100% inner hair cell loss shows a reduction in frequency smearing as the neural activity dies away altogether; see Figure 6.5d. If the hair cells are totally destroyed then they cannot fire at all in this region. It is for this reason that the neural firing at 1.6kHz decays below the spontaneous level.

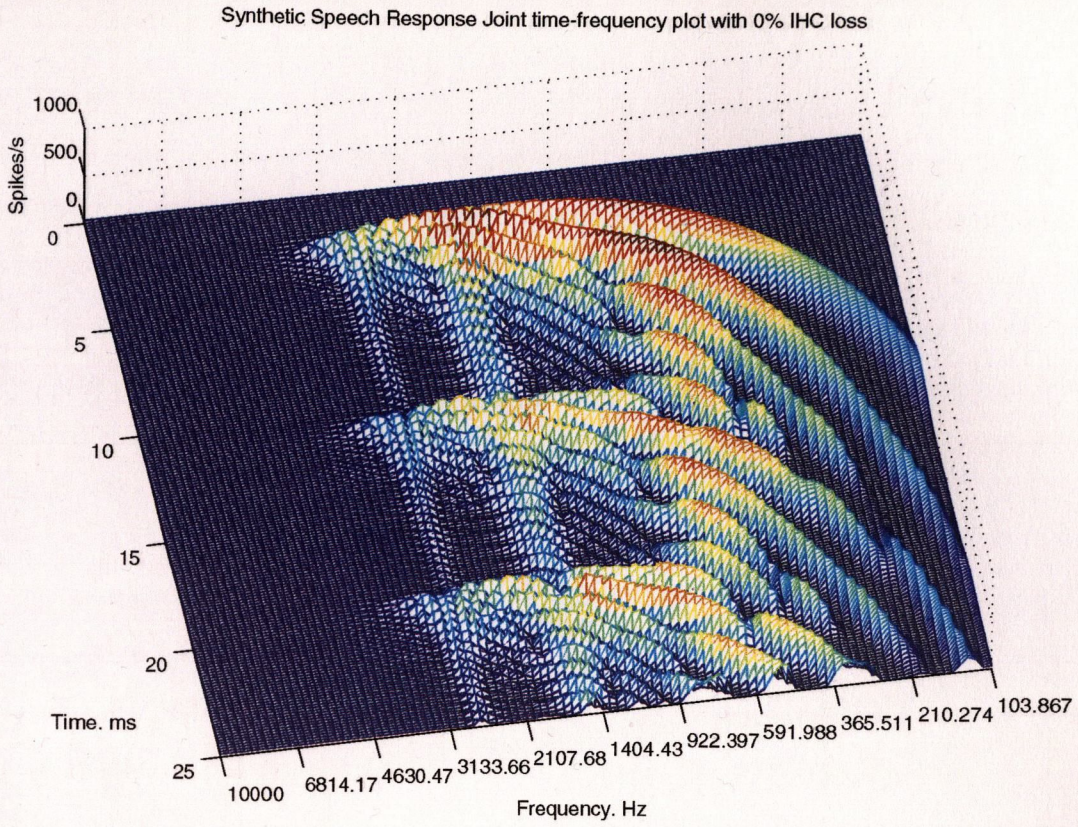


Figure 6.5a Response to Synthetic Speech with 0% IHC Loss.

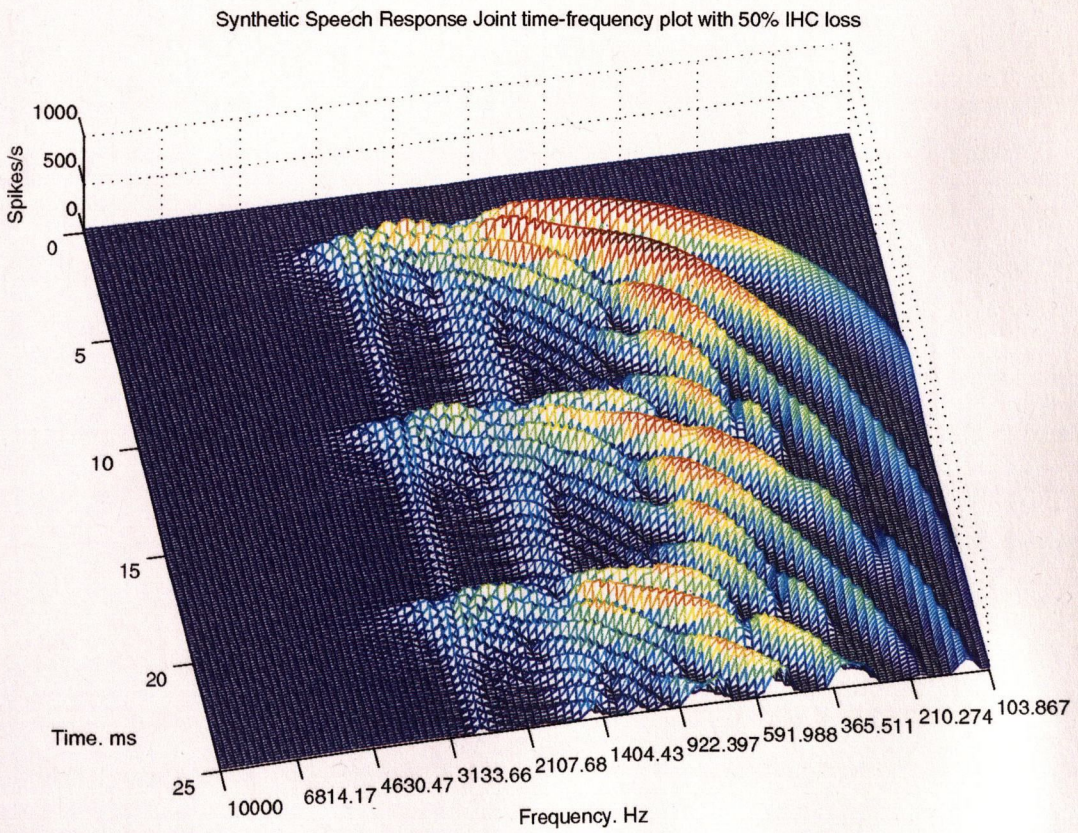


Figure 6.5b Response to Synthetic Speech with 50% IHC Loss.

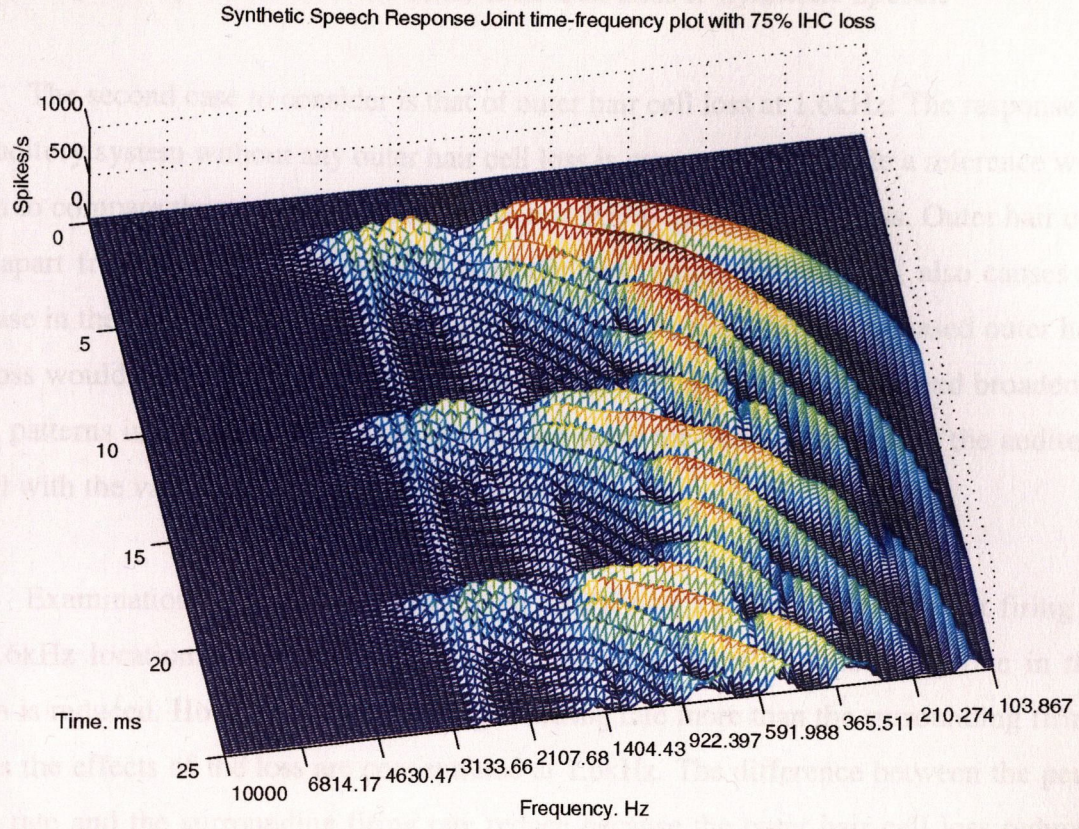


Figure 6.5c Response to Synthetic Speech with 75% IHC Loss.

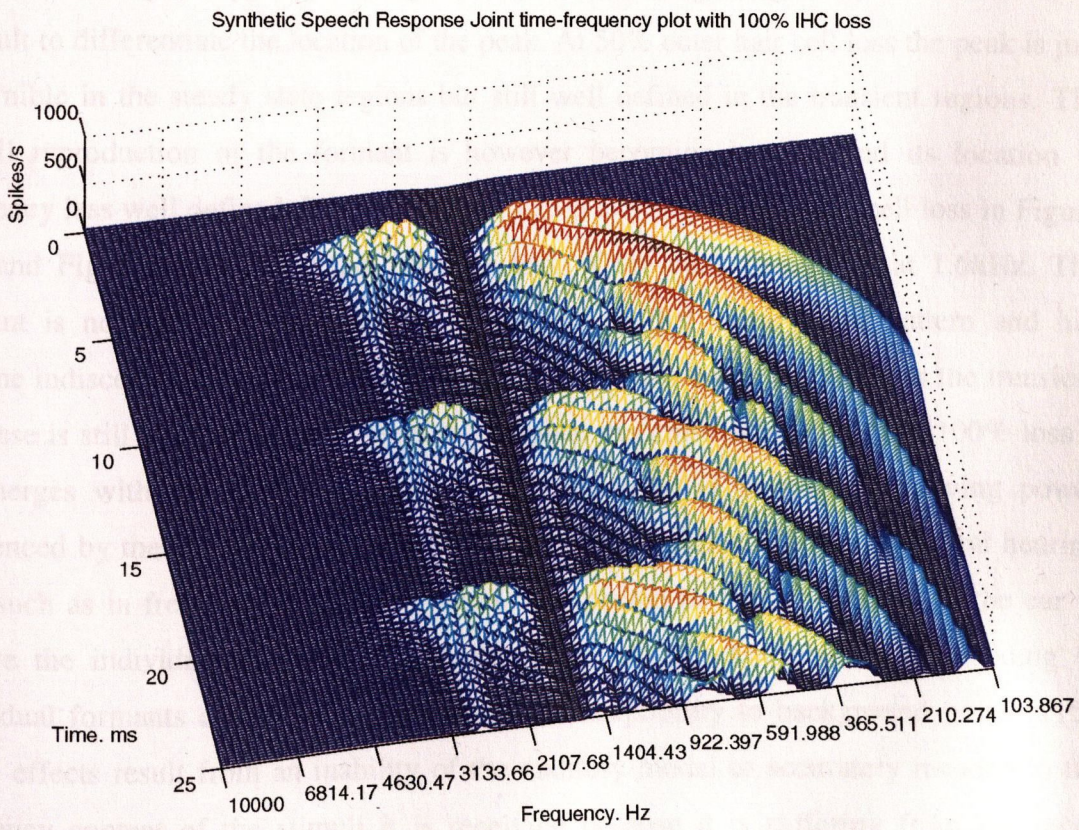


Figure 6.5d Response to Synthetic Speech with 100% IHC Loss.

6.3.2.2 Response of the Model with Outer Hair Cell Loss to Synthetic Speech

The second case to consider is that of outer hair cell loss at 1.6kHz. The response of the auditory system without any outer hair cell loss is generated to provide a reference with which to compare the response for 50%, 75% and 100% outer hair cell loss. Outer hair cell loss, apart from causing a non-linear loss of dynamic range compression, also causes an increase in the threshold and broadening of the cochlear tuning curves. Increased outer hair cell loss would therefore be expected to cause reduced inner hair cell firing and broadened firing patterns in the area affected. Figures 6.6a-d show the neural response of the auditory model with the various amounts of outer hair cell loss.

Examination of Figure 6.6a and Figure 6.6b shows a reduction in the peak firing at the 1.6kHz location. With 50% outer hair cell loss the overall neural firing rate in the region is reduced. However, it affects the peak firing rate more than the surrounding firing rate as the effects of the loss are concentrated at 1.6kHz. The difference between the peak firing rate and the surrounding firing rate reduce because the outer hair cell loss reduces the gain at the peak of the tuning curve thus reducing the inner hair cell response at the peak and the broadening of the tuning curves enhances the surrounding response. The net effect is that the peak appears to merge with the surrounding firing and it is therefore more difficult to differentiate the location of the peak. At 50% outer hair cell loss the peak is just discernible in the steady state regions but still well defined in the transient regions. The overall reproduction of the formant is however becoming broader and its location in frequency less well defined. Progression on to 75% and 100% outer hair cell loss in Figure 6.6c and Figure 6.6d clearly indicates the loss of the peak firing rate at 1.6kHz. The formant is no longer being faithfully reproduced in the neural firing pattern and has become indiscernible from the surrounding firing. At 75% outer hair cell loss the transient response is still at a high enough level to stimulate the inner hair cells but at 100% loss it too merges with the surrounding firing. This reduction in frequency resolving power (evidenced by the change in shape of the effective critical bands in sensorineural hearing loss, such as in frequency smearing) would be expected to limit the ability of the ear to resolve the individual formants of speech signals, to affect normal 'place' coding of individual formants and to cause an increased susceptibility to background noise [115]. These effects result from an inability of the auditory model to accurately respond to the frequency content of the stimuli it is receiving because it is suffering from frequency smearing due to outer hair cell loss.

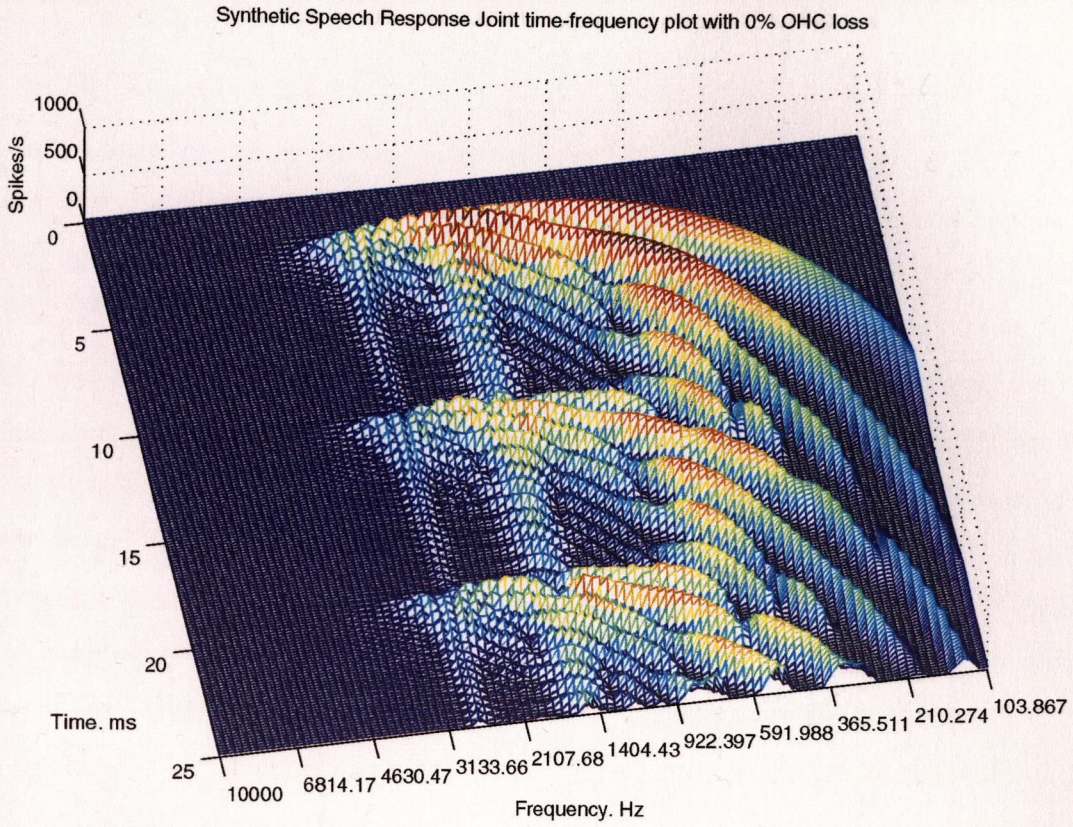


Figure 6.6a Response to Synthetic Speech with 0% OHC Loss.

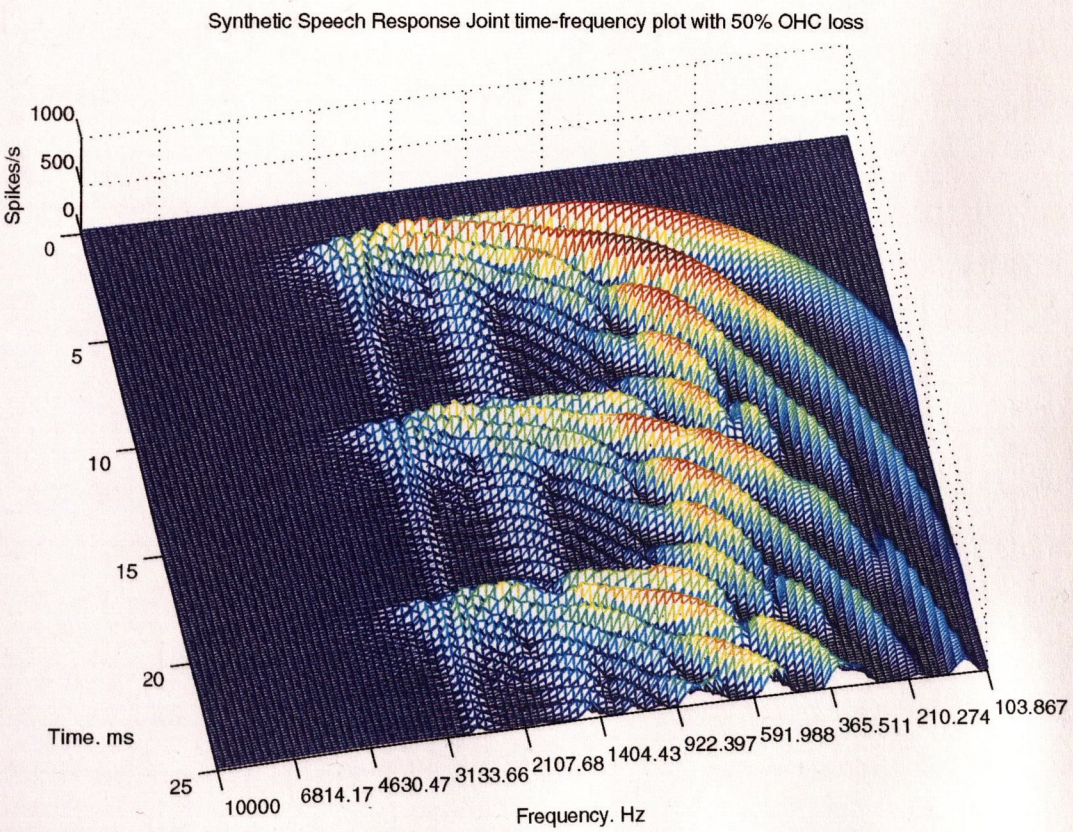


Figure 6.6b Response to Synthetic Speech with 50% OHC Loss.

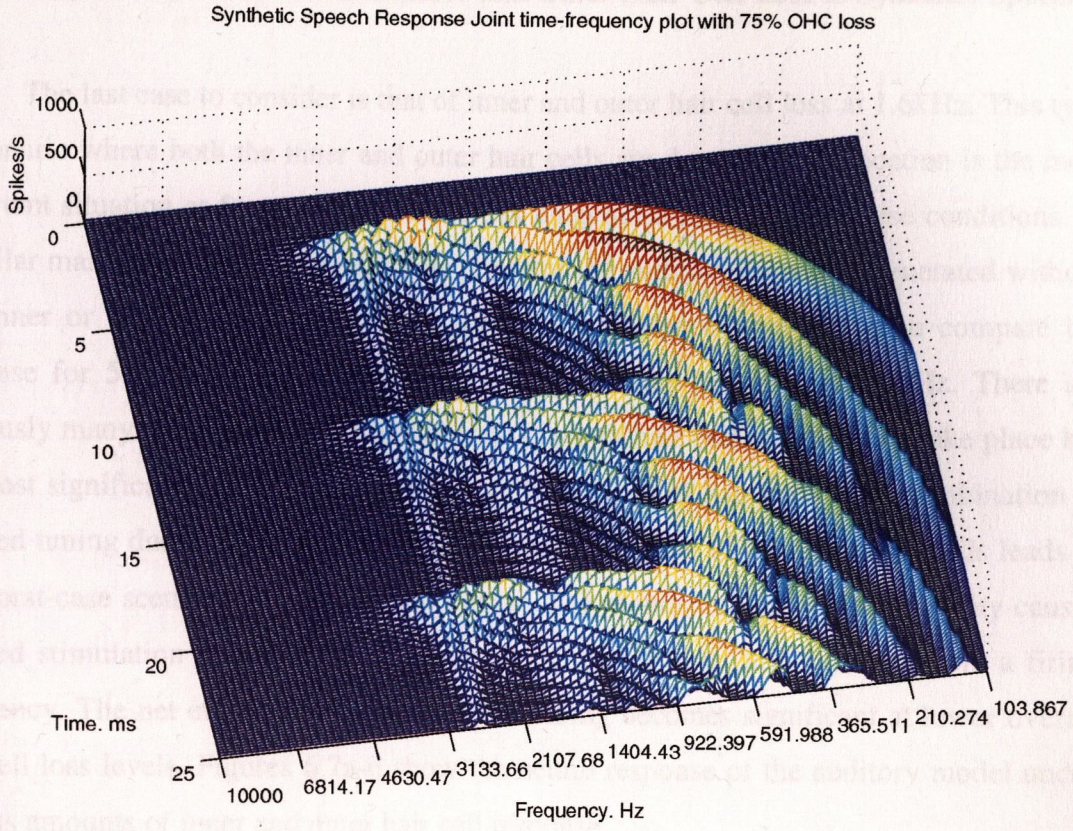


Figure 6.6c Response to Synthetic Speech with 75% OHC Loss.

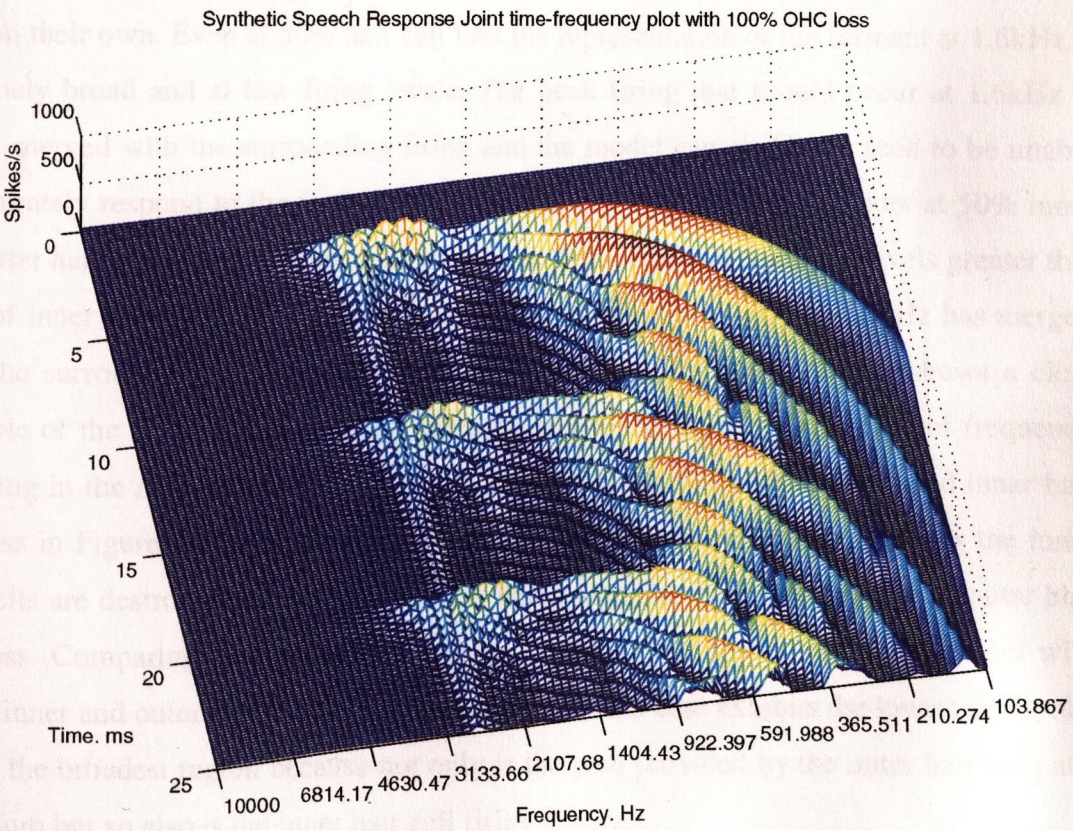


Figure 6.6d Response to Synthetic Speech with 100% OHC Loss.

6.3.2.3 Response of the Model with Inner and Outer Hair Cell Loss to Synthetic Speech

The last case to consider is that of inner and outer hair cell loss at 1.6kHz. This type of scenario where both the inner and outer hair cells are damaged at a location is the most important situation as frequency smearing has its greatest effect under these conditions. In a similar manner as before, the neural response of the auditory system is generated without any inner or outer hair cell damage to provide a reference with which to compare the response for 50%, 75% and 100% inner and outer hair cell loss at 1.6kHz. There are obviously many combinations of the different types of hair cell loss that can take place but the most significant is that of uniform loss of both types of hair cells. The combination of reduced tuning due to outer hair cell loss and reduced firing of the inner hair cells leads to the worst-case scenario of frequency smearing. The reduced outer hair cell activity causes reduced stimulation of the inner hair cells, which themselves are suffering from a firing deficiency. The net effect is that frequency smearing becomes significant at lower overall hair cell loss levels. Figures 6.7a-d show the neural response of the auditory model under various amounts of inner and outer hair cell response.

Considering the neural firing patterns in Figures 6.7a-d, both inner and outer hair cell loss have an even more significant effect than impairment of the inner or outer hair cells on their own. Even at 50% hair cell loss the representation of the formant at 1.6kHz is extremely broad and at low firing levels. The peak firing that should occur at 1.6kHz is totally merged with the surrounding firing and the model can clearly be seen to be unable to accurately respond to the formant as expected. The firing rate that occurs at 50% inner and outer hair cell loss is at a level similar to the firing rate that occurs at levels greater that 75% of inner or outer hair cell loss alone. Even the transient firing at 1.6kHz has merged with the surrounding neural activity. Progression onto 75% hair cell loss shows a clear example of the formant being totally smeared and is the most obvious case of frequency smearing in the results. Considering the worst-case scenario of 100% outer and inner hair cell loss in Figure 6.7d, the neural firing reduces below the spontaneous rate as the inner hair cells are destroyed and the threshold is being significantly increased by the outer hair cell loss. Comparing the response in Figure 6.7d with the responses of the model with 100% inner and outer hair cell loss alone shows that this case exhibits the lowest firing rate across the broadest region because not only is the gain provided by the outer hair cells at a minimum but so also is the inner hair cell firing ability.

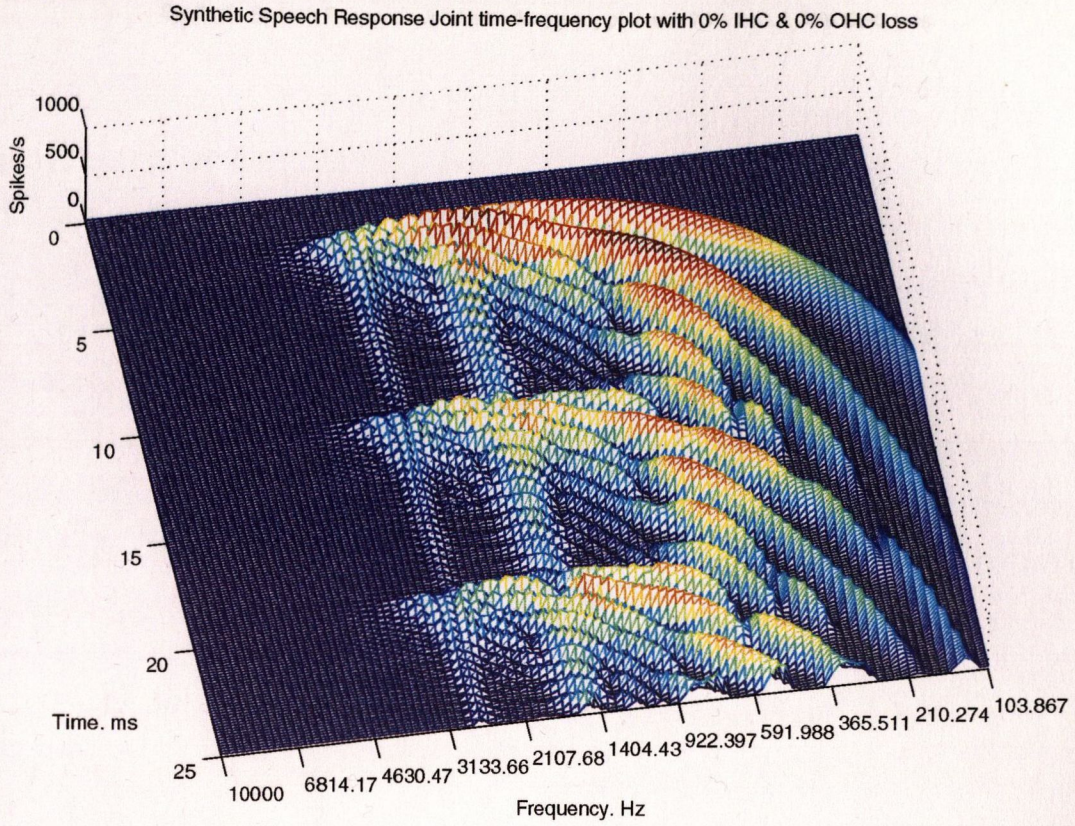


Figure 6.7a Response to Synthetic Speech with 0% IHC & 0% OHC Loss.

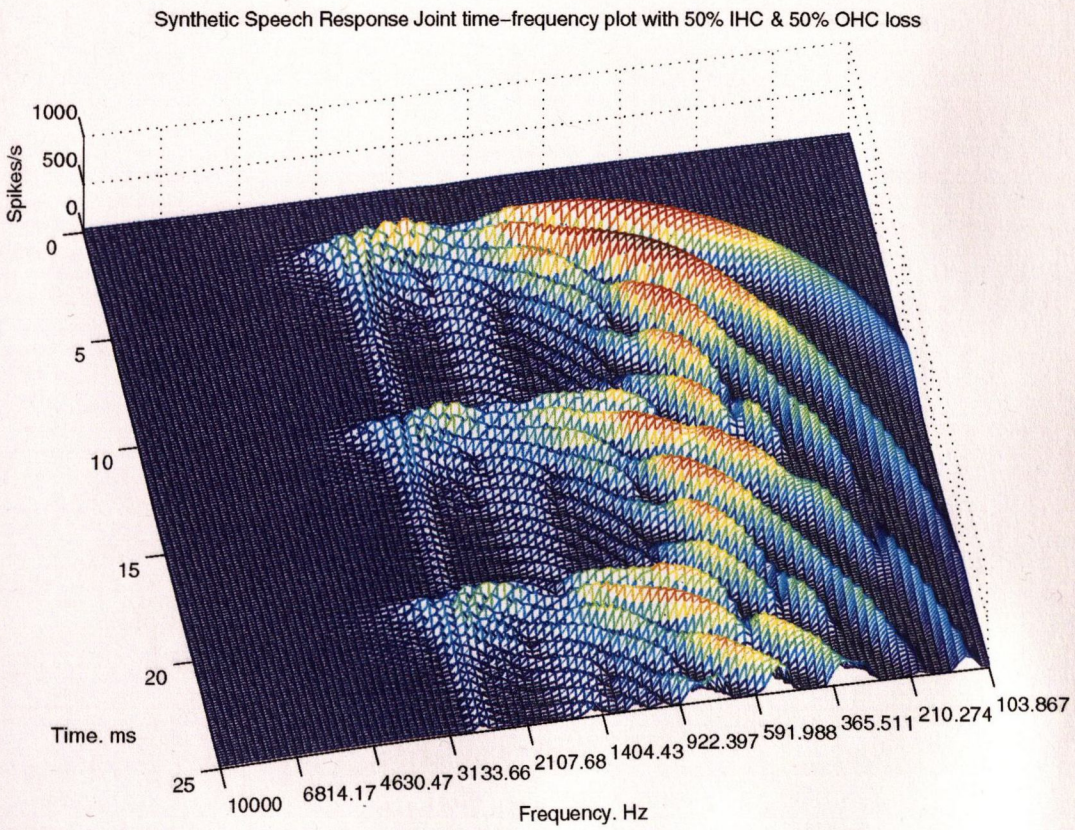


Figure 6.7b Response to Synthetic Speech with 50% IHC & 50% OHC Loss.

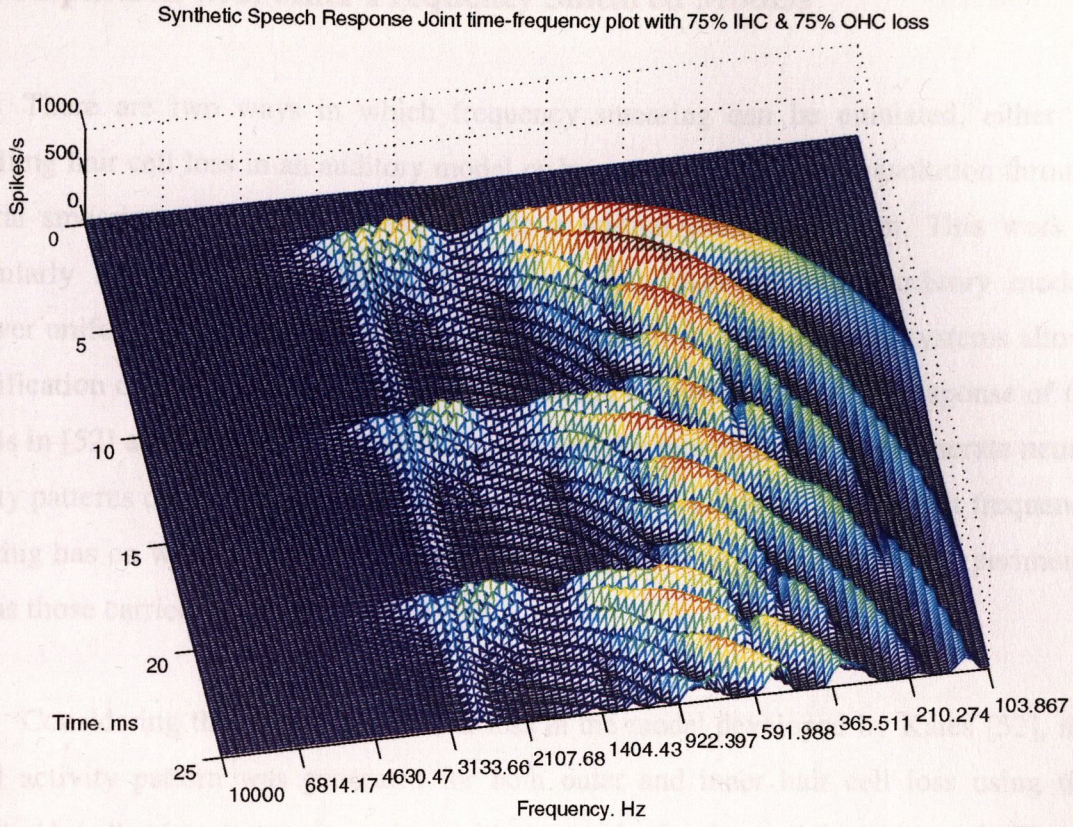


Figure 6.7c Response to Synthetic Speech with 75% IHC & 75% OHC Loss.

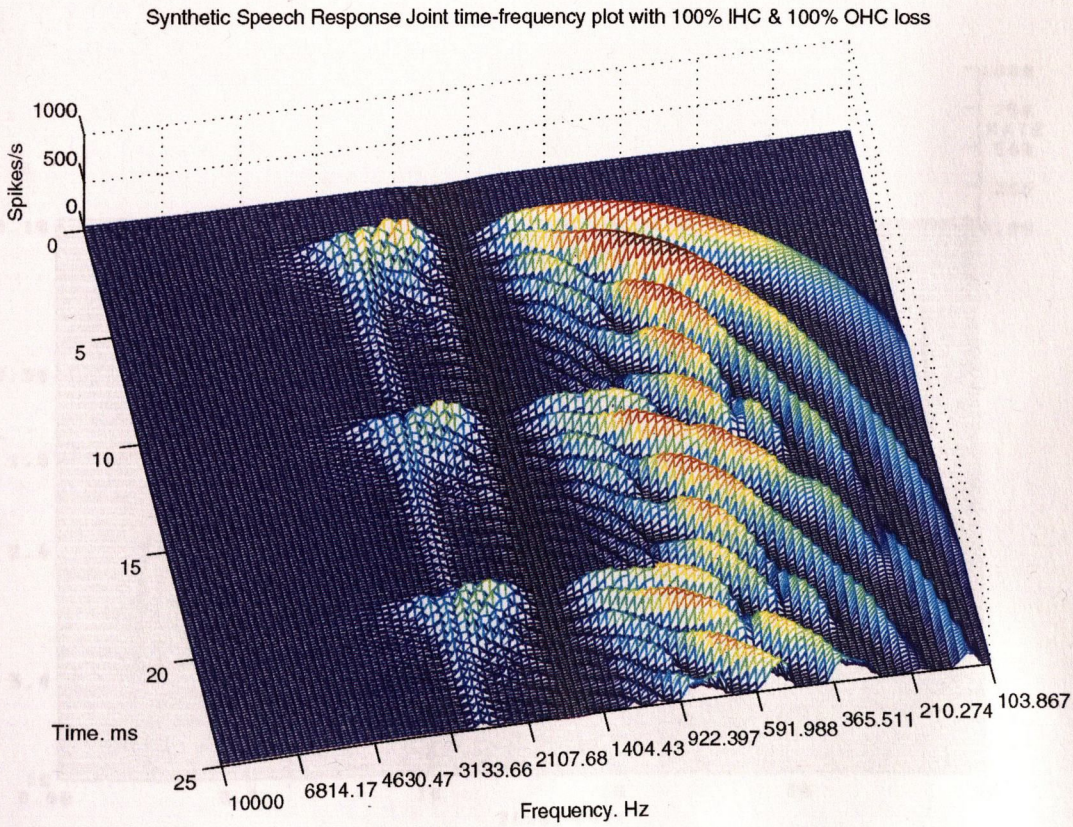


Figure 6.7d Response to Synthetic Speech with 100% IHC & OHC Loss.

6.4 Comparison with other Frequency Smeared Models

There are two ways in which frequency smearing can be emulated, either by modelling hair cell loss in an auditory model or by reducing frequency resolution through spectral smearing of a stimulus presented to a normal auditory system. This work is particularly focused on the frequency smearing that takes place in auditory models however uniform spectral smearing of stimuli presented to normal auditory systems allows quantification of the effects smearing has on acoustic comprehension. The response of the models in [52] and [17] with hair cell loss are especially interesting as they generate neural activity patterns comparable to the outputs of the model in this work. The effect frequency smearing has on word recognition in normal auditory systems is evaluated in experiments such as those carried out in [102].

Considering the effects of hair cell loss in the model developed by Kates [52], the neural activity pattern was generated for both outer and inner hair cell loss using the syllable /da/ allowing direct comparison with the results for the model in this work. Figure 6.8 shows the response with complete *outer hair cell loss*. The outer hair cell action in this model is modelled as level dependent feedback and it is this feedback that is reduced.

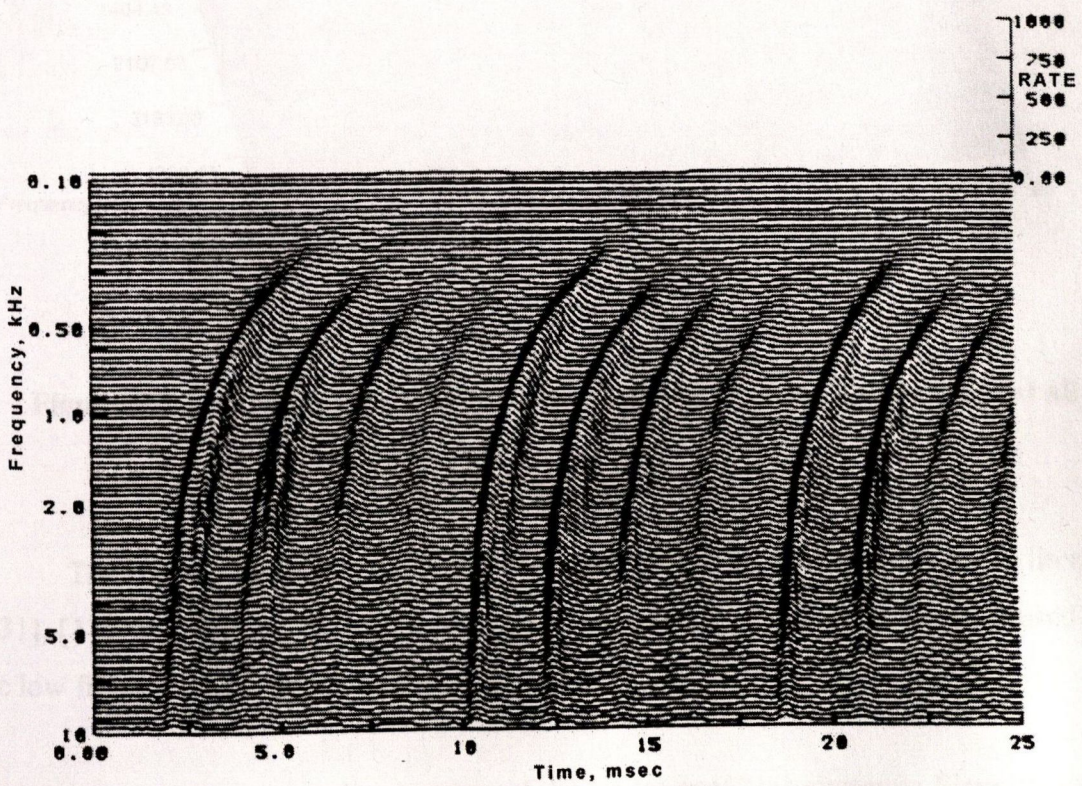


Figure 6.8 Response to synthetic speech with complete outer hair cell loss [52]

Examination of Figure 6.8 shows that loss of the outer hair cell feedback causes the mechanical system in the model to become completely linear and exhibit broadening of the cochlear filters with a resulting reduction in frequency selectivity. The detailed timing information can be seen to be completely lost and the firing rate becomes totally synchronised with the first formant. This is to be expected as the threshold shift at higher frequencies is greater than at lower frequencies; thus the higher formants will become almost entirely lost [131]. The model shows the required spectral smearing effects but the neural activity at higher frequencies is somewhat elevated compared to the response of the model in this work to the same stimulus under the same impairment conditions, see Figure 6.9.

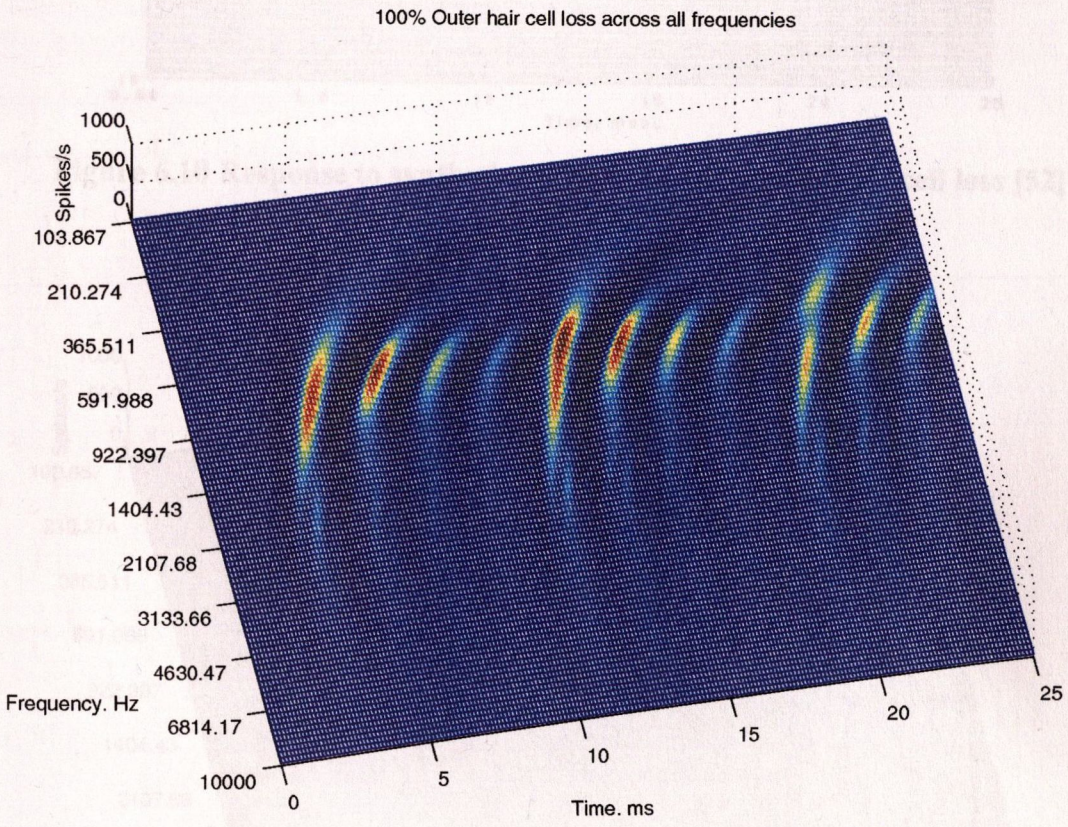


Figure 6.9 Response of the auditory model with total outer hair cell loss at all frequencies.

The response in Figure 6.9 is in better agreement with recent results in the literature [131], [17], exhibiting significant reduction in high frequency response synchronised with the low frequency first formant in the stimulus.

In a similar manner the response of the model in [52] is shown in Figure 6.10 with a tenfold increase in the stereocilia resistance in the inner hair cells.

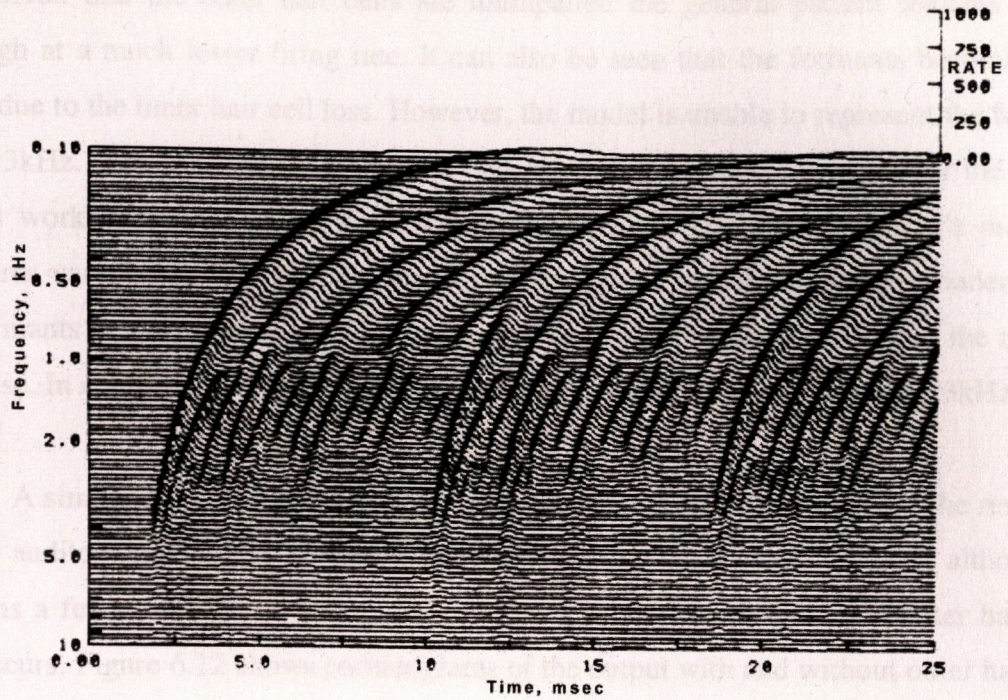


Figure 6.10 Response to synthetic speech with 90% inner hair cell loss [52]

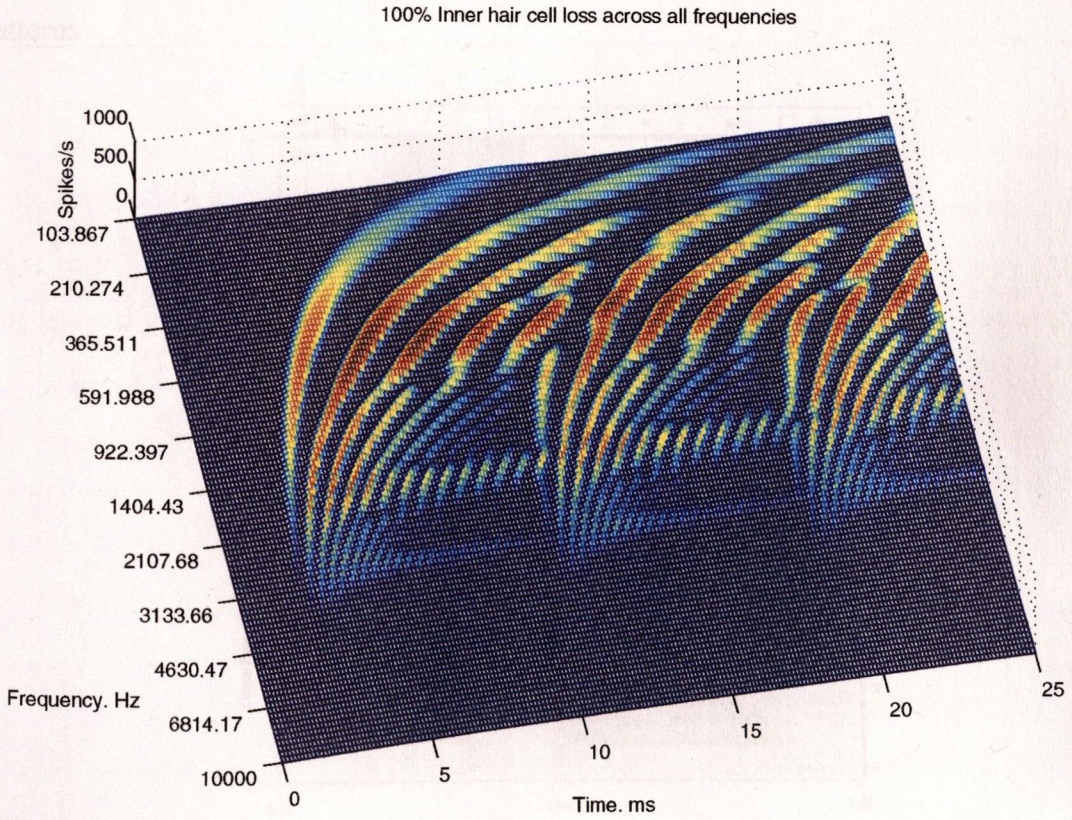


Figure 6.11 Auditory model response with 90% inner hair cell loss at all frequencies.

The increase in resistance is justified for modelling inner hair cell damage as breakage and blockage of the stereocilia will reduce the number of available conductance channels. A tenfold increase in resistance is equated to approximately 90% inner hair cell

loss. Given that the outer hair cells are unimpaired the general pattern remains intact, although at a much lower firing rate. It can also be seen that the formants become quite broad due to the inner hair cell loss. However, the model is unable to represent the formant above 3kHz. 90% inner hair cell loss has been modelled for all frequencies for the model in this work and its response is shown in Figure 6.11. The current model's response compares very favourably with the response from Kate's model exhibiting broadening of the formants at all frequencies and an overall lowering of the firing activity in the models response. In contrast the model here is capable of generating the formant above 3kHz.

A similar comparison can be made with the more recent work in [17]. The response of this auditory model was generated for the utterance "greasy wash", which although it contains a few syllables, illustrates the concept of what takes place when outer hair cell loss occurs. Figure 6.12 shows cochleograms of the output with and without outer hair cell loss. A cochleogram is the auditory image generated by performing temporal integration of the neural firing rate [17], [55] and helps identification of features in the neural firing patterns

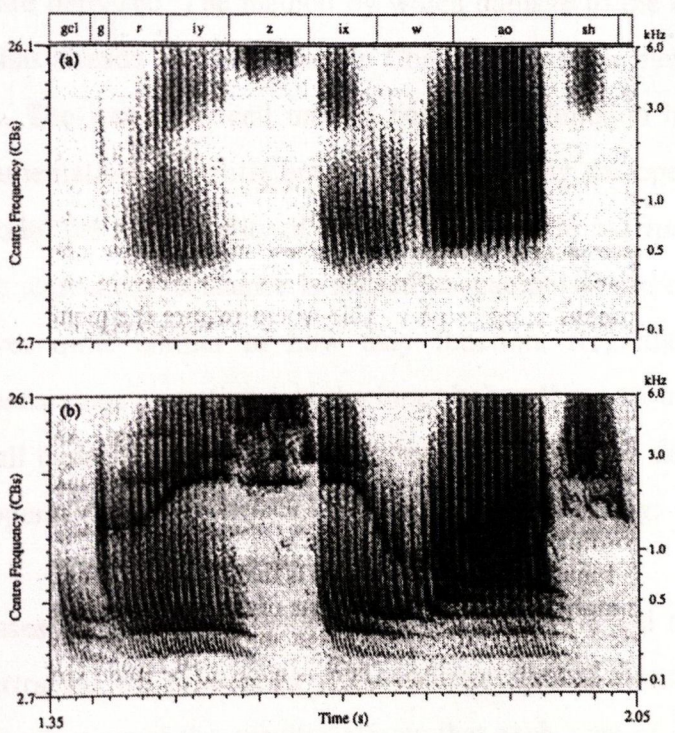


Figure 6.12 Cochleograms for the utterance "greasy wash" with (above) and without (below) total outer hair cell loss respectively [17].

It is worth noting that in cochleograms the darker the area the more intense the response in that region. A detailed analysis of the data in Figure 6.12 is give in [17] but it

suffices to say that the unimpaired response shows sharp tuning and good synchronisation with the formants in the speech as seen in the lower diagram. In contrast the impaired model exhibits reduced tuning, reduced high frequency response and loss of synchronisation in the response to the formants within the speech, as seen in the upper diagram in Figure 6.12. Complex speech such as this has not been presented to the model in this work[2] but the responses of the model to syllables have been shown to exhibit similar responses in both the impaired and unimpaired modes of operation. Qualitative comparison shows good overall agreement between the responses from the model in this work and from the model in [17].

6.5 Summary and Conclusions

In summary, the unimpaired auditory model can be made to emulate an auditory system suffering from frequency smearing by incorporating damage to the inner and outer hair cells. There are three possible scenarios in which frequency smearing may occur, when inner hair cell damage occurs, when outer hair cell damage occurs and when both types of hair cells are damaged. The method by which damage to the outer hair cells has been implemented has already been specified in Chapter 5 and this method has been used in this Chapter too. The inner hair cell models being used function in slightly different ways as they fundamentally target different ways of emulating the operation of the inner hair cell. In one model the permeability of the cell is altered to manipulate the cell firing and in the other the stereocilia resistance is altered to manipulate the cell firing. The two models are however quite similar in how they function. Regardless of the model, specification of a percentage operation sets the inner hair cell model used to fire at some percentage of its full capacity. The effect of implementing hair cell damage is to get the auditory model to operate as if it is suffering from frequency smearing.

All three cases of hair cell damage have been emulated and the response of the model has been extracted for tone burst and synthetic speech stimuli. The tone burst responses allow examination of the specific effects that each type of damage causes and requires specific knowledge of the effects of frequency smearing to identify its presence. The necessary details have been specifically targeted in the Introduction. The response of

---

[2] The computational burden that would arise using such stimuli would incur an extremely long simulation time given the current implementation of the model. This issue is addressed in the final Chapter.

the model to synthetic speech illustrates the effects each type of damage has on the formants that make up the speech. In particular the effects of frequency smearing can be most easily seen in these responses especially with 75% inner and outer hair cell loss. In all cases, but most especially in the case of damage to both the inner and outer hair cells, the model emulates an auditory system suffering from frequency smearing.

In general, the work in the literature focuses on presenting spectrally smeared stimuli to normal listeners to try and quantify the effects spectral smearing would have on impaired listeners. There are few auditory models that deal with hair cell damage but the model in this work compares very favourably to those that do allow for such impairment. In particular, the works of [52] and [17] have been examined here, as they allow direct comparison with the responses in this work. Both models exhibit the effects that hair cell damage would have on the response of the auditory system. However, the model in this work best compares with the recent work in [17], in keeping with the current knowledge in this area. In general the model in this work shows very good correlation with the known responses of the auditory system suffering from frequency smearing due to hair cell damage.

Chapter 7 - Recruitment Compensation

7.1 Introduction

The development of a model for recruitment within the scope of a model for the unimpaired auditory system not only allows exploration of recruitment and its effects on the auditory response but also facilitates the development of compensatory mechanisms for this impairment. Such compensatory functions must however be considered in terms of the causes of the impairment and especially the effects introduced by the impairment, as it is these effects that must be counteracted.

In depth coverage of the causes and effects of recruitment have been dealt with in the Chapter 5. However in summary, loss of the outer hair cell function within the cochlea causes an *increase in the threshold of audibility* and the *non-linear loss of dynamic range compression of the signal excitation stimulating the inner hair cells*, resulting in an *abnormally rapid growth of loudness level.* This essentially identifies two key factors that must be overcome to compensate for recruitment. Provision must be made to compensate for the increase in the threshold of audibility and to re-introduce compression of the dynamic range of the signal excitation stimulating the inner hair cells. Only when both these measures are in place can the effects of recruitment be counteracted.

Of the two effects mentioned, a frequency selective increase in auditory threshold is by far the most well known auditory problem. It can be simply identified by means of an audiogram and compensated for by means of selective amplification. Until the recent introduction of digital hearing aids this generally involved amplification of all signals being presented to the auditory system, including amplification of noise providing only minimal benefit. Current digital hearing aids, however, are capable of frequency selective amplification, providing amplification only at required frequencies and capable of rejecting background noise, thus providing a far superior solution. Provided the increase in threshold is solely due to conductive hearing loss or reduced inner hair cell function such mechanisms can be of considerable benefit. Provision for frequency selective amplification, to counteract the effects of an increase in auditory threshold, in a compensatory function for recruitment is a relatively simple matter when dealt with in a digital signal-processing framework.

However, if loss of outer hair cell function accompanies the loss of inner hair cell function, which is often the case, then recruitment will also occur and previously mentioned amplification mechanisms are unable to compensate for the loss of dynamic range compression. In fact they antagonise the problems introduced by recruitment by providing excessive amplification. It is well known that many users of hearing aids tend to reduce the recommended level of amplification provided by their hearing aids, complaining of sounds being excessively loud. It can be easily understood why this would be the case in the event that an individual was suffering from a non-linear loss of dynamic range compression on top of an increase in auditory threshold. Recruitment by its very nature is much more difficult to identify and quantify, requiring more elaborate testing procedures and mechanisms to counteract its effects. Chapter 8 deals with the issues of identification and quantification of recruitment; in this Chapter a mechanism for recruitment compensation is developed.

In general, there has been considerable effort in the development of automatic gain control systems capable of rejecting noise and providing reasonable benefit to the user. The past decade has seen a significant move towards the development of compression devices to provide enhanced automatic gain control avoiding such features as clipping etc [8]. These systems typically involve some form of linear compression to compress the amplified signal presented to the auditory system, while ensuring not to exceed the uncomfortable threshold of the user. The very latest compensation mechanisms use wide dynamic range compression which map the input signal range into the residual dynamic range of the hearing impaired listener [156]. Such systems use linear compressive functions to perform this mapping on the basis that signals above a certain limit must be audible and all signals near the upper limit must be maximally compressed to avoid exceeding the uncomfortable threshold level, everything in between is linearly compressed between these two limits. However, outer hair cell loss causes a non-linear loss of dynamic range compression and these systems cannot hope to adequately cater for such a loss. These systems do not directly target the problem of recruitment (or even claim to do so); in fact, there are very few such devices that even attempt to do this. The form of compression above will provide minimal assistance to a recruitment sufferer and will in general antagonise the problem by providing either too much compression, in which case the recruited system will not respond, or too little, in which case the recruited system will over respond. It is out of this background that the development of the following compensation system arises.

The general methodology taken here is best considered in terms of its theoretical approach and implementation of this approach. Evaluating the impairment reveals that a loss of outer hair cell function in a region of the basilar membrane introduces detrimental effects for the reception of signals that maximally stimulate these regions. This inability of the auditory system to respond in such a region can also be considered from the point of view of the stimulating signal. A stimulus having frequency content within the frequency band that maximally excites the impaired region will not be responded to in the appropriate manner within such a region. If this frequency band is isolated within the stimulating signal and appropriate corrective measures introduced into the stimulating signal (to counteract the inability of the impaired region to respond appropriately to it) before the stimulating signal interacts with the impaired region, then it may be possible to cause an interaction of the stimuli and the impaired region that better approximates the response that would have occurred in the region if it were not damaged. This analysis suggests two primary system requirements.

1. A mechanism for efficiently dividing a signal into band-limited sub-signals and reconstructing a corrected signal from these sub-signals prior to presentation of the corrected signal to the auditory system, subject to the conditions of being distortion free and within a tolerable delay.

2. A mechanism for introducing corrective measures into the band-limited sub-signals prior to reconstruction of the stimulating signal.

Subsequent sections in this Chapter deal with the development of these two sub-systems and an evaluation of the resulting compensation mechanism under various conditions of impairment. Clearly both the defect mechanisms and compensation mechanisms require characterisation based on initial conditions provided by evaluation of the impairment; these matters are dealt with in Chapter 8.

7.2 Compensation Function Development

Before developing the compensatory measures for the effects introduced by recruitment, several other factors must be considered. These factors primarily come to light when more detailed consideration is given to the specific measures that must be implemented, namely the degree to which each effect may exist and the range of

frequencies over which they may have their influence must be identified. This is the type of information that emerges from an evaluation of a particular instance of an impairment; for the purpose of development, this information serves only to specify the ranges of operation over which compensatory measures must be capable of operating. Targeting of compensatory measures to counteract a specific instance of an impairment requires quantification of the increase in auditory threshold and the loss of compression, along with the frequency ranges affected. This information is then used in the characterisation of an instance of the compensatory function.

As indicated in the introduction, the recruitment compensation function consists of two main systems, for clarity it is best to deal with them separately.

7.2.1 Generation of Band-Limited Sub-Signals

Loss of outer hair cell function in the auditory system is generally not an isolated incidence and more usually several regions are affected, thus several bands of frequencies suffer from the effects of recruitment to differing degrees. Considering this situation, the system for isolating the impaired sub-bands must be able to divide the frequency range of the incoming stimuli into several bands so that a suitable corrective measure may be applied to each sub-band signal. At the far end, this system must then recombine the compensated sub-signals with the remainder of the signal content that does not require correction, to produce a signal that the auditory system can respond to appropriately even though it is impaired. Significantly, the reconstructed signal must be free from *alias-distortion*, *amplitude-distortion* and *phase-distortion*, all of which place stringent requirements on the system. Such a system capable of isolating specific sub-bands, each of different bandwidth, would require a custom designed section for each band. The variable bandwidths and responses would require extremely complex reconstruction algorithms, the implementation of which would be extremely costly in integrated circuit real estate and would not be particularly flexible. It would not be feasible to develop such an un-restricted system given the constraints of hearing aid size and cost. Another approach is therefore required that places restrictions on the implementation of the system for sub-band generation and signal reconstruction so that a compromise may be reached.

Taking these matters into consideration, if the bandwidth of all the frequency bands is fixed then the frequency range of the auditory system can be uniformly divided up into

frequency bands using an *analysis-synthesis* system based on the *modulation* of a *prototype filter*, provided the perfect-reconstruction (PR) property is enforced. (PR implies a system is free from alias, amplitude and phase distortion). Provided the frequency range of the auditory system is sufficiently divided then the impaired bands can be isolated and suitable corrective measures applied only to those bands requiring attention. At the output the synthesis bank then allows for efficient reconstruction of the signal components. The resulting computational complexity of such a system reduces to the cost of implementing the prototype plus the modulation costs.

The use of modulated prototype filters to efficiently generate analysis-synthesis filter banks is relatively well known, being introduced in the mid-seventies [157], but has primarily been used in the development of *approximate* reconstruction systems, better known as *pseudo quadrature mirror filter banks* (pseudo-QMF banks). Of particular interest for the current situation is the more recent extension of cosine modulated pseudo-QMF banks to cosine modulated perfect reconstruction systems [158], [159], [160], [161].

A brief introduction will follow to highlight the important features of cosine modulated pseudo-QMF banks supplemented by a more thorough explanation of how an approximate reconstruction system can be extended to form a cosine modulated perfect reconstruction system. This will then be followed by an examination of how the cosine modulated PR system is used to generate band-limited sub-signals in the recruitment compensation mechanism.

7.2.1.1 Cosine modulated pseudo-QMF banks

In general, QMF banks are generated using exponential modulation of a prototype low pass filter (cut off frequency $2\pi/M$, M is the number of sub-bands) as this can be efficiently implemented using Discrete Fourier Transforms (DFT) thus forming a DFT filter bank [162]. Considering the analysis filter bank and applying this approach, if $h[n]$ represents the real prototype coefficients and $h_k[n]$ the resulting coefficients of the exponential modulation of $h[n]$, then

$$h_k[n] = h[n]e^{j\frac{2\pi kn}{M}} = h[n]W^{-kn} \tag{7.1}$$

$$H_k(z) = H(zW^k) \tag{7.2}$$

$$W = W_M = e^{-j\frac{2\pi}{M}} \tag{7.3}$$

where $H(z)$ and $H_k(z)$ are the z-domain representations of the prototype filter and its uniformly shifted versions generated through exponential modulation and M represents the number of sub-channels to be generated. This method generates filter coefficients that are generally complex even if the prototype is real, therefore the outputs in the sub-channels can be complex signals even if the original input is real. Direct implementation of such a DFT filter bank forces the prototype to be of length M and hence have very limited cutoff and stop band attenuation. If the prototype filter is represented in the form of a type 1 polyphase decomposition then

$$H(z) = \sum_{l=0}^{M-1} z^{-l} E_l(z^M) \tag{7.4}$$

$$E_l(z) = \sum_{n=-\infty}^{\infty} h[Mn + l] z^{-n}, \text{ with } 0 \le l \le M-1 \tag{7.5}$$

The resulting uniformly shifted versions of the prototype can then be expressed as

$$H_k(z) = H(zW^k) = \sum_{l=0}^{M-1} z^{-l} W^{-kl} E_l(z^M) \tag{7.6}$$

This type of system can be quite efficiently implemented in the form shown in Figure 7.1.

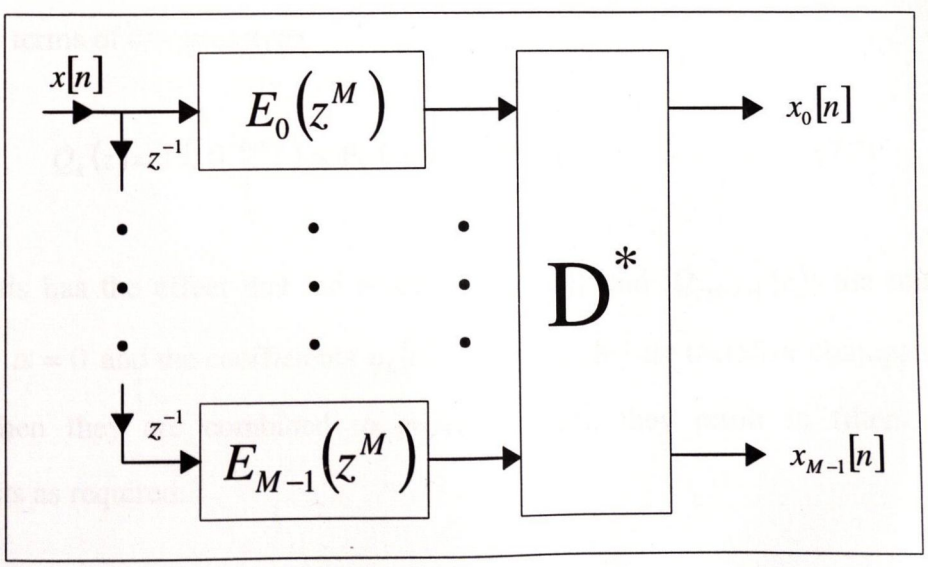


Figure 7.1 Polyphase DFT analysis filter bank.

In Figure 7.1, D^* represents the conjugate of an MxM DFT matrix generated from elements $W_M^{km}, 0 \le k,m \le M-1$. Using this polyphase form allows the length of the prototype to be increased beyond length M by extension of the M polyphase elements, thus allowing sharper cut off and increased stop band attenuation yet producing only M sub-bands. It is from this background that cosine modulation arises.

A class of filters with real coefficients can be derived using cosine modulation rather than exponential modulation, by firstly obtaining $2M$ complex filters using exponential modulation and combining appropriate pairs [161]. Replacing M with $2M$ in the above equations generates a uniform-DFT filter bank with 2M filters. Clearly the dimensions of the conjugate DFT matrix must increase to be $2Mx2M$ and the number of polyphase components to $2M$ components. Let $p[n]$ represent a real coefficient low pass prototype filter (cut off frequency $\pi/2M$) and $p_k[n]$ the exponentially modulated versions of $p[n]$. Given that $p[n]$ is a real coefficient prototype, its frequency response, $|P(e^{j\omega})|$, is symmetric with respect to $\omega = 0$. It can therefore be readily shown that the modulation responses are right shifted versions of the prototype by an amount $k\pi/M$ and that the responses $|P_k(e^{j\omega})|$ and $|P_{2M-k}(e^{j\omega})|$ are images of one another with respect to $\omega = 0$. Thus $|P_k(e^{j\omega})|$ and $|P_{2M-k}(e^{j\omega})|$ are ideal candidates to be combined to get real coefficient filters of pass bandwidth $2\pi/M$. Making a change of $z \to zW^{0.5}$ in the z-domain representation of the prototype implements a right shift of $\pi/2M$ and ensures the resulting filters, after combination, have the same bandwidths. The complex filters $Q_k(z)$can be defined in terms of this prototype.

$$Q_k(z) = P(zW^{k+0.5}) \text{ with } 0 \le k \le 2M-1 \qquad (7.7)$$

This has the effect that the responses $|Q_k(z)|$ and $|Q_{2M-1-k}(z)|$ are images with respect to $\omega = 0$ and the coefficients $q_k[n]$ and $q_{2M-1-k}[n]$ are therefore conjugates of each other. When they are combined to generate $h_k[n]$, they result in filters with real coefficients as required.

7.2.1.2 Approximate cancellation of alias distortion

It is clear that the bandwidth of each of the outputs of the analysis bank would be M times narrower than the bandwidth of the original signal; therefore the signal can be decimated M-fold to reduce the overall computational load and make the structure more efficient. Typically the output of each of the analysis filters $H_k(z)$ is decimated and then interpolated before entering the synthesis filters $F_k(z)$; however, this process is not without its drawbacks as it gives rise to alias components $H_k(zW_M^l)X(zW_M^l)$ and their images, which must be removed by the synthesis filters. The synthesis filters must also retain the unshifted components $H_k(z)X(z)$. To accomplish this, the pass band of the synthesis filters must in general coincide with those of the analysis filters.

Note: Generally the noble identities are used when such filters are being implemented in polyphase form thus increasing the efficiency of the system further. See Appendix A1.

Convenient expressions from [161] can be used to specify both the analysis and synthesis filters in terms that allow specification of necessary and sufficient conditions for approximate elimination of aliasing. By defining

$$U_k(z) = c_k P(zW^{k+0.5}) = c_k Q_k(z) \qquad 0 \le k \le M-1 \qquad (7.8)$$

$$V_k(z) = c_k^* P(zW^{-(k+0.5)}) = c_k^* Q_{2M-1-k}(z) \qquad 0 \le k \le M-1 \qquad (7.9)$$

the M modulated analysis filters can then be expressed as

$$H_k(z) = a_k U_k(z) + a_k^* V_k(z) \text{ with } 0 \le k \le M-1 \qquad (7.10)$$

Using this notation, the synthesis filters, which are also derived from the prototype filter, can be expressed as

$$F_k(z) = b_k U_k(z) + b_k^* V_k(z) \text{ with } 0 \le k \le M-1 \qquad (7.11)$$

where a_k, b_k and c_k are unit magnitude constants used to characterise the above expressions. The outputs of the synthesis filters have components $H_k(zW_M^l)X(zW_M^l)$ for $0 \le l \le M-1$, but with high enough stop band attenuation only some of these components are significant, the remainder although not zero are ignored, hence the "approximate alias cancellation". With some significant mathematical endeavour it can be shown that the outputs of $F_k(z)$ and $F_{k-1}(z)$ share the common alias component $X(zW_M^{\pm k})$ and that applying the following constraint causes cancellation of the significant alias components when the outputs are added [161].

$$a_k b_k^* = -a_{k-1} b_{k-1}^* \quad \text{with } 1 \le k \le M-1 \tag{7.12}$$

7.2.1.3 Cancelling phase distortion

Phase distortion can be completely eliminated if the *resulting combination* of the analysis and synthesis filter pairs are designed to be linear phase. This can be accomplished if the relationship in Equation (7.13) can be satisfied

$$F_k(z) = z^{-(N-1)} H_k(z^{-1}) \tag{7.13}$$

where N is the length of the prototype filter. Considering that both $H_k(z)$ and $F_k(z)$ are defined in terms of $U_k(z)$ and $V_k(z)$ it can be shown that if $P(z)$, the prototype filter, is constrained to be linear phase, that is $p[n] = p[N-1-n]$, then defining c_k as in Equation (7.14) ensures that both $U_k(z)$ and $V_k(z)$ are constrained to be linear phase.

$$c_k = W^{(k+0.5)N/2} \tag{7.14}$$

Consequently it can be shown that Equation (7.13) can be fulfilled if b_k is constrained as

$$b_k = a_k^* \tag{7.15}$$

Finally a constraint must be placed on a_k. Considering Equation (7.15) and the alias cancellation constraint in Equation (7.12) it can be shown that the necessary constraint on

a_k must take the form

$$a_k = e^{j\theta_k} \text{ where } \theta_k = (-1)^k \frac{\pi}{4} \text{ and } 0 \le k \le M-1 \qquad (7.16)$$

7.2.1.4 Minimising amplitude distortion

Amplitude distortion arises if the combination of the analysis and synthesis filter pairs are not exactly flat. This form of distortion cannot be eliminated but it can be reduced to an acceptable level through minimisation of an objective function representing the non-flatness of the analysis/synthesis pairs during the design of the prototype. The development of such objective functions can be found in the literature and it suffices to say that this form of distortion can be reduced to reasonable levels in pseudo-QMF banks. Of primary importance will be the *elimination* of this form of distortion in the PR-system as this would cause significant problems in the design of the corrective measures being applied in each sub-band.

7.2.1.5 Modulated Filter Expressions

Given all the constraints required to minimise the various distortions, closed form expressions for the modulated analysis and synthesis filters can be constructed based on these constraints and Equations (7.10-11). After some mathematical manipulation the following expressions can be derived

$$h_k[n] = 2p[n]\cos\left(\frac{\pi}{M}(k+0.5)\left(n-\frac{N-1}{2}\right)+\theta_k\right) \qquad (7.17)$$

$$f_k[n] = 2p[n]\cos\left(\frac{\pi}{M}(k+0.5)\left(n-\frac{N-1}{2}\right)-\theta_k\right) \qquad (7.18)$$

Clearly it can be seen that the ultimate process of constructing the analysis and synthesis filter banks can be achieved through cosine modulation of a single prototype filter [161].

Sections 7.2.1.1 to 7.2.1.5 are designed to give a brief explanation of the process of cosine modulated pseudo-QMF banks and are by no means a thorough treatment of the

subject; they are, however, required to lay the foundation for the development of cosine modulated PR-systems; significantly greater detail can be found in [163] and [161].

7.2.1.6 Cosine modulated perfect reconstruction systems

It has recently been shown that FIR filter banks satisfying perfect reconstruction can be obtained by the modulation of a linear phase prototype, length $N = 2mM$ (M is the number of sub-bands and m is an arbitrary positive integer), when certain constraints are imposed on the prototype [158], [164]. To begin with, the development of pseudo-QMF banks is a precursor to the development of PR-systems as it is the specification of the constraints that enhances a pseudo-QMF system to become a PR-system. Considering Equations (7.4-6), the prototype can be expressed as

$$H(z) = \sum_{q=0}^{2M-1} \sum_{p=0}^{m-1} h[q+2pM] z^{-(q+2pM)} = \sum_{q=0}^{2M-1} z^{-q} G_q\left(z^{2M}\right) \qquad (7.19)$$

where $G_q(z)$ are the $2M$, type 1 polyphase components of the prototype filter. Considering Equations (7.16 & 7.17) the modulation component can be summarised as

$$C_{k,l} = 2\cos\left(\frac{\pi}{M}(k+0.5)\left(l-\frac{N-1}{2}\right)+(-1)^k\frac{\pi}{4}\right) \qquad (7.20)$$

Note: $C_{k,(l+2pM)} = (-1)^p C_{k,l}$ due to the periodicity of the cosine modulation.

Combining Equations (7.17, 7.19 & 7.20), the analysis filters can be expressed as

$$H_k(z) = \sum_{q=0}^{2M-1} \sum_{p=0}^{m-1} h[q+2pM] C_{k,(q+2pM)} z^{-(q+2pM)} \qquad (7.21)$$

Ultimately this can be reduced to the following expression, which relates the modulated analysis filters to the polyphase elements of the prototype filter in the z-domain.

$$H_k(z) = \sum_{q=0}^{2M-1} C_{k,q} z^{-q} G_q\left(-z^{2M}\right) \qquad (7.22)$$

$$\tilde{G}_k(z)G_k(z)+\tilde{G}_{M+k}(z)G_{M+k}(z)=\frac{1}{2M}, \quad 0\le k\le M-1 \quad (7.27)$$

That is

"$E(z)$ is lossless if and only if appropriate pairs of polyphase components of $H(z)$ are power complementary", [160]. (See Appendix A2)

Once the polyphase component matrix is lossless a suitable synthesis bank can always be found such that the overall analysis/synthesis system guarantees perfect reconstruction. Clearly therefore, the polyphase component matrices in pseudo-QMF banks are only approximately lossless resulting in approximate reconstruction. This fundamentally means that the pseudo-QMF system can be enhanced to be a perfect reconstruction system if the prototype filter being used to generate such a system is designed so that appropriate pairs of polyphase components are power complementary (PC).

7.2.1.7 Design procedure for a prototype filter guaranteeing perfect reconstruction

The previous sections have laid out the criterion for the generation of a cosine modulated perfect reconstruction QMF bank. All that is required to do now is design the prototype given the above constraints and decide on an implementation strategy. Considering the design of the prototype, there are certain requirements both for a PR-system and for these systems to be useful in the development of a recruitment compensation system.

- Clearly, appropriate polyphase pairs must be power complementary to generate a PR-system.
- High stop band attenuation is required for suitable isolation of the sub-bands.
- The transition width between the pass band and stop bands of the prototype must be sufficiently narrow to ensure minimal cross-over between adjacent pass bands in the filter banks and guarantee no cross-over of non-adjacent bands within the pass band of any non-adjacent filters in the filter banks.

It is generally agreed in the literature that these types of goals can best be achieved through optimisation. There are several ways of performing this optimisation most of which involve some form of spectral factorisation whereby one of the polyphase

components of the PC pairs $\{G_k(z) \;\; G_{M+k}(z)\}$ is optimised and the other is computed by spectral factorisation [166]. These methods are known to be extremely computationally expensive and significant work has been carried out into the development of alternate methods [160], [167], [168] and [169]. Primary among these efforts has been the exploitation of two channel lossless lattices due to the fact that any FIR bounded real pair of transfer functions $\{P(z) \;\; Q(z)\}$ satisfying

$$\tilde{P}(z)P(z)+\tilde{Q}(z)Q(z)=1, \forall z \tag{7.28}$$

can always be realised as a non-recursive, cascade of two-channel lossless lattice structures [170] of the form shown in Figure 7.2.

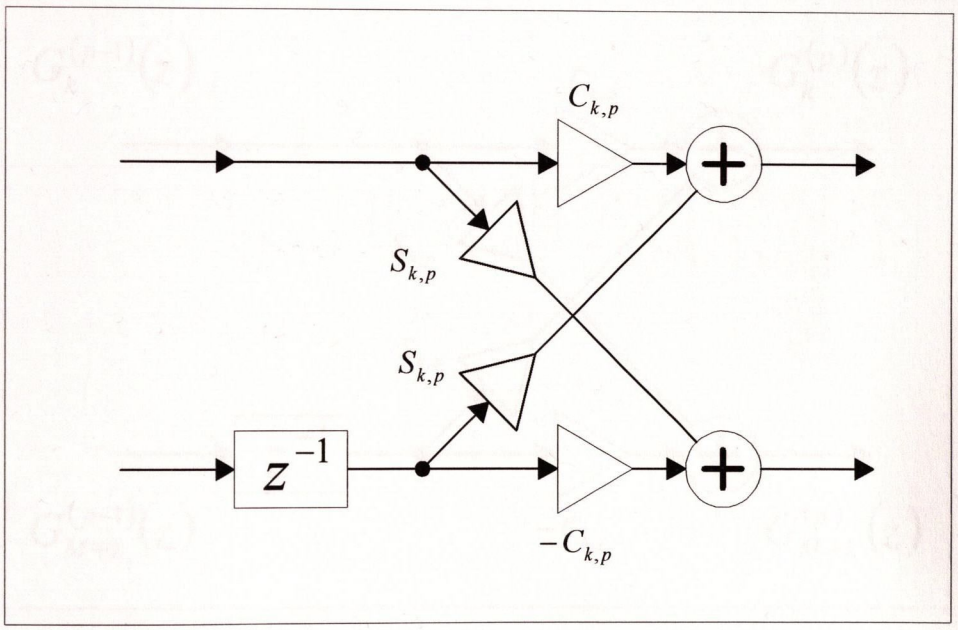


Figure 7.2 4-multiplier 2-channel lossless lattice structure.

In Figure 7.2 $S_{k,p}=\sin(\theta_{k,p})$ and $C_{k,p}=\cos(\theta_{k,p})$ where $\theta_{k,p}$ is the characterising parameter for the lattice section. Given that the lossless lattice structure guarantees that Equation (7.28) holds, if the polyphase component pair $\{G_k(z) \;\; G_{M+k}(z)\}$ are designed using 2-channel lossless lattice structures then they inherently must be power complementary, as required. The optimisation process for such structures directly involves the lattice parameter and by the very nature of these structures the PC property of the polyphase components is therefore guaranteed regardless of the process used to optimise $\theta_{k,p}$. Since the PC property is guaranteed it does not have to be included in the

design of the prototype as it is a beneficial side effect of lossless lattices. It is worth noting that this is still the case in the presence of coefficient quantisation [160], [171].

Considering the prototype more closely, if the number of sub-bands M is restricted to be *even* and since the prototype is required to have linear phase, the impulse response $h[n]$ will be symmetric, therefore only half the coefficients have to be determined. In the lossless lattice framework this means that only half the number of lattices need to be optimised, as the rest of the polyphase components can be easily determined. Therefore only $M/2$ lattices need to be optimised to generate sufficient polyphase components to completely specify the prototype filter. Figure 7.3 illustrates generation of the k^{th} pair of PC polyphase components.

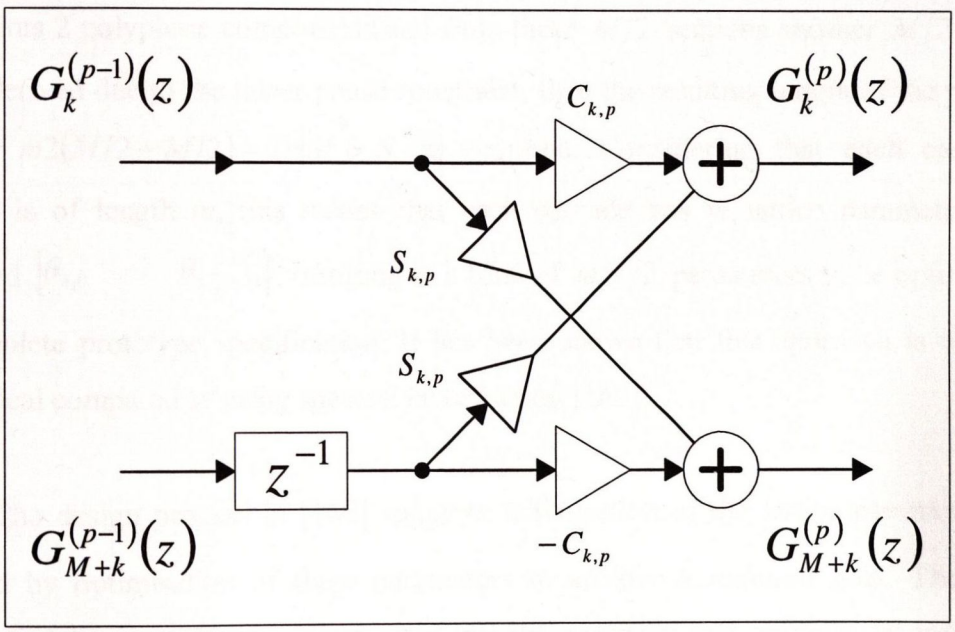


Figure 7.3 Generation of the k^{th} pair of PC polyphase components using lossless lattices.

It can be clearly seen from Figure 7.3 that cascading such sections together increases the length of the polyphase components and thus the prototype. Examination of this cascaded approach shows that initialisation of each of the k^{th} component pairs $G_k^0(z)$ and $G_{M+k}^0(z)$ must occur to begin the process. These functions are initialised as

$$G_k^0(z) = \cos(\theta_{k,0}) \tag{7.29}$$

$$G_{M+k}^0(z) = \sin(\theta_{k,0}) \tag{7.30}$$

Considering the lossless lattice in Figure 7.3, the general transfer function for the lattice can be expressed in matrix form, from [160], as

$$\begin{bmatrix} G_k^{(p)}(z) \\ G_{M+k}^{(p)}(z) \end{bmatrix} = \begin{bmatrix} \cos(\theta_{k,p}) & \sin(\theta_{k,p}) \\ \sin(\theta_{k,p}) & -\cos(\theta_{k,p}) \end{bmatrix} \begin{bmatrix} G_k^{(p-1)}(z) \\ z^{-1} G_{M+k}^{(p-1)}(z) \end{bmatrix} \tag{7.31}$$

where

$$1 \le p \le m-1 \text{ and } 0 \le k \le \left[\frac{M}{2}\right] - 1 \tag{7.32}$$

Addition of each additional lattice section increases the order of the prototype by 1. If there are m lattice sections cascaded and $M/2$ sets of these sections, where each section implements 2 polyphase components and from these $M/2$ sections another $M/2$ sections can be derived due to the linear phase constraint, then the resulting length of the prototype will be $m2(M/2+M/2) = 2mM = N$ as required. Considering that each cascade of sections is of length m, this means that each cascade has m lattice parameters to be optimised $\left[\theta_{k,0} \quad \cdots \quad \theta_{k,(m-1)}\right]$, resulting in a total of $mM/2$ parameters to be optimised for the complete prototype specification. It has been shown that this approach is extremely economical compared to using spectral factorisation [160].

The design process in [160] suggests initialisation of the lattice parameters $\theta_{k,p}$ followed by optimisation of these parameters to achieve a required goal. The general initialisation begins with an almost ideal impulse response and optimises it towards the required goals using a two-step optimisation process. Using standard optimisation routines [172] the lattice parameters are optimised to minimise the stop band energy of the prototype. This requires evaluation of the stop band energy at each iteration of the optimisation process.

$$\Phi_1 = \int_{(\pi/2M)+\delta)}^{\pi} \left|H(e^{j\omega})\right|^2 d\omega \tag{7.33}$$

This results in a *minimum energy* solution. The resulting prototype is then used as a starting point for the minimisation of the maximum stop band response to yield a *minimax*

solution. This requires evaluation of the following function at each iteration of the optimisation process.

$$\Phi_2 = \max\left|H\left(e^{j\omega}\right)\right|, \omega \in \{(\pi/2M + \delta), \pi\} \tag{7.34}$$

where δ represents the width of the transition band between the edge of the pass band and the edge of the stop band. These processes do indeed generate prototype filters capable of generating PR-systems. However the resulting PR-systems have some undesirable features, as follows

1. The filter banks resulting from the final prototype have significant overlap of adjacent pass bands and even allow overlap of non-adjacent pass bands.
2. The stop band attenuation of the filters in the filter banks is insufficient even with high order prototypes.
3. The optimisation for high order prototypes is extremely computationally intensive.

These problems would appear to make the resulting systems unsuitable for use in the current application; however, this is not the case. The main problem lies in the initialisation process, which result in unsatisfactory optimisation performance and unsuitable prototype filters.

Generally, each lattice parameter can take on any value in the range $\{-\pi, \pi\}$ due to the periodicity of the trigonometric functions in the lattices. When the number of lattice parameters to be optimised is small, i.e. the prototype is relatively short, the optimisation routines can converge relatively easily. The reason for this is that the small parameter space generates energy and minimax spaces that are relatively simple, with very few local minima between which the optimisation routines can get *trapped*. The optimisation routines used [172] are designed to search for the global minimum of functions, in this case the energy and minimax functions, and be able to recognise and *escape* from local minima to eventually converge on the global minimum. However, the optimisation routines are based on a quasi Newton algorithm that uses gradients to converge on the global minimum and it is important that the initial starting point be sufficiently near to the global minimum to successfully converge on it. When the prototype is short, it is sufficient to initialise the lattice parameters to generate an almost ideal response to ensure the starting point is

sufficiently near the global minimum for the optimisation routines to converge on the global minimum.

However when the size of the parameter space increases, the sizes of the energy and minimax spaces increase and become increasingly complex, containing significant numbers of local minima between which the optimisation can become trapped and produce a false global minimum. Initialising the lattice parameters to generate an almost ideal response does not start the optimisation routines sufficiently near to the global minimum so as to avoid being trapped between local minima, the number of which is significantly increased. A more elaborate initialisation scheme is therefore required to start the optimisation process sufficiently close to the global minimum for it to be located.

The way to avoid the problem of false convergence is to have better initialisation of the optimisation process. If the prototype filter is designed closer to a more realistic filter then the optimisation process becomes more of a refining process. A closer examination of the optimisation process shows that the effective result of the optimisation routines, in terms of the prototype filter response, is to manipulate the lattice parameters so that the initial response is honed into a response that has a minimum finite transition band, flat pass band response and minimum stop band energy to give maximum stop band attenuation. The greater the number of parameters the longer the prototype will be and hence the better the response. However, there is a price to pay for this, increased convergence complexity and longer times for convergence.

There are several standard algorithms available for the design of real FIR filters; however, it is specifically require that the following can be specified for the prototype

- The pass band edge
- The stop band edge
- Stop band attenuation
- Pass band ripple

Considering the above requirements, the Parks-McClellan algorithm based on the Remez exchange algorithm and Chebyshev approximation generates the best fit between the desired and actual frequency responses. This algorithm generates filters that are optimal in the sense that they minimise the maximum error between the desired frequency response

and actual frequency response [173]. They do however generate filters with equiripple behaviour. This however is not a problem as the optimisation process can minimise the pass band ripple to within tolerable levels. It is therefore possible to generate an initial prototype filter using the Parks-McClellan algorithm and using this as the starting point for the optimisation process, produce a prototype with the desired characteristics.

7.2.1.8 Prototype filter generation

The sampling rate for the system is 40kHz. Therefore considering the Nyquist condition the highest frequency that can be validly represented within the system is 20kHz which is sufficiently high to cover the frequency range of the human auditory system. Arbitrarily dividing this frequency range up into 1kHz bands means that $M = 20$ in analysis/synthesis system. Considering Figure 7.4.

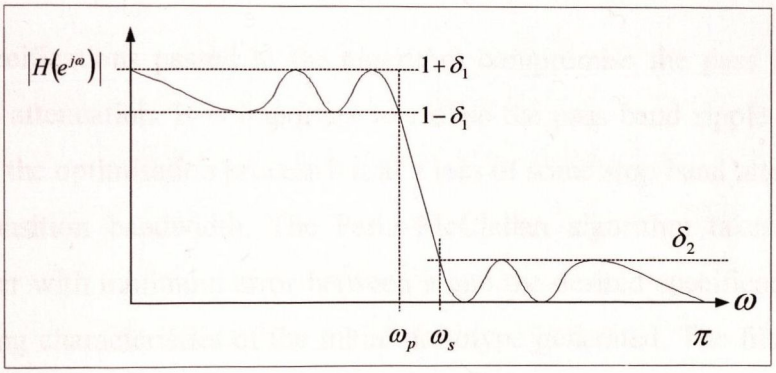


Figure 7.4 General low pass filter characteristic.

δ_1	is the peak pass band ripple
δ_2	is the peak stop band ripple
ω_p	is the pass band edge
ω_s	is the stop band edge
$\omega_s - \omega_p$	is the transition bandwidth
$-20\log_{10}(1-\delta_1)$	is the peak pass band ripple in dB
$-20\log_{10}(\delta_2)$	is the minimum stop band attenuation

If the sub-bands are to have bandwidth 1kHz then the prototype must have a bandwidth of 500Hz because the resulting analysis/synthesis filters have twice the prototype bandwidth. If the prototype response was an ideal one then the pass/stop band

edge would be 500Hz but since the filter being generated is a real filter, a trade off will be made between transition bandwidth, pass band ripple and stop band attenuation. Considering that the optimisation routine will flatten the pass band by trading off stop band attenuation, the initial prototype must have greater stop band attenuation than that required by the final prototype. After experimentation with the various possibilities the following table specifies the values used in the Parks-McClellan algorithm

Parameter	Value
δ_1	1dB max.
ω_p	350Hz
ω_s	700Hz
Stop band Attenuation	112dB min.

Table 7.1 Desired initial prototype specifications.

The specifications passed to the algorithm compromise the pass band ripple for high stop band attenuation. It is important to realise the pass band ripple is significantly reduced during the optimisation process but at a loss of some stop band attenuation and an increase in transition bandwidth. The Parks-McClellan algorithm takes this data and generates a filter with minimum error between it and the desired specifications. Table 7.2 lists the resulting characteristics of the initial prototype generated. The filter order is also constrained to be integer multiples of $2M$.

Parameter	Value
Filter Length, N	480
δ_1	0.4848dB
ω_p	350Hz
ω_s	700Hz
Stop band Attenuation	115dB

Table 7.2 Initial prototype characteristics.

The filter characterised by the values in Table 7.2 is then used as the starting point for the optimisation process. Having run the complete optimisation process on the initial prototype the resulting final prototype has specifications shown in Table 7.3.

Parameter	Value
Filter Length, N	480
δ_1	0.01589dB
ω_p	375Hz
ω_s	775Hz
Stop band Attenuation	103.5dB

Table 7.3 Initial prototype characteristics.

Examination of the prototype reveals an extremely low pass band ripple, high stop band attenuation and a transition bandwidth that is small enough to avoid overlap of non-adjacent pass bands. The net result of these processes is a prototype filter that can be cosine modulated to generate analysis/synthesis filter banks having suitable stop band attenuation, minimal pass band overlap of adjacent filters and no overlap of non-adjacent filters in the pass bands. Signals presented to this system will be divided into 20 sub-bands each of bandwidth 1kHz and subsequently these sub-bands will be recombined to reconstruct the original signal, free from alias, amplitude and phase distortion. Comparison of the output of this system to the input shows that the output is just a time-delayed version of the input. Considering the sampling rate of the system is 40kHz, the analysis and synthesis filters are the same length as the prototype, $N = 2mM = 480$ and they are FIR, the net group delay of the system is 12ms. Figures 7.5 shows the frequency response of the final prototype.

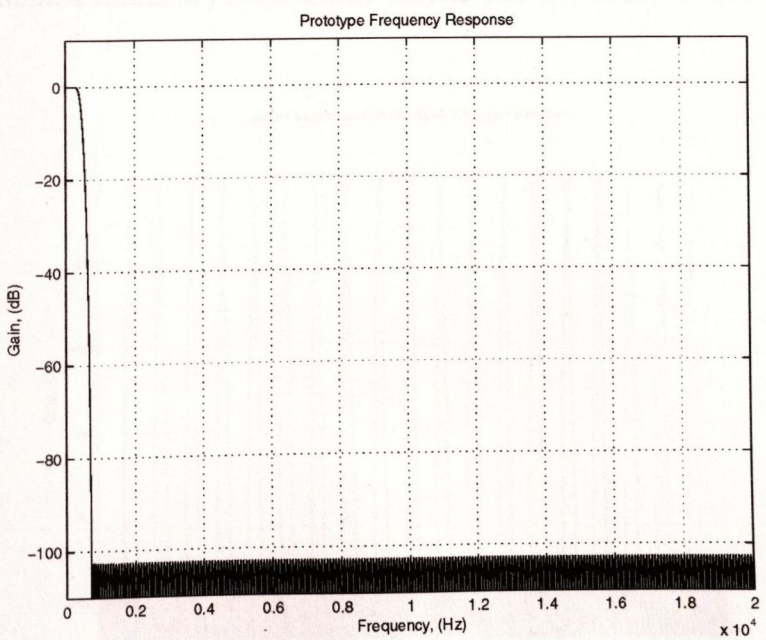


Figure 7.5 Frequency response of the prototype filter.

7.2.1.9 Analysis and synthesis filter bank generation

Having generated a suitable prototype filter all that remains is to specify a system for performing the modulation to generate the analysis and synthesis filter banks. Considering the specification of the analysis filter bank in Equation (7.22), this lends itself very efficiently to implementation of the prototype in its polyphase form. Figure 7.6 illustrates the layout for this process in both direct form and using the noble identity, which is more efficient.

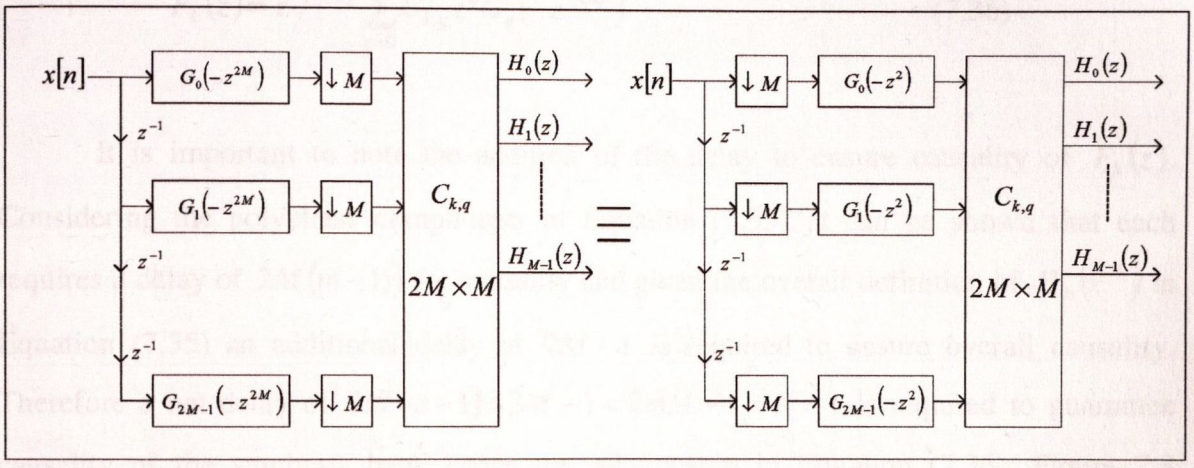


Figure 7.6 The Analysis Filter Bank, RHS employs the Noble identity.

Implementing this system generates $M = 20$ analysis filters and hence 20 sub-bands. The combined frequency response of the analysis filters can be seen in Figure 7.7.

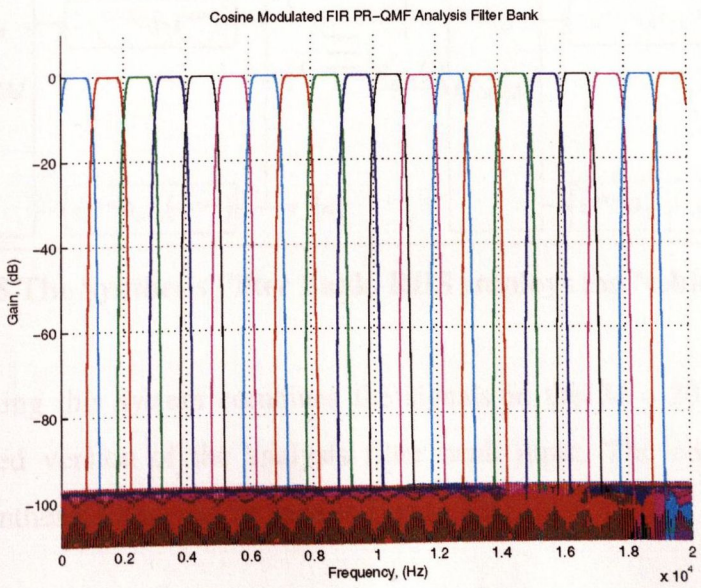


Figure 7.7 Frequency response of the analysis filter bank.

Specification of the prototype also allows similar implementation of the synthesis filter bank. Considering the synthesis filters must obey the relationship $F_k(z)=cz^{-(N-1)}H_k(z^{-1})$ an explicit z-domain closed representation of the synthesis filter can be derived with $c=1$ as

$$H_k(z^{-1})=\sum_{q=0}^{2M-1}C_{k,q}z^{q}G_q(-z^{-2M}) \tag{7.35}$$

$$F_k(z)=z^{-(N-1)}\sum_{q=0}^{2M-1}C_{k,q}z^{q}G_q(-z^{-2M}) \tag{7.36}$$

It is important to note the addition of the delay to ensure causality of $F_k(z)$. Considering the polyphase components of Equation (7.35), it can be shown that each requires a delay of $2M(m-1))$ for causality and given the overall definition of $H_k(z^{-1})$ in Equation (7.35) an additional delay of $2M-1$ is required to ensure overall causality. Therefore a net delay of $2M(m-1)+2M-1=2mM-1=N-1$ is required to guarantee causality of the synthesis bank hence the relationship in Equation (7.36). Figure 7.8 illustrates the layout for this process in both direct form and using the noble identity, which is more efficient.

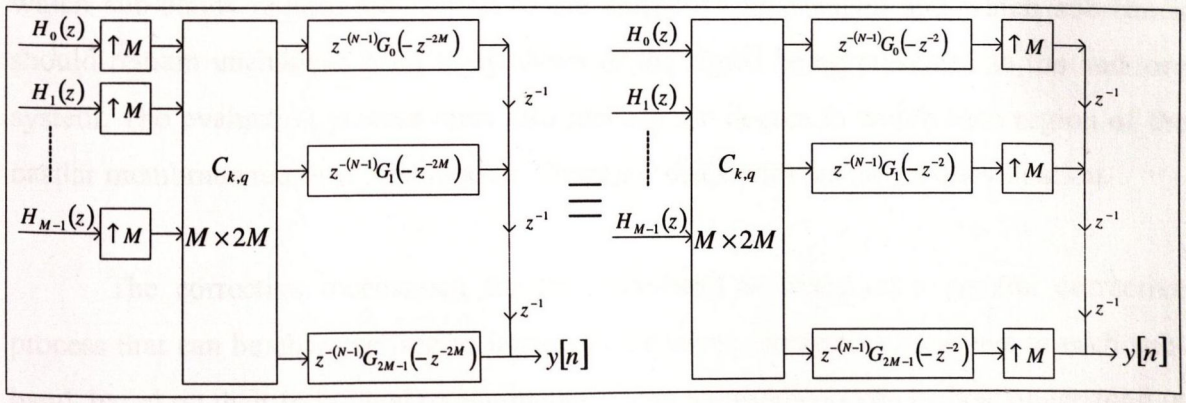


Figure 7.8 The Synthesis Filter Bank, RHS employs the Noble identity.

Implementing this system combines the signals in the $M=20$ sub-bands into a single time-delayed version of the analysis filter bank input. The combined frequency response of the synthesis filters can be seen in Figure 7.9.

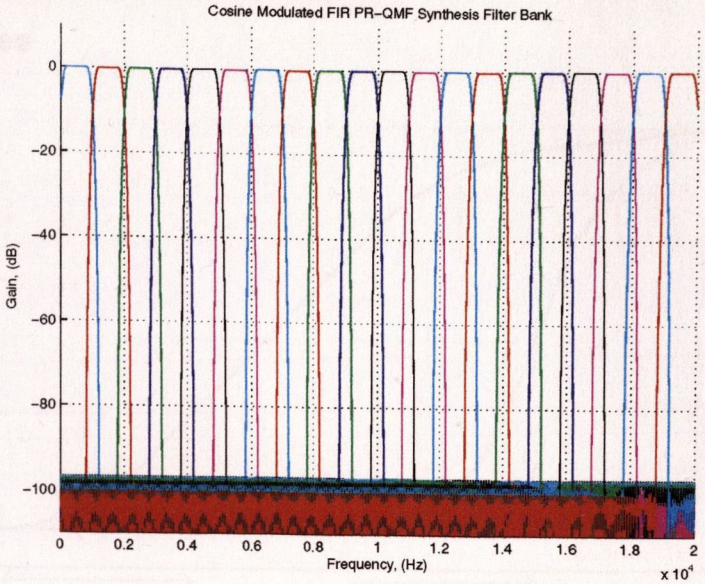


Figure 7.9 Frequency response of the synthesis filter bank.

7.2.2 Generation of the Compensation Measures for the Sub-Band Signals

The system described above can divide acoustic stimuli into M sub-band signals, some of which must be modified to pre-introduce compensation for the deficiencies of the auditory system due to outer hair cell damage at regions along the basilar membrane. Clearly the evaluation process must identify which regions of the basilar membrane are impaired and hence which frequency bands need intervention. This in turn will identify which sub-bands require application of the corrective mechanism and which sub-bands should remain unchanged prior to synthesis of the signal being presented to the auditory system. The evaluation process must also identify the degree to which each region of the basilar membrane requires intervention. Chapter 8 deals with the issues of evaluation.

The corrective mechanism for each sub-band is based on a general corrective process that can be characterised to impart the necessary corrective measures to each sub-band, based on their individual requirements. These requirements can be best understood in terms of the generalised plot in Figure 7.10 which illustrates the neural response of an impaired and unimpaired auditory system as described in Chapter 5, Section 5.3.2.

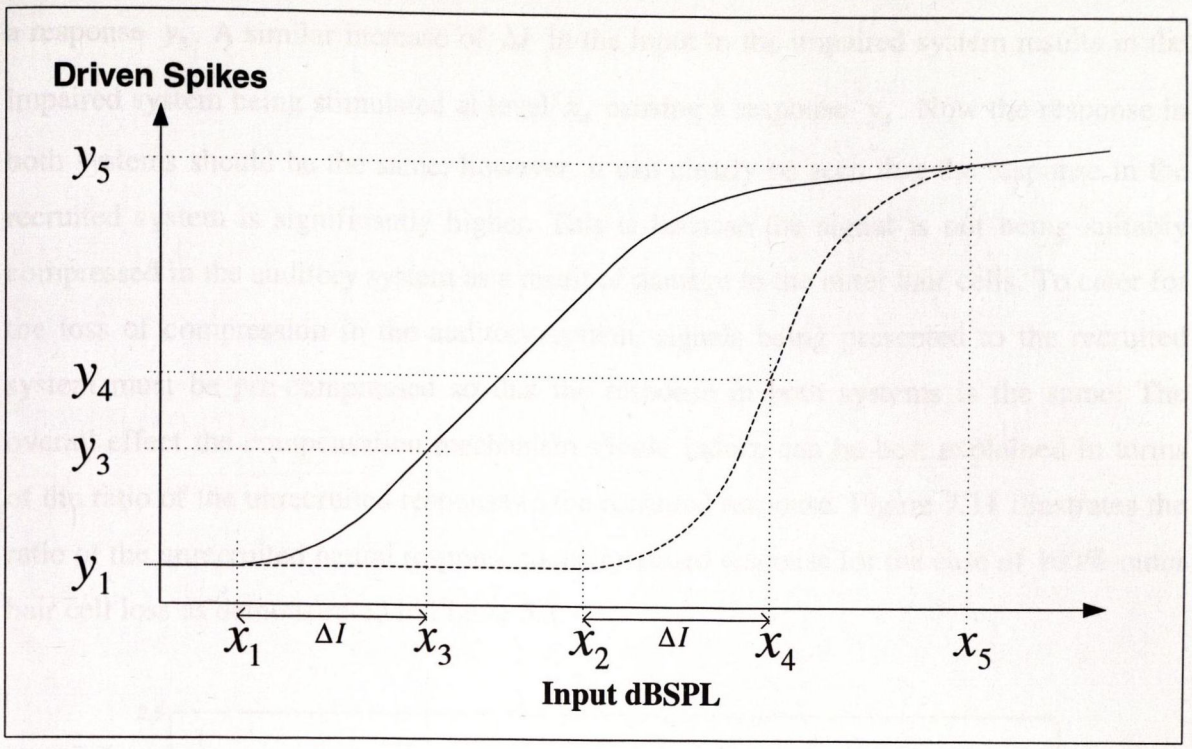


Figure 7.10 Idealised neural response. (Solid plot is unimpaired)

Examining the unrecruited response plot (solid plot) in Figure 7.10, signals with levels below x_1 elicit a driven response y_1 that is the zero driven level, as these signals are below auditory threshold. The levels x_1 to x_5 span the range of the human auditory system and signals with levels within this range elicit driven responses between y_1 and y_5. In order that the recruited system respond to signals just above the normal threshold of hearing in a similar manner to the response in the unrecruited system, suitable gain must be applied to the signal presented to the recruited system. In this case a signal (at threshold) presented to the unrecruited system at level x_1 would have to be at level x_2 to elicit the same response in the recruited system (dashed plot). This can be done by increasing the level of the signals being presented to the recruited system by $x_2 - x_1$. Importantly, such gain is only applied to signals at or above threshold, as signals below threshold should not elicit a response. A system such as this can overcome the problem of the increased auditory threshold associated with outer hair cell loss.

However, once the signal level increases above auditory threshold, the non-linear loss of dynamic range compression due to outer hair cell damage causes the loudness level to grow abnormally rapidly. Considering Figure 7.10, if the signal level is increased by an amount ΔI above threshold, the input to the unimpaired system will be at level x_3, causing

a response y_3. A similar increase of ΔI in the input to the impaired system results in the impaired system being stimulated at level x_4 causing a response y_4. Now the response in both systems should be the same; however, it can clearly be seen that the response in the recruited system is significantly higher. This is because the signal is not being suitably compressed in the auditory system as a result of damage to the outer hair cells. To cater for the loss of compression in the auditory system, signals being presented to the recruited system must be pre-compressed so that the response in both systems is the same. The overall effect the compensation mechanism should induce can be best explained in terms of the ratio of the unrecruited response to the recruited response. Figure 7.11 illustrates the ratio of the unrecruited neural response to the recruited response for the case of 100% outer hair cell loss as demonstrated in Figure 5.3.

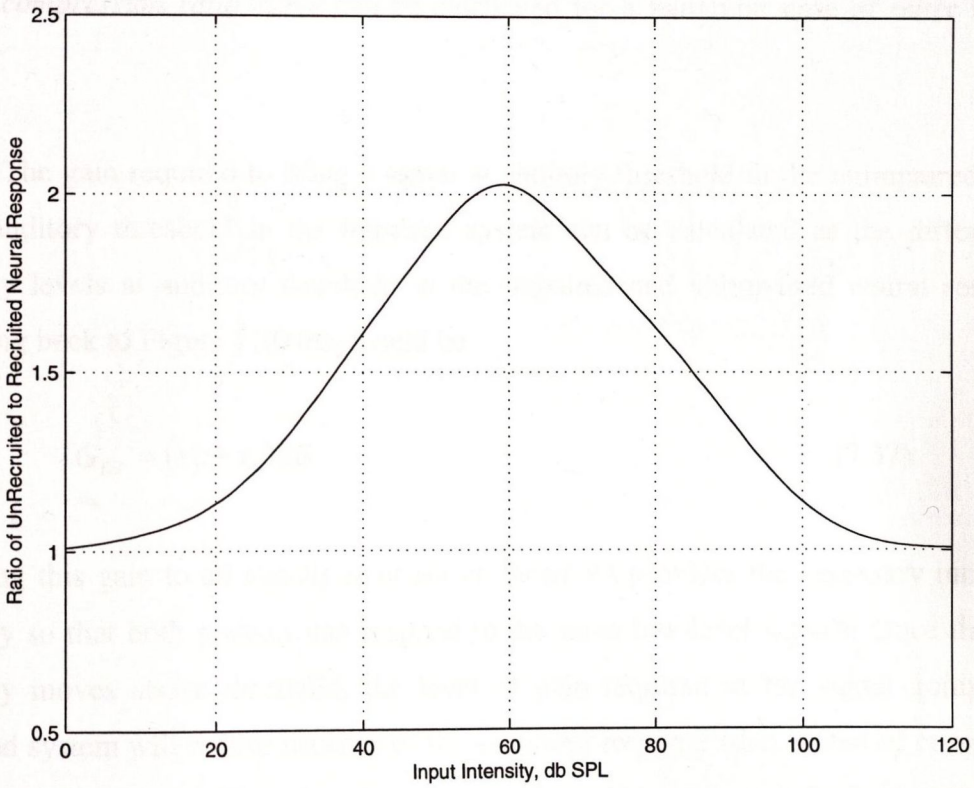


Figure 7.11 Ratio of the unrecruited neural response to the recruited neural response with 100% outer hair cell loss. (Neural Response Ratio Curve)

It can be seen from Figure 7.11 that the ratio of the neural responses of the unimpaired and impaired systems are almost unity for very low signal levels and very high signal levels and therefore both systems are approximately equally capable of responding at such levels. However, the significant increase in the ratio over the middle range of levels is a measure of the inability of the recruited system to accurately respond to these middle

levels. It is important to note that most human communication falls within this middle band and hence the impairment will have significant detrimental effects on communication. The effect of the compensation mechanism should be to reduce this ratio to unity across the whole auditory range thus making the neural responses of both systems equal. It can clearly be seen from Figure 7.11 that varying levels of compression are required across the range of input intensities.

7.2.2.1 Compression requirements

Considering the compression requirements of the compensation mechanism more closely, it can be seen that the system must identify the intensity and decide whether or not compression is required and if so how much to apply. Given the data presented in Figure 7.10 a *compression ratio curve* can be calculated for a particular case of outer hair cell loss.

The gain required to bring a signal at auditory threshold in the unimpaired system up to auditory threshold in the impaired system can be calculated as the difference in intensity levels at auditory threshold in the impaired and unimpaired neural responses. Referring back to Figure 7.10 this would be

$$G_{TH} = (x_2 - x_1)dB \tag{7.37}$$

Applying this gain to all signals at or above threshold provides the necessary increase in intensity so that both systems can respond to the same low-level signals. Once the signal intensity moves above threshold, the level of gain required in the signal going to the recruited system will reduce because of the enhanced response (due to loss of compression in the auditory system). Therefore the signal level must be compressed prior to entering the recruited system such that the level is high enough to elicit a response yet low enough so that the response is at the same level as in the unrecruited system. The compression ratio required for any input intensity can calculated from

$$C_{RATIO} = \frac{I_R(R_{NC})}{I_R(R_C)} = \frac{I_S + G_{TH}}{I_R(R_C)} \tag{7.38}$$

where C_{RATIO} is the compression ratio, $I_R(R_{NC})$ is the intensity level (including threshold compensation gain G_{TH}) required to cause a response R_{NC} in the recruited system prior to compression, $I_R(R_C)$ is the intensity level (including threshold compensation gain G_{TH}) required to cause a response R_C in the recruited system after compression and I_S is the original signal level. The compression ratio can be derived in conjunction with Figure 7.12.

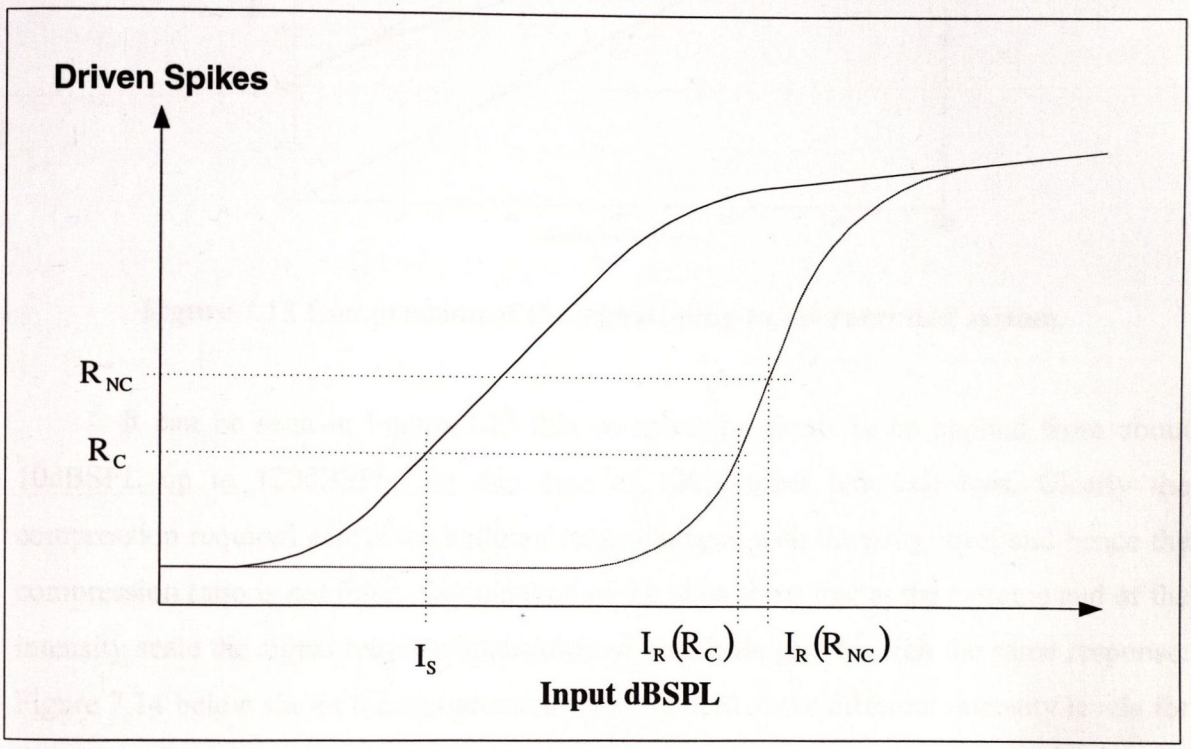


Figure 7.12 Compression ratio calculation evaluation.

Considering the intensity level $I_R(R_{NC})$ can be calculated as I_S is known and G_{TH} can be calculated from the recruited and unrecruited neural response plots. The required unrecruited response is the response a signal at level I_S will cause in the unrecruited system and therefore is the required level in the recruited system, R_C. Using the recruited neural response plot and R_C, the compressed intensity level $I_R(R_C)$ can be measured and hence the compression ratio can be calculated for each value of I_S. It is this compression the compensation mechanism must apply when the signal intensity level is at level I_S, or more precisely when the signal level going to the recruited system is at level $I_R(R_{NC})$, after the compensation mechanism caters for the increased threshold. Figure 7.13 shows how the input intensity level must be compressed to elicit the same response in the recruited system as in the unrecruited system.

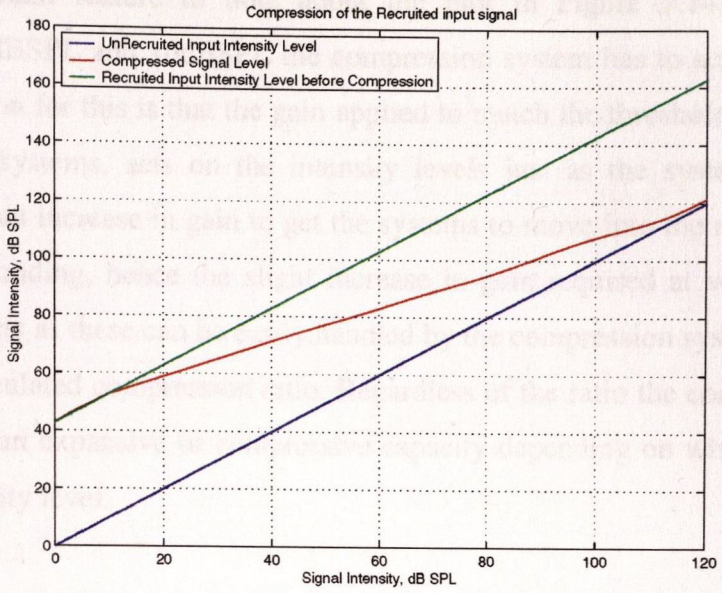


Figure 7.13 Compression of the signal going to the recruited system.

It can be seen in Figure 7.13 that compression needs to be applied from about 10dBSPL up to 120dBSPL, for this case of 100% outer hair cell loss. Clearly the compression required across the auditory range changes with intensity level and hence the compression ratio is *not* fixed. Examination of the plots show that at the extreme end of the intensity scale the signal requires application of very little gain to elicit the same response. Figure 7.14 below shows the compression ratio required at the different intensity levels for the case of 100% outer hair cell loss at 1kHz and shows the variable nature of the compression required.

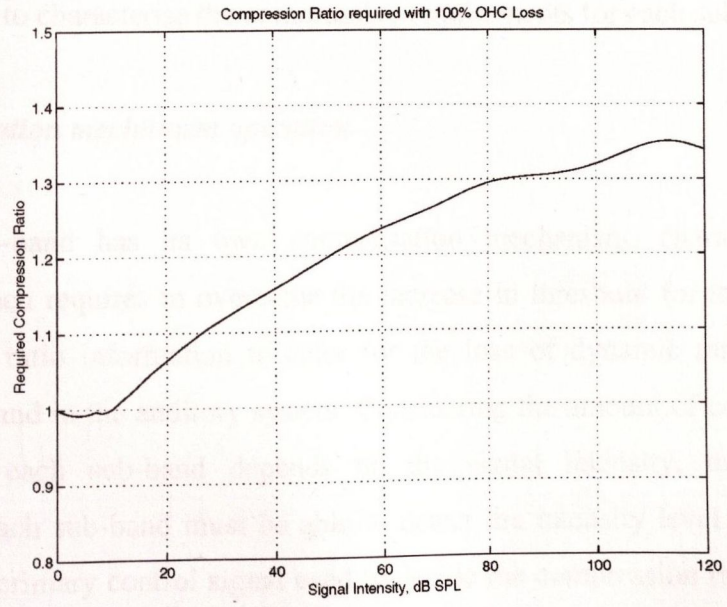


Figure 7.14 Compression ratio required for 100% outer hair cell loss at 1kHz.

An important feature to note about the plot in Figure 7.14 is that between approximately 0dBSPL and 10dBSPL the compression system has to act as an expansion system. The reason for this is that the gain applied to match the thresholds of the recruited and unrecruited systems, acts on the intensity levels just as the systems are about to respond, it takes an increase in gain to get the systems to move into the region where they are actually responding, hence the slight increase in gain required at very low intensity levels. Such effects as these can be easily handled by the compression system as it depends solely on the calculated compression ratio. Regardless of the ratio the compressive system can act in either an expansive or compressive capacity depending on what is required for the current intensity level.

It can be seen therefore that each sub-band requires a characteristic compression ratio curve, which specifies how much compression is required at particular intensity levels. To generate this compression information for each sub-band, the recruited and unrecruited neural responses for each sub-band are required. Clearly each sub-band does not have an infinitesimally small bandwidth, therefore the neural responses at a single frequency will not provide the best characterisation of the sub-band. A better approximation can be achieved by taking an average of the neural responses at both of the sub-band edges and at the centre of the sub-band for each of the M sub-bands. Given that the site of outer hair cell damage covers a region rather than a single location the use of an averaged characteristic neural response for each sub-band is an appropriate compromise. Having measured the average neural response for each sub-band, compression ratio curves can be calculated to characterise the compressive requirements for each sub-band.

7.2.2.2 Compensation mechanism operation

Each sub-band has its own compensation mechanism, characterised by the particular gain each requires to overcome the increase in threshold for that sub-band and the compression ratio information to cater for the loss of dynamic range compression within that sub-band in the auditory system. Considering the amount of compression to be introduced into each sub-band depends on the signal intensity, the compensation mechanism for each sub-band must be able to detect the intensity level in the sub-band. This level is the primary control signal used to decide the compression ratio to be applied based on the characteristic compression ratio data for each sub-band. It is important that the signal level used to set the amount of compression provided in each sub-band is

primarily based on the spectral content and signal level dynamics of the sub-band signal rather than the complete input signal. The reason for this is that the long term spectral shape of the complete signal is dominated by energy associated with low frequency components and as a result the energy associated with high-frequency components would be uncorrelated with the control signal used to determine the compression level, resulting in incorrect levels of compression being introduced. Ideally the system should be able to measure the instantaneous level of the signal in the sub-band and use this to manipulate the gain, however this poses some problems. If the gain is controlled with a rapidly changing signal level, modulation distortion can occur along with pumping and breathing (noticeable changes in background noise). On the other hand if the gain is controlled by long-term level measurements, then intense isolated transients may cause dropouts in the processed signal, as short intense signal components will have too much influence on the gain. A compromise must therefore be made that involves a combination of the short term and long-term signal levels of the complete signal and the signal levels within each sub-band. In this way the compression for each sub-band can be determined using a weighted combination of a dual-time constant complete signal level and a short-term sub-band level.

Such an approach to level detection has been used in a high quality audio application [110] and in high quality compression strategies [5], [174]. In these systems, the signal level is determined using fuzzy control rules based on trapezoidal membership functions to calculate the combinations of the short and long-term signal levels. A weighted combination of this overall level is then combined with the short-term root-mean-square (RMS) level within each sub-band to determine the signal level used in the calculation of the compression to be introduced into each sub-band signal. The detection of the short and long term signal levels are computed using the data in Table 7.4 and the RMS levels in the sub-bands are calculated using a first order recursive average over a period of 50ms within each sub-band.

	Rise Time	**Fall Time**
Short Term	4ms	150ms
Long Term	325ms	1000ms

Table 7.4 Signal level detection parameters.

The trapezoidal membership function used to determine the ratio of the short and long-term signal levels can be simply represented using the plot in Figure 7.15. The red

plot represents the weighting on the short-term level and the blue plot the weighting on the long-term level.

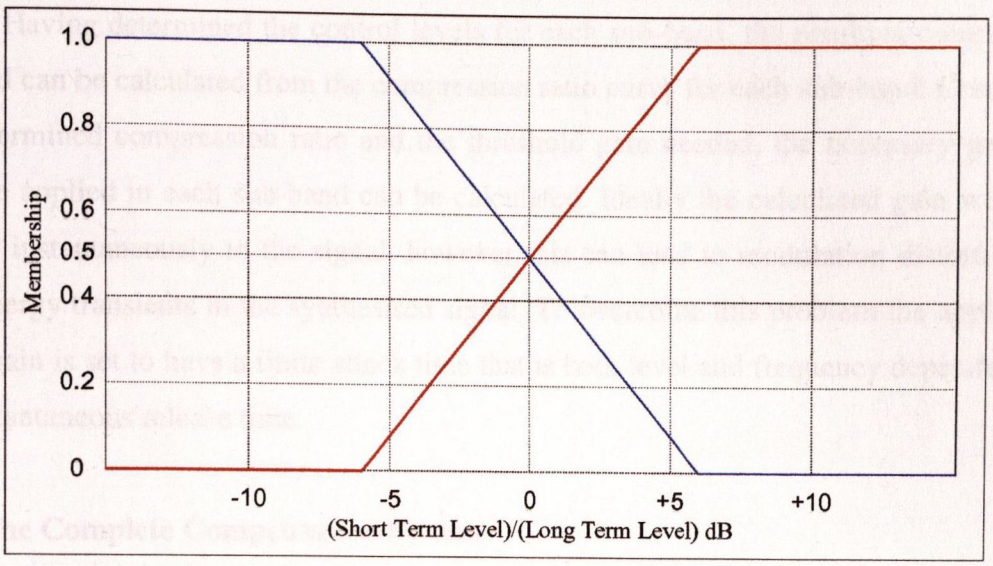


Figure 7.15 Trapezoidal membership function.

Once the level of the complete signal and the RMS level in the sub-bands have been calculated, a weighted combination of the two can be used to either emphasise control based on the complete signal or on the sub-band signal or on a combination of the two within each sub-band. The equations below govern the level detection system.

$$MainLevel = FuzzyParam * ShortTermLevel + (1 - FuzzyParam) * LongTermLevel \quad (7.39)$$

$$SBCL_n = MainLevel - C * (MainLevel - SubBandrmsLevel_n) \quad (7.40)$$

Equation (7.39) is used to calculate the main signal level as a weighted combination of the short-term level and the long-term level. The weighting factor, *FuzzyParam* is the fuzzy control value determined from the trapezoidal membership function shown in Figure 7.15. Given that the short-term and long-term levels are measured, this parameter can be easily evaluated and used in the determination of the *MainLevel*; this level represents the level of the original signal. Equation (7.40) is used to calculate the sub-band control level used in the determination of the compression required. Since each sub-band is independent, each has its own control level based on the main signal level and the current RMS level within the band. The parameter C is used to determine whether the main level or the sub-band level has a greater bias on the overall control level. Considering the requirements of the current system the weighting leans more heavily on the side of the sub-band signal level.

However, the influence of the main signal level assists in dampening any rapid fluctuations in signal level within each sub-band.

Having determined the control levels for each sub-band, the resulting compression required can be calculated from the compression ratio curve for each sub-band. Combining the determined compression ratio and the threshold gain needed, the necessary gain that must be applied in each sub-band can be calculated. Ideally the calculated gain would be applied instantaneously to the signal, however this can lead to modulation distortion and high-energy transients in the synthesised signal. To overcome this problem the application of the gain is set to have a finite attack time that is both level and frequency dependent and has instantaneous release time.

7.2.3 The Complete Compensation System

The cosine modulated PR-QMF system provides separation of the complete signal into sub-band signals, which it subsequently re-combines to synthesise a single signal to present to the impaired auditory system. Compensation mechanisms are provided for each sub-band to alter the signals within the sub-bands so the response of the impaired system to the complete compensated signal is as close as possible to the response of the unimpaired system to the signal prior to compensation. The complete layout of the compensation system can be seen in Figure 7.16.

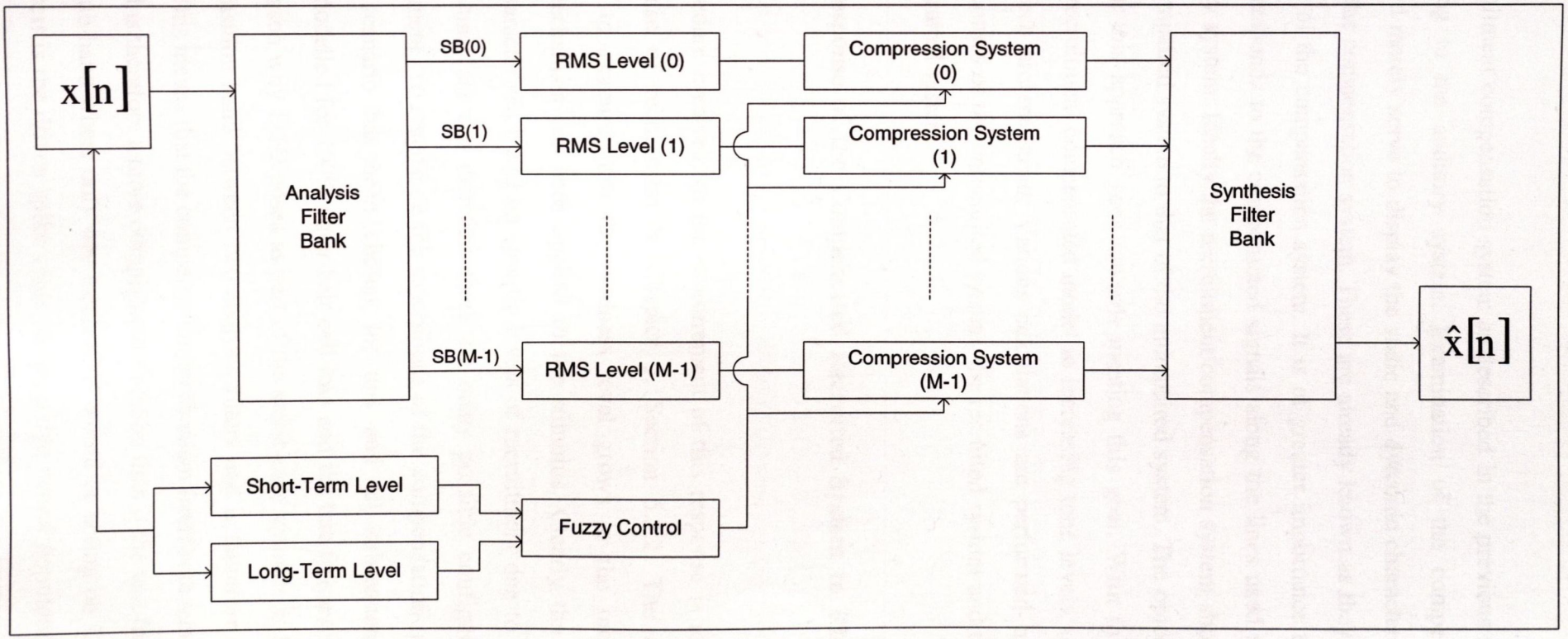


Figure 7.16 Layout of the Analysis/Synthesis Compensation System.

7.3 Simulation of the Recruitment Compensated Auditory Model

The recruitment compensation system as described in the previous sections act on the stimuli going to the auditory system. Examination of the compensated signals themselves would merely serve to display the static and dynamic characteristics imparted to the signal by the compensation system. These are already known as they are defined in the specification of the compensation system. It is of greater importance to see how the auditory model responds to the compensated signals along the lines used to examine the recruited auditory system. Ideally the recruitment compensation system should restore the response of the impaired system to that of the unimpaired system. The objective here is to examine how far this approach goes towards meeting this goal. With this is mind the response of the recruitment compensated model to increasing tone levels, tone bursts and to synthetic speech are measured. Various calculations are performed on this data to compare the responses of the unrecruited system, the recruited system and the recruitment compensated recruited system.

7.3.1 Neural Response of the Compensated Recruited System to Rising Stimulus Level

The procedure involved for the measurement of this response is identical to that used for the case of recruitment in Chapter 5, Section 5.3.2. The object of this measurement is to examine how the loudness level grows in the model when the compensation mechanism has been applied to the stimulus. Clearly the compensation mechanism is characterised based on specific levels of recruitment due to outer hair cell loss within the inner ear and therefore there are many possible configurations for the modelled recruitment. To emphasise the capabilities of the compensation mechanism, the worst possible scenario has been chosen for this and all subsequent simulations. Recruitment is modelled for 100% outer hair cell loss and the test frequency is chosen to be 1kHz. The reason why 1kHz poses as part of the worst-case scenario is because it is at centre of the transition band between two analysis filters and is therefore present within two sub-bands. This means that the compensation mechanism from both sub-bands will act on the signal and is therefore a more complicated situation than if the test frequency was at the centre of a sub-band where only one mechanism would be acting on it. As before, of particularly interest is the driven spike count in the active neural population. Figure 7.17 shows the driven neural responses for the unimpaired system, the system suffering from

100% outer hair cell loss and the system suffering from 100% outer hair cell loss with the compensation mechanism in place.

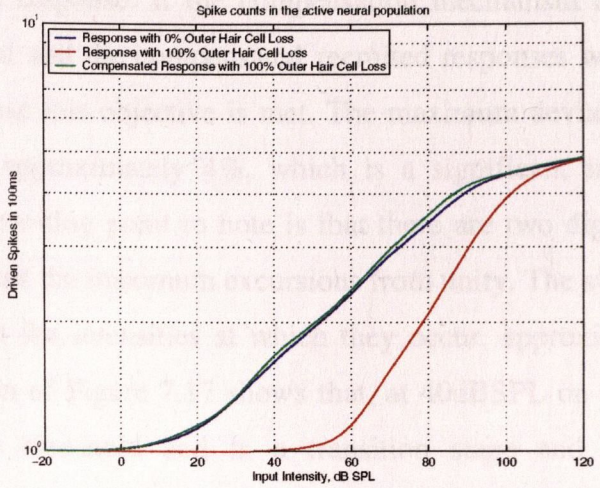


Figure 7.17 Driven spike curves before and after compensation

In Figure 7.17 the blue curve is the response of the unimpaired system, the red curve is the response of the system suffering from 100% outer hair cell loss (i.e. recruited) and the green curve is the response of the system suffering from 100% outer hair cell loss with the compensation mechanism having been applied. Examination of the responses reveals how satisfactorily the compensation mechanism works even in the worst possible situations. Closer examination of the responses shows that the compensated response does not exactly restore the impaired system to its unimpaired state and has a tendency to overestimate the gain required. However, this over-estimation can be minimised through optimisation of the compensation mechanism. Figure 7.18 gives a more precise indication of how close the compensation response approaches the unimpaired response.

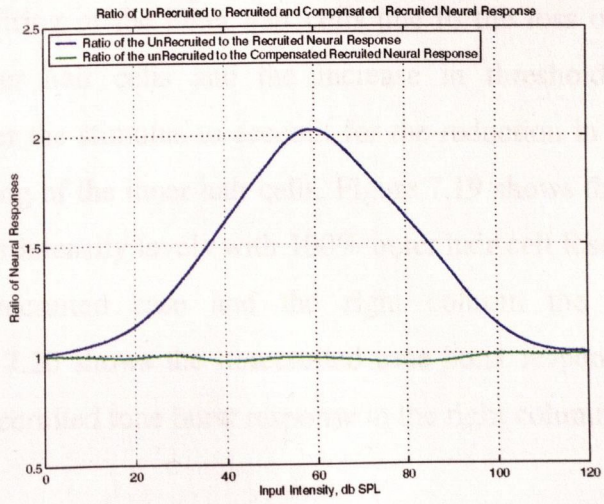


Figure 7.18 Ratio of the Unrecruited Neural Response to the Recruited and Compensated Recruited Neural Responses.

The blue curve represents the ratio of the unrecruited response to the recruited response and the green curve represents the ratio of the unrecruited response to the compensated recruited response. If the compensation mechanism operated perfectly the ratio of the unrecruited and the compensated recruited responses would be unity. Figure 7.18 indicates how close this objective is met. The maximum deviation from unity in the compensated case is approximately 4%, which is a significant improvement over the recruited case. An interesting point to note is that there are two dips in the compensated ratio curve that represent the maximum excursions from unity. The size of these dips is not what is interesting but the intensities at which they occur, approximately 40dBSPL and 85dBSPL. Examination of Figure 7.17 shows that, at 40dBSPL on the recruited plot, the system is just above threshold and in a transition stage and at 85dBSPL on the compensated plot the system is progressing towards its saturated state and in a transition state. At these points the rate of change of the compression is higher than at any other intensity and shows that the compensation mechanism has the greatest difficulty at these points. Clearly it performs an admirable job even under these conditions, given that the maximum deviation from an ideal response is only 4%.

7.3.2 Tone Burst Response of the Compensated Recruited System

The tone burst response is measured as before for intensity levels progressing from 30dBSPL up to 120dBSPL, at a frequency of 1kHz. This is performed once again under the worst-case conditions of 100% outer hair cell loss and at a sub-band transition frequency. The object of this test is to see how far the compensation system goes towards correcting the firing of the inner hair cells. The reason for this is that the outer hair cell loss adversely affects the firing of the inner hair cells due to the loss of compression of the excitation of the inner hair cells and the increase in threshold. The compensation mechanism should alter the stimulus to account for the reduction in outer hair cell action and restore normal firing of the inner hair cells. Figure 7.19 shows the tone burst response at 1kHz for the various intensity levels with 100% outer hair cell loss in all cases. The left column shows the recruited case and the right column the recruited case with compensation. Figure 7.20 shows the unrecruited tone burst response in the left column and the compensated recruited tone burst response in the right column.

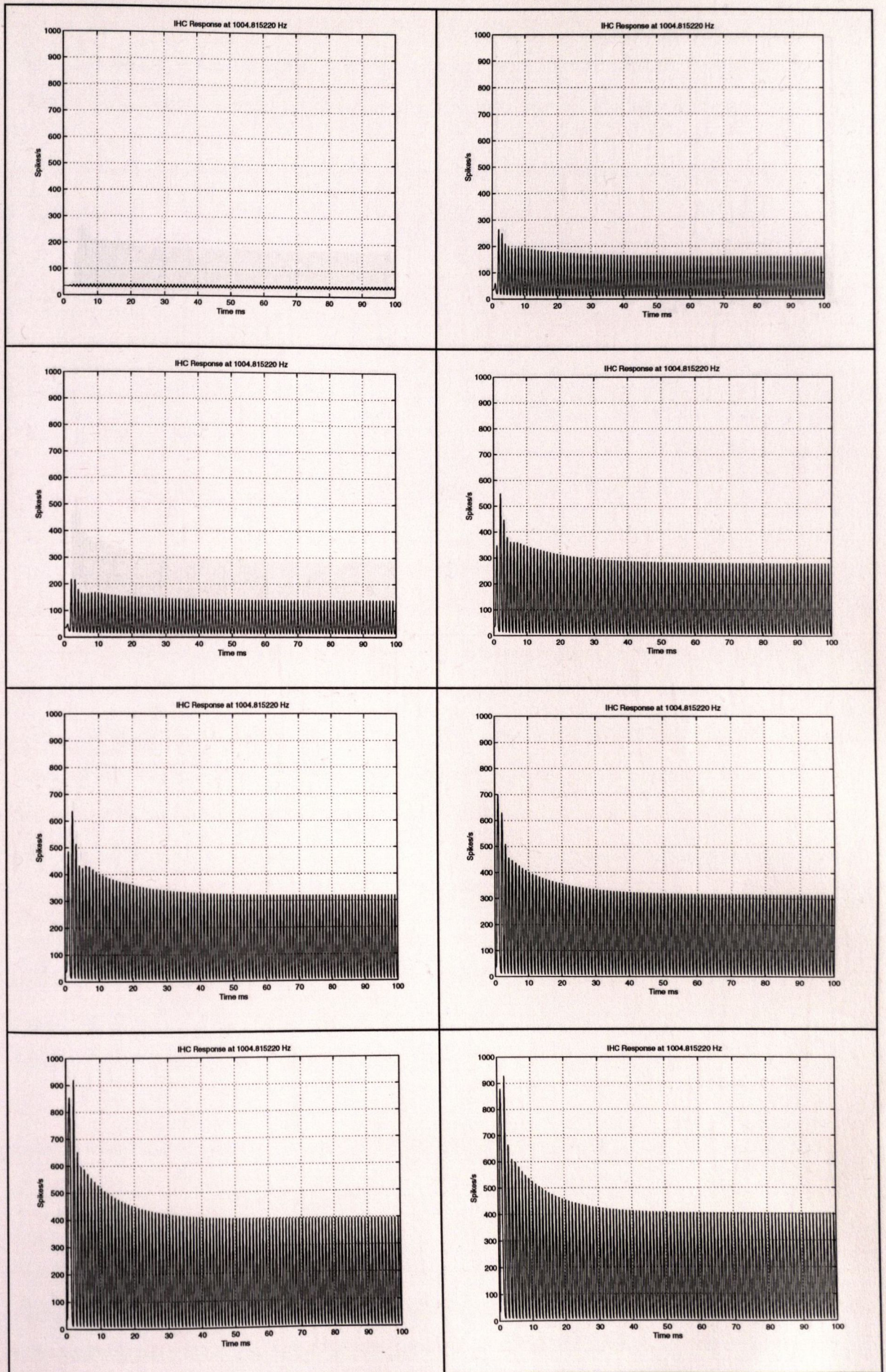


Figure 7.19 1kHz tone burst response at 30, 60, 90 & 120dBSPL top to bottom. Left is without compensation, right is with compensation.

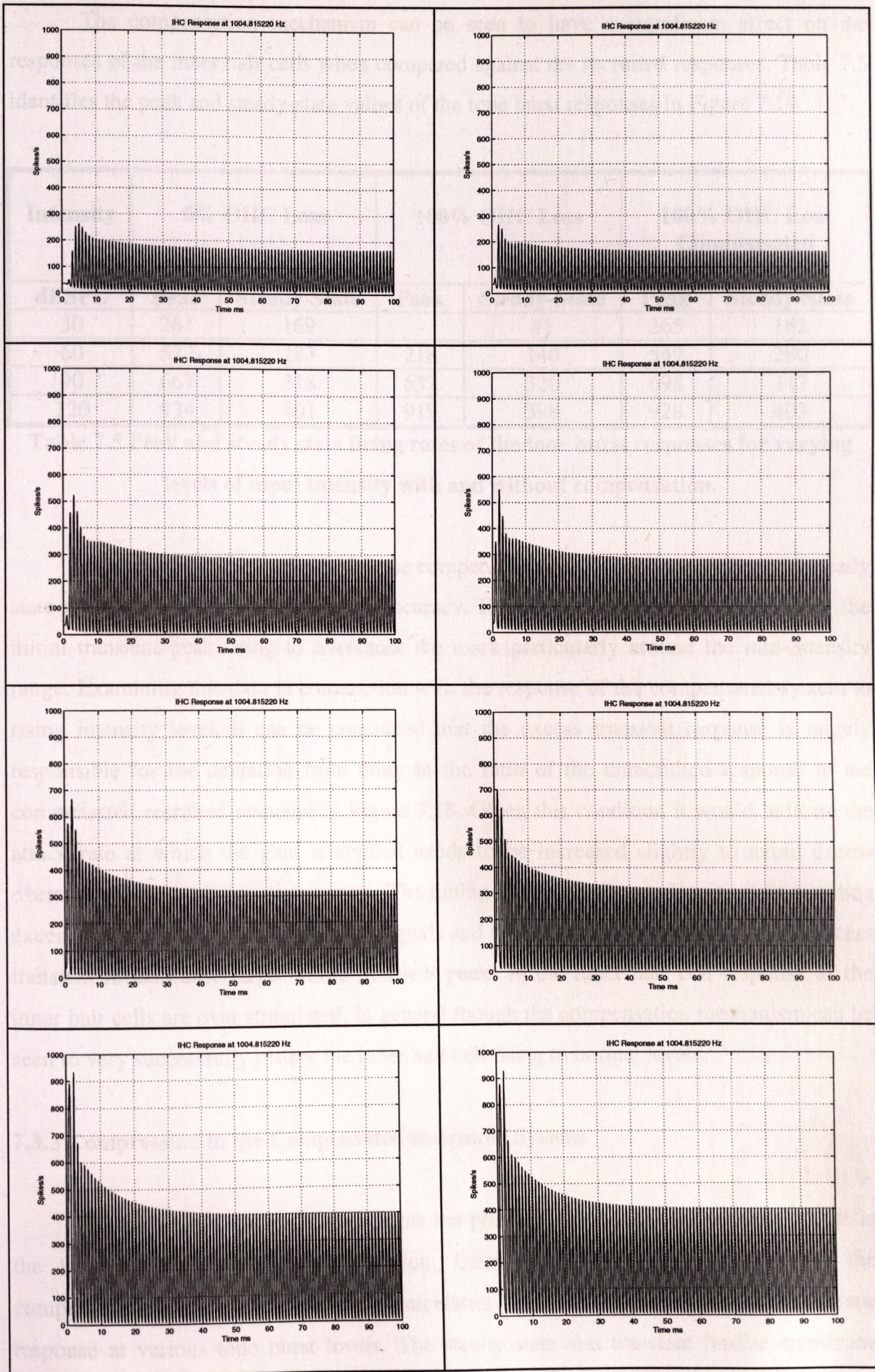


Figure 7. 20 1kHz tone burst response at 30, 60, 90 & 120dBSPL top to bottom. Left is unrecruited, right is recruited with compensation.

The compensation mechanism can be seen to have a significant effect on the responses of the inner hair cells when compared against the recruited responses. Table 7.5 identifies the peak and steady-state values of the tone burst responses in Figure 7.19.

Intensity	0% OHC Loss		100% OHC Loss		100% OHC Loss Compensated	
dBSPL	Peak	Steady-State	Peak	Steady-State	Peak	Steady-State
30	261	169	-	41	265	162
60	522	283	218	140	549	280
90	667	318	637	320	698	317
120	934	401	919	398	928	403

Table 7.5 Peak and steady state firing rates of the tone burst responses for varying levels of input intensity with and without compensation.

Examining the data reveals that the compensation mechanism can restore the steady state responses with a high degree of accuracy. However, it has a tendency to allow the initial transient peak firing to overshoot the mark particularly around the mid-intensity range. Examining this data in conjunction with the response of the compensated system to rising intensity level, it can be concluded that the excess transient response is largely responsible for the deviation from unity in the ratio of the unrecruited response to the compensated recruited response in Figure 7.18. Given this condition it would indicate the attack rate at which the gain is applied needs to be increased slightly to avoid excess overshoot in the compensation system. The tuning of this attack rate is critical in avoiding excess transients both in the sub-band signals and the synthesised stimulus. Clearly excess transient levels cause larger initial transient peaks in the inner hair cell response as the inner hair cells are over stimulated. In general though the compensation mechanism can be seen to very successfully restore the inner hair cell firing to normal levels.

7.3.3 Compression in the Compensated Recruited System

It has been shown in Chapter 5 that the primary effect of outer hair cell damage is the loss of dynamic range compression. Using the same approach as before the compression in the system has been calculated from the measured basilar membrane response at various tone burst levels. The steady state and transient basilar membrane responses are taken with 100% outer hair cell loss and at a frequency of 1kHz, thus testing the system under worst-case conditions. Table 7.6 shows the steady state and transient

compression ratios for the unrecruited system, the recruited system and the compensated recruited system.

Tone Burst Level, dBSPL	0% OHC Loss		100% OHC Loss		100% OHC Loss Compensated	
	Transient	Steady State	Transient	Steady State	Transient	Steady State
30=>60	2.35:1	2.43:1	1.43:1	1.39:1	2.23:1	2.30:1
60=>90	2.31:1	2.38:1	1.43:1	1.45:1	2.13:1	2.27:1
90=>120	1.93:1	2.17:1	1.31:1	1.42:1	2.01:1	2.19:1

Table 7.6 Compression Ratios for the Unrecruited, Recruited and Recruited Compensated System.

Examination of the results in Table 7.6 clearly shows that the compensation system has a significant beneficial effect on the compression in the recruited system both in the transient and steady state mode of operation. The compensation mechanism can be seen to provide a very close restoration of the compression ratio below 90dBSPL but tends to over compress the signals at higher levels. The reason for the seemingly excessive compression going from 90dBSPL to 120dBSPL is by design because at no point should the human auditory system be exposed to such extreme levels. All hearing aids should ensure that the uncomfortable threshold level of the user should never be exceeded and forcing the compression to be become excessively high as the intensity levels approach the uncomfortable threshold level ensures this will be avoided.

7.3.4 Neural Response of the Compensated Recruited System to Synthetic Speech

The above tests indicate that the compensation mechanism does indeed successfully introduce a mechanism for the compensation of recruitment. The final test of the compensated system involves the stimulation of the system with a more complicated stimulus, synthetic speech. The system is stimulated with the same synthetic syllable /da/ as before to facilitate comparison. The conditions are set so that 100% outer hair cell loss is specified at 1kHz and the response is generated with and without the compensation mechanism in place. Figure 7.21 shows the response of the recruited system and Figure 7.22 shows the response of the compensated recruited system.

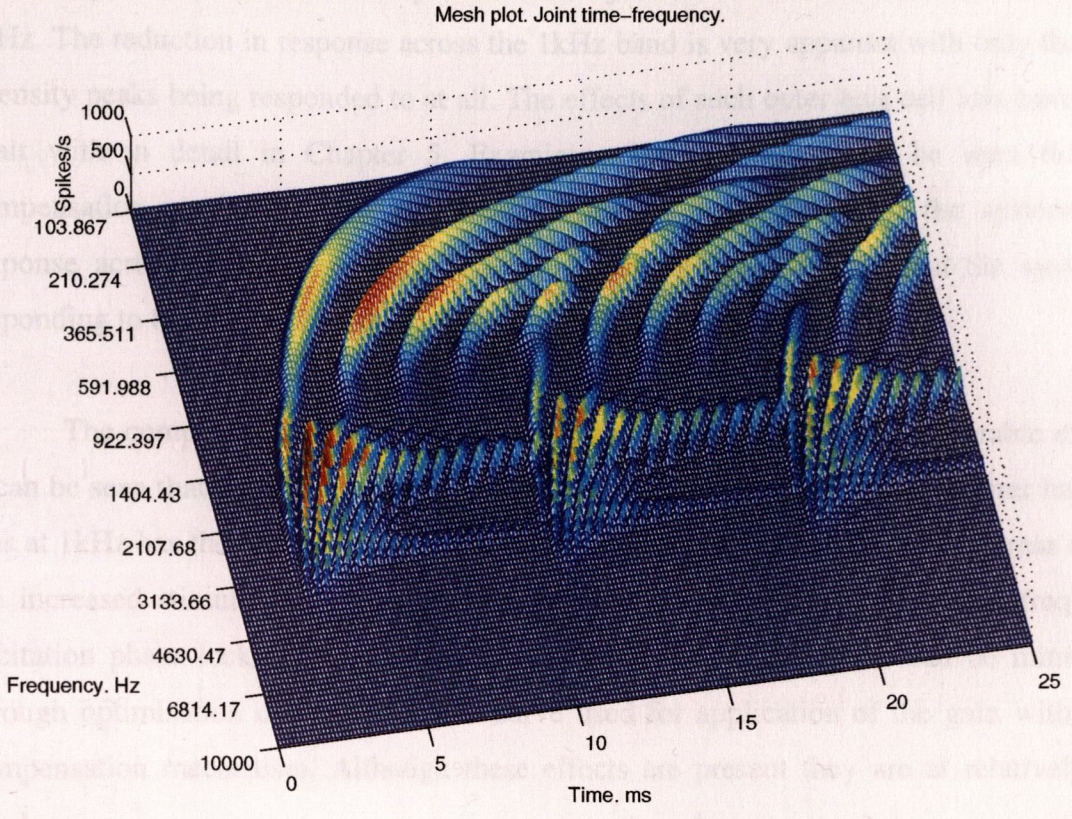


Figure 7.21 Neural Response with 100% OHC loss at 1kHz, without Compensation

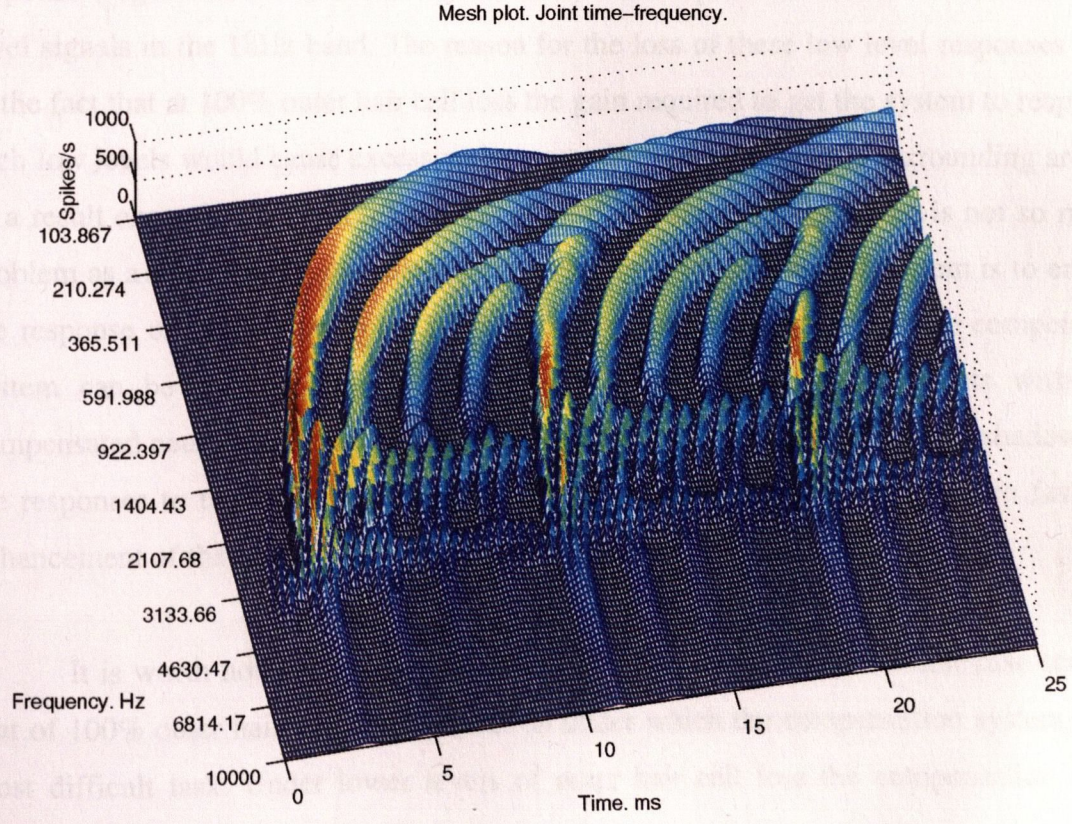


Figure 7.22 Neural Response with 100% OHC loss at 1kHz, with Compensation

Figure 7.21 shows the response of the system with 100% outer hair cell loss at 1kHz. The reduction in response across the 1kHz band is very apparent with only the high intensity peaks being responded to at all. The effects of such outer hair cell loss have been dealt with in detail in Chapter 5. Examining Figure 7.22 it can be seen that the compensation mechanism has a dramatic effect on the response of the system. The response across the impaired band has been enhanced significantly and the system is responding to both the high and low intensity components of the signal.

The compensation mechanism does, however, introduce some undesirable effects. It can be seen that the compensation mechanism although compensating for outer hair cell loss at 1kHz has the effect of causing increased excitation of higher frequency areas due to the increased stimulation level at 1kHz; this can be seen in the excess high frequency excitation phase locked to the lower frequency signal. These effects can be minimised through optimisation of the attack rate curve used for application of the gain within the compensation mechanism. Although these effects are present they are at relatively low levels.

Another feature that can be detected on careful comparison of the unimpaired response (Figure 4.38) and the compensated system response is the loss of extremely low level signals in the 1kHz band. The reason for the loss of these low level responses is due to the fact that at 100% outer hair cell loss the gain required to get the system to respond to such low levels would cause excessive increases in stimulation in the surrounding area and as a result excessive loudness. This loss of very low intensity responses is not so much a problem as a trade off. The primary requirement of the compensated system is to enhance the response of the system to the primary components within speech. The compensation system can be seen to achieve this given the presence of the formants within the compensated neural response. The low level components will be greatly overshadowed by the responses to the main components and hence to an extent are neglected in favour of enhancement of the main signal content.

It is worth noting that the responses shown above are for the worst case scenario, that of 100% outer hair cell loss, a situation under which the compensation system has its most difficult task. Under lower levels of outer hair cell loss the compensation system performs even more impressively.

7.4 Other Recruitment Compensation Systems

Generally speaking methods for recruitment compensation can be divided into three different categories based on the approach used.

1. Linear amplification with automatic gain control.
2. Amplitude compression with equalisation.
3. Parametric compression of speech parameters.

Each of these methods uses an inherently different technique to try and compensate for the effects of loudness recruitment. However, none are consistently good for the full range of sounds found in a language like English for all listening environments [4]. Even within each broad category there are several variations on the general themes.

7.4.1 Linear Amplification with Automatic Gain Control

The approach employing linear amplification with automatic gain control uses a base level of amplification to raise as much of the speech signal above the threshold of hearing as possible while applying automatic gain control to ensure all signals remain below the threshold of pain. This form of linear amplification is the type typically found in hearing aids [85]. Studies performed on this approach have shown it to be relatively successful under controlled laboratory conditions [175]; most notably this technique is successful in steady-state background conditions where only minor to medium uniform loss of dynamic range has occurred. In other studies examining linear amplification, a marked improvement of performance has been found under conditions of flat loss. However, it has also been shown that linear amplification is significantly less effective for conditions of sloping loss (the type often found when recruitment is present) [85]. It was concluded that loudness recruitment and threshold elevation have a more detrimental effect on speech intelligibility in the presence of fluctuating background sounds [85] and that only part of the decrease of performance produced by actual cochlear hearing loss can be compensated by conventional linear hearing aids [85], [72]. In general this approach is unsuitable for recruitment compensation.

7.4.2 Amplitude Compression with Equalisation

The second approach, employing amplitude compression and equalisation operates to compress the normal dynamic range of hearing into the reduced range of the impaired subject with the net effect being that the impaired listener experiences sounds of comparable loudness to that of the unimpaired listener. In general, the specification of the parameters for the compression in the majority of such schemes is based on normal conversation without any temporal variation [4]. This approach was originally conceived by Villchur [89] who showed that expansion followed by attenuation was a reasonable model of loudness recruitment and therefore this approach using amplitude compression and equalisation operates to invert the effects of recruitment as modelled using expansion and attenuation. However, having compression parameters that do not vary with time results in this scheme being unable to provide sufficient amplification to low intensity signals and causes severe clipping of high intensity signals. More successful advanced modifications of this approach have recently evolved incorporating both intensity and frequency dependence into the specification of the amplitude compression employed.

Several novel structures based on this idea have emerged which aim to overcome the limitations of the most basic form of amplitude compression. One such structure [6] proposes to "*place the long term average speech spectrum within the loudness contours of the impaired hearing at a position similar to its position within normal hearing.*" The approach used involves a level-dependent mixing structure that applies threshold gain at the lowest input levels and tends towards unity gain for the highest input levels. In summary, the architectural structure of the process involves passing the input through two linear filters whose transfer functions *place* the loudest and quietest expected inputs. The two outputs are then weighted as a function of input intensity to produce a final output. Most importantly, the mixing algorithm results in a smoothly varying transfer function continuum in between the lowest and highest input levels to ensure smooth transitions between output levels. The specific details can be found in [6] but it suffices to say that the amount of compression gain is both intensity and frequency dependent. In comparison to the fixed frequency shaping and amplification provided by linear hearing aids as specified by the National Acoustics Laboratory (NAL) standard [176], this structure is significantly better at addressing the dependence of loudness on input intensity and frequency. This is not surprising given the typical static nature of linear hearing aid responses. One of the primary goals of this method was simplicity of implementation, however, this does involve

certain trade offs. Firstly, the frequency dependent nature of the compression suffers because the limited length and simplicity of the filters used in placing the input provides crude coverage of the frequency range of speech resulting in an approximate placement of the input level. The net effect is that the required gain is only approximately applied to the input across the frequency range covered by the filters. This could be overcome with more complex filters of longer length but would greatly increase the complexity and latency of the scheme. Secondly, the linear mixing algorithm depends on the loudness contours of the reduced dynamic range being nearly equally spaced, something that will happen only with a flat hearing loss. If the loss is sloping, then the distribution of loudness contours in the reduced dynamic range will not be uniform and a significantly more complex mixing algorithm will be required. Unfortunately, this is typically the case. In general, this method shows a very innovative and efficient approach to addressing the problem of loudness recruitment but in its present state is of limited capacity due to its inability to cater for more realistic hearing losses.

Subsequently a very recent and more advanced compensation method has been proposed in [106], [107], [108] that uses the work of [6] as its basis. In this method the importance of the frequency dependent nature of the compressive gain is recognised and addressed using multiband processing for the input placement. This allows more precise control of the compressive gain applied at different frequencies providing much greater accuracy of input placement. At its core this method does not, however, focus on suitable location of the long-term average speech spectrum within the impaired dynamic range but more so on mapping real world signals into this reduced dynamic range. There is a clear distinction in that the long-term average speech spectrum is a good general specification for speech but a poor representation of the rapidly fluctuating acoustic stimuli associated with real world signals. In summary, this approach uses a frequency- and magnitude-dependent non-linear compensator to apply the necessary compressive gain to the input, in conjunction with an echo canceller to allow provision of large gains in the compensator. The compensator is generated by additively weighting a set of arbitrarily defined time-invariant band pass filters, where the weighting function used to perform the non-linear compression is determined from a predefined non-linear frequency-dependent compression curve. The actual weighing applied is dependent on the input level to each sub-band. In this way, the non-linear compression applied to the input signal is both frequency and magnitude dependent. Given the feedback involved in this method, when trying to compensate for slight impairment instability is not a problem; however, for profound

impairment the compensator is required to provide substantial gain and the closed-loop system used in this method can become unstable. To prevent this an echo canceller is incorporated to maintain stability of the system. The detailed specifications for this method can be found in [106], [107], [108]. Limited results are available for this compensation model (it is very recent 1998-1999) but it has addressed the key issues associated with loudness recruitment and has been shown to be capable of providing gain variations of over 60dB. This work presents a hearing aid concept involving recruitment compensation and acoustic feedback cancellation with reduced computational complexity. While still under development it does appear to provide a promising approach to using amplitude compression for recruitment compensation. A point to note is that the level dependence of the compression is dependent on the level of the entire input presented to the system. In this sense the compression is level dependent but this does not specifically address the signal level within each of the sub-bands when determining the weighting for each sub-band. As a result, the compression in each band will be dependent on the long-term spectral shape of the input, which is dominated by the low frequency content of the signal. However, the short-term spectral shape of speech is usually significantly different from the long-term spectral shape and thus the compression applied will be uncorrelated with the high frequency energy in the input [110], [5]. It would seem logical that the signal level within each sub-band should be the predominant level considered when determining the weighting for each sub-band but as has been noted this is a work in progress.

Lastly, in this section on amplitude compression, are the compensation mechanisms that deal directly with multiband processing. This approach typically deals with analysis/synthesis systems that separate a signal into sub-bands and then recombines the modified sub-band signals to generate a compensated output. The compensation mechanism is typically applied directly to the sub-band signals prior to output reconstruction. [5], [61] and [109] are recent examples of such systems. Detailed technical specifications for the processing blocks used in [61] and [109] remain proprietary as they are used in the core of a recruitment-compensating hearing aid provided by the GN ReSound Corporation. However, it is known that this system is based on two band compression design generically similar to the work in [177]. Considering the work in [5], this proposes a multiband compression strategy for a digital hearing aid based directly on the multiband analysis/synthesis approach. This system specifies static characteristics to describe the system response to steady-state signals, incorporating the threshold of hearing and uncomfortable loudness level along with ratios of compression and expansion within

each sub-band. A level detection system used by the compression mechanism determines the dynamic characteristics of the system. The level detection system determines a control signal for each sub-band from which the compression applied in each sub-band is determined. Importantly, the detection system considers both the complete input signal level and the level within each sub-band when determining the control level for each sub-band. Special allowance is made for biasing of the control signal for each sub-band in favour of the overall signal level or particular sub-band level allowing flexible characterisation of the control placed in each sub-band. Although the compressive function is specific to each sub-band, the typical compressive functions used apply fixed compression ratios i.e. a linearly varying compression level. However, the compression variation required to compensate for loudness recruitment is non-linear in nature. This system will therefore typically overestimate the compression required at low input levels and underestimate the compression required at high input levels. An improved system for determining the compression functions would greatly enhance this system. In generally though, this approach to amplitude compression is extremely flexible and addresses the frequency and level dependent nature of a compensation mechanism for loudness recruitment.

7.4.3 Parametric Compression of Speech Parameters

Technically speaking this approach could be considered as another variation of multichannel amplitude compression in that it involves a form of analysis of the input into various channels, these channels are modified to compensate for loudness recruitment and then recombined to produce an output. Where this general approach differs from the other forms of amplitude compression is that the analysis stage typically performs some form of frequency analysis the output of which is used to characterise a sinusoidal model for speech. The speech is then represented as the sum of sinusoids with various amplitudes, frequencies and phases. Each sinusoidal component corresponds to a separate frequency band and the amplitude parameters are given compression gain, based on an intercomponent masking model of both normal and impaired hearing. The masking model incorporates the listener's threshold of hearing and pain. The amount of gain applied to each frequency component is based on preserving the relative distance above masking threshold and allows the algorithm to adjust to the time-varying qualities of the input speech. The modified frequency components are then used to synthesise the output. There are two main variations on this theme depending on the underlying analysis performed.

The first employs a Fast Fourier Transform (FFT) to perform the analysis e.g. [7] and the second employs a Fast Wavelet Transform (FWT) to perform the analysis e.g. [4]. The specific details of each method can be found in the associated references but it suffices to say they both follow a similar pattern as described above. Where the methods differ is at the analysis level and in how the compression gain is applied. The type of calculation used to determine the compression gain is the same for both the FFT and FWT processes. However, for the FFT process the compression gain is dependent on and applied to the sinusoidal amplitude parameters, whereas for the FWT process the compression gain is dependent on and applied to the wavelet coefficients. In essence both methods closely resemble multiband amplitude compression and as expected both perform on par with the amplitude compression strategies laid out in Section 7.4.2 [111]. A specific advantage of these types of methods is the fact that the frequency resolution can be significantly increased without excessive increases in the numerical complexity of the systems. In the FFT case, the length of the FFT can be increased to provide increased resolution; for the case of the FWT, the system of decomposition would have to move to wavelet packets but would allow decomposition of the signal into frequency bands as fine as desired. Greater frequency resolution would allow more precise specification of the compression applied to each component used in the synthesis of the output of the compensation systems with the aim of providing more suitable loudness recruitment compensation.

7.4.4 Linear versus Non-linear Processing

Throughout the development of compensation mechanisms for loudness recruitment, using the approaches mentioned above and their variations, there have been a multitude of comparisons carried out between the different approaches to try and find which method provides the best relief. This is a rather complex task as no one testing protocol can isolate the benefits provided by each system, making it quite difficult to perform direct comparison of the various methods. The recent work in [2] tries to provide an in depth analysis of a representative set of analyses that have been carried out by many authors along with their own sets of tests. The main theme of most of the testing procedures has been to try and evaluate how the different mechanisms affect the intelligibility of speech. Clearly it would be gratifying to see one method providing significant better intelligibility over all other methods, however this is not the case. Considering all the current approaches to loudness recruitment compensation, they can broadly be categorised into one of two types, either they are linear-amplification systems

or non-linear amplification systems. Linear-amplification systems cover all methods involving linear amplification, with and without frequency shaping. Non-linear amplification systems encompass all systems involving some form of non-linear compressive amplification. In general the debate comes down to which of these two categories is the best approach to adopt. Examination of the current state of knowledge in the area cannot prove conclusively that either approach is superior; however, there is a growing favour for the non-linear approach.

Under laboratory conditions and very specific types of loss (namely flat loss) linear methods have been shown to outperform non-linear methods [2], specifically linear methods can perform very well when optimally fitted to specific input signals and levels where they can typically outperform non-linear methods. The performance of linear systems however degrades substantially for inputs deviating from the optimum level [105], deviations which continually occur for real world signals. It has been shown that under real world situations only part of the decrease of performance produced by actual cochlear hearing loss can be compensated for by conventional linear amplification [85], [72]. Under real world conditions and situations of both flat and sloping hearing loss, the non-linear approach is as good as the linear approach and in many situations outperforms linear methods [85], [2], [178] and as such appears to be gathering favour considering the increasing number of compensation approaches being developed.

It is significant to note that variations on the non-linear processing approach are incorporated into the most recently developed hearing aids of the world's main hearing aid manufacturers to provide multiband compression. See Table 7.7.

Manufacturer	**Hearing Aid**
Siemens	MUSIC, InteliVenience, FRONTIER, Some non-programmable K-AMP
GN ReSound	Danavox Danalogic, ReSound Digital 5000 series
Oticon	DigiFocus II
Unitron	Digital Sound F/X

Table 7.7 Hearing Aid Manufacturers using non-linear processing.

Many of the currently available digital hearing aids are software programmable to allow optimal configuration for the user. In general, the hearing aids indicated above allow maximum flexibility for configuration of the compression employed in an attempt to

compensate for loudness recruitment. However, of all the hearing aids available only those provided by Siemens and GN ReSound have compression strategies flexible enough to begin tackling the problem of loudness recruitment and these come with a significant price tag.

Considering the multitude of comparisons performed between different methods and the benefits of multiband processing [179], [180], the approach taken in this work tries to tap into the positive aspects of the different methods in an attempt to harness their beneficial elements. The approach taken in this work can be considered a form of hybrid of the two possible approaches. Specifically the system proposed applies linear amplification to low level signals to ensure all signals above threshold in the normal system are above threshold in the impaired system; however, once the signals rise above threshold, the amplification becomes a non-linear compressive gain that is both frequency and intensity dependent. The compressive function not only depends on the frequency and intensity but also on the damage characteristics of the basilar membrane to ensure the neural information on the auditory nerve is the same in both the impaired and unimpaired systems. At present clinical intelligibility tests of this compensation mechanism have not been performed so it is difficult to appraise the effectiveness of this system relative to others. However, the simulations that have been carried out indicate that it fulfils the set goals in terms of the auditory model and its recruited form, that of restoration of the neural output of the model. Clearly such intelligibility tests need to be carried out to precisely evaluate the viability of the mechanism described in this Chapter; however, such tests are beyond the scope of the current work and will be re-visited in the concluding Chapter.

7.5 Summary and Conclusions

In summary it has been shown that a compensation mechanism for recruitment can be developed and tested using a model for the unimpaired and impaired auditory systems. The compensation mechanism consists of two sub-systems

- A system for the separation of stimuli into their component sub-bands followed by recombination of the sub-signals to synthesise a stimulus.

- A system for the introduction of compensatory measures into the sub-signals.

The system for the separation of stimuli into their sub-bands and the re-synthesis of the compensated stimuli uses a perfect-reconstruction cosine modulated analysis-synthesis filter bank to perform this operation. The design of this system depends on an optimised prototype low pass filter whose polyphase component matrix guarantees to be lossless thus ensuring perfect reconstruction, free from alias distortion, phase distortion and amplitude distortion. The system for the introduction of compensatory measures into the sub-band signals provides detection of the signal level prior to sub-band separation and after sub-band separation to generate an overall control level used in the determination of the level of compression required for each channel. The control level determined for each band is a weighted combination of the level of the complete signal and the level within each sub-band. The control level in each sub-band is used in conjunction with the characteristic compression curve for each sub-band to determine the required compression to apply in each sub-band. Having calculated the control level, the threshold gain and the compression required, the compensation system then applies suitable gain to the sub-band signals to introduce the necessary effects for optimal response in the impaired auditory system.

The net result of this compensatory mechanism is a pre-processor for signals going to the impaired auditory model that attempts to minimise the difference between the response of unimpaired system to uncompensated signals and the response of the impaired system to compensated signals. The compensation system has received significant testing and has shown that even under the worst-case conditions, it can significantly restore the dynamic range compression within the impaired model. As it stands the compensation system has not undergone subjective testing, notably the estimation of the benefits such a system may provide in increasing the intelligibility for an impaired user have not been determined. However, such testing is beyond the scope of the current work.

Chapter 8 - Model Characterisation

8.1 Introduction

The models described in the previous chapters emulate the functioning of the auditory system in its unimpaired form, in its recruited form, in its frequency smeared form and finally in its recruited form with recruitment compensation in place. Specifically, the unimpaired auditory model has been developed and configured based on available experimental data and as such has been characterised using this data. The models of the impaired auditory system have been developed to emulate varying degrees of impairments and must be characterised with the specific details of a particular instance of an impairment before they can emulate that impairment. In a similar manner the recruitment compensation system can only compensate for a particular instance of recruitment once it has been characterised based on the particular modelled instance of recruitment.

The process of model characterisation involves identifying some of the possible methods by which the information necessary to configure the different models may be acquired from an impaired auditory system. Only after the specific details of an impairment have been acquired can the models be configured to emulate what is taking place in the system. Once a suitable auditory model has been configured then the process of characterisation of a compensation mechanism can begin. The following sections deal with characterisation of the auditory model in its various forms and finally considers characterisation of the recruitment compensation mechanism.

8.2 Characterisation of the Unimpaired Auditory Model

At the time of writing, the current unimpaired auditory model has been designed and configured to emulate as closely as possible the currently available data on the auditory system. It would be naive to think that the current model is perfect and exactly replicates the processes of the human auditory system because a complete understanding of this very complex system is unavailable and is one of the main reason such models are needed, to gain a better understanding of this system. The evolution of knowledge in this area is an ongoing process involving modelling, experimentation and theorisation to gain a clearer understanding of the different aspects involved in the hearing process. As this understanding evolves, auditory models must also evolve to reflect the current state of

knowledge of the biological system to provide the best possible emulation of the hearing process with which future developments in the area may be made. The unimpaired auditory model laid out in this work has been built up in modular form to facilitate incorporation of such updated information as the understanding of the auditory system expands. Clearly the better the unimpaired auditory system can be modelled the better impairments of this system can be understood, modelled and alleviated. Chapter 4 has dealt with the design of the unimpaired auditory model and how it relates to the current experimental and modelling data available on the auditory system.

8.3 Characterisation of the Recruited Auditory Model

Before considering methods for acquiring recruitment information from an auditory system, it is important to specify exactly what information is required to characterise the auditory model so that it emulates a system suffering from recruitment. It is known that outer hair cell damage results in recruitment in the auditory response and this has been incorporated into the recruited auditory model through specification of a percentage of outer hair cell damage along the modelled basilar membrane. The basilar membrane is a continuous structure represented in a discrete model and it would be unreasonable to expect individual specification of outer hair cell damage for every discrete location along the basilar membrane. It is for this reason the recruited auditory model only allows specification of outer hair cell damage at a finite set of frequencies in the auditory range and employs cubic spline interpolation to calculate the damage at surrounding frequencies; this is not an unreasonable compromise as damage typically occurs in a continuous fashion.

A characterisation method is therefore required to identify the percentage of outer hair cell damage at particular locations within the auditory system. If this information is then used to configure the recruited auditory model then the model should be able to emulate the recruited auditory system. Considering the location of the outer hair cells, invasive direct measurement of the damage to the outer cells is not possible and alternate non-invasive methods must be considered. With this is mind three different auditory tests are considered.

- The alternate binaural loudness balance test. (ABLB)
- The loudness growth in 1/2-octave bands test. (LGOB)

- The distortion product otoacoustic emission test. (DPOAE)

These tests evaluate the effects outer hair cell damage has on the auditory response and the results are used to identify the amount of outer hair cell damage present in the auditory system. Considering the tests themselves the ABLB test is relatively well known among the battery of audiometric tests whereas the LGOB and DPOAE tests are not as well known.

8.3.1 The Alternate Binaural Loudness Balance Test

Before performing any of the specialised tests for recruitment a standard pure-tone audiogram test should be carried out to identify the presence of a sensorineural impairment, along with the threshold level and uncomfortable loudness level of the patient. The threshold level and the uncomfortable loudness levels are required to set the ranges of the test procedures.

The ABLB test is a well-known clinical test targeted at identifying the abnormal growth of loudness with intensity in the impaired human auditory system. There are two different forms of this test depending on whether the hearing loss is unilateral or bilateral; however, the general methodology for both forms is the same. In particular, this test requires the patient to compare the loudness of a tone that is heard within normal limits with a tone for which the patient has an impaired threshold. These comparisons are performed at suprathreshold levels and the test requires balancing the loudness of the two tones. The ABLB test is applied to patients with unilateral loss where tones are presented alternately to both the impaired and unimpaired ears at the same frequency and the intensity of the tone in the impaired ear is altered so its perceived loudness is the same for both ears. The alternate form of this test is the monaural loudness balance test (MLB) and is applied to patients with binaural hearing impairment. In this case two tones are alternately presented at different frequencies to one ear, where the threshold of one tone must be within normal limits, and the intensity of the other tone is altered so its perceived loudness is the same for both frequencies. Regardless of the method used the tests are carried out at various reference intensity levels and frequencies across the normal range of hearing. The results at each frequency characterise how loudness grows at that frequency in the impaired situation relative to the unimpaired situation. In this way a map of loudness growth can be constructed for the impaired auditory system that relates intensity at specific

frequencies with the loudness perceived. The specifics of the testing procedures can be found in many of the audiology reference texts, e.g. [15], [181].

In order that sufficient information be acquired to characterise the model the test procedure would typically have to cover octave frequencies from 250Hz up to 8kHz across the intensity range of the patient from the threshold of hearing up to the loudness discomfort level. Given that at each frequency approximately six intensities would have to be evaluated this would lead to 36 test points and at approximately 30 seconds per test point, would lead to a test time of around 18 minutes (without training or re-testing any point as a check). Figure 8.1 illustrates the typical type of curve produced by the ABLB test for a recruited auditory system at a single frequency [182].

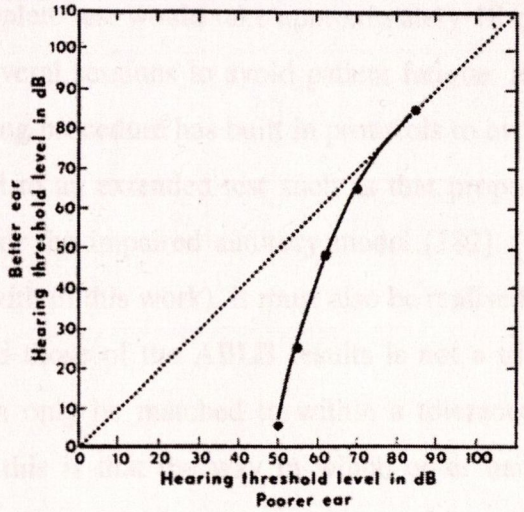


Figure 8.1 Typical result of the ABLB test.

Considering the solid curve in Figure 8.1, the loudness of a tone at 50dBSPL in the impaired ear is approximately the same as the loudness of a 5dBSPL tone in the unimpaired ear, showing the increase in auditory threshold in the impaired ear. Increasing the intensity in the unimpaired ear by 10dB (e.g. from 10 to 20dBSPL) requires only an increase of approximately 2dB in the impaired ear, showing that the rate of growth of loudness is increased in the impaired ear, i.e. the ear is suffering from recruitment.

The data provided by this type of test specifies how loudness grows at octave frequencies from 250Hz to 8kHz in the impaired situation relative to the unimpaired situation. Applying the same test tones across the same range of frequencies to the unimpaired auditory model provides unimpaired threshold and loudness growth references

for the test intensities in the unimpaired model. As the percentage outer hair cell damage is increased in the auditory model the threshold and the rate of growth of neural firing rate will increase relative to the reference levels measured in the unimpaired model. Once the increase in threshold and the relative rate of growth of loudness is the same in the model as in the results from the ABLB test then the model should be operating in a similar manner to the impaired ear. Applying this procedure at all the frequencies in the ABLB test will provide sufficient data for the model to be able to interpolate the damage levels for the entire basilar membrane in the model.

There are, however, several factors that must be considered. It would be insufficient to only perform the balance procedure once at each test point; more realistically it would have to be preformed several times to get a more accurate average result. Given that a complete test would take approximately 18 minutes, the testing would have to take place in several sessions to avoid patient fatigue. It is important to note that the standard ABLB testing procedure has built in protocols to avoid fatigue, adaptation etc. which would be applied to an extended test such as that proposed for acquisition of the necessary information for the impaired auditory model [182]. (The details of the testing protocols are not dealt with in this work). It must also be realised that matching the growth curves in the model and those of the ABLB results is not a trivial matter and generally speaking the curves can only be matched to within a tolerance, be it a small tolerance (~5%). The reason for this is that the way in which outer hair cell damage affects the dynamic compression in the model has been constructed based on averaged results from impaired ears, which will never exactly replicate the responses of an individual ear, it will, however, get very close.

In general, this method will provide sufficient data to characterise the recruited auditory model and involves a test that is currently available to the audiologist. The test requires extension over a larger range of frequencies but does not require additional knowledge or training on the part of the audiologist; a most beneficial factor.

8.3.2 Loudness Growth in 1/2-Octave Bands

The LGOB test is a more recently developed extension of a previous set of procedures based on absolute loudness judgements of presentations of ½-octave bands of periodic noise centred at 0.25, 0.5, 1.0, 2.0 and 4.0kHz that have been randomised over

level and frequency [183]. The presentation at each level is designed to evoke a subjective judgement from the patient given a pre-defined set of loudness ratings (one of: Very Soft, Soft, Ok, Loud, Very Loud, Too Loud). The specific details of the assessment procedure are defined in [183]. However, it is useful to note that it incorporates a two-phase approach, a "limits and practice" phase and a "data collection phase". The first identifies the threshold of audibility and the loudness discomfort level and provides a practice session for the patient to acquaint them with the procedure, the second involves the actual data acquisition phase which follows the same procedure as experienced in the practice session and uses the upper and lower limits to bound the intensity range. Considering the patient must provide subjective judgements on the loudness it is very important they understand the rating system in order that accurate results be obtained and that reasonable consistency is maintained in the test-retest data for both normal and impaired subjects. The complete procedure is stated to take less than 30 minutes for five frequencies. The typical form of the results is shown in Figure 8.2.

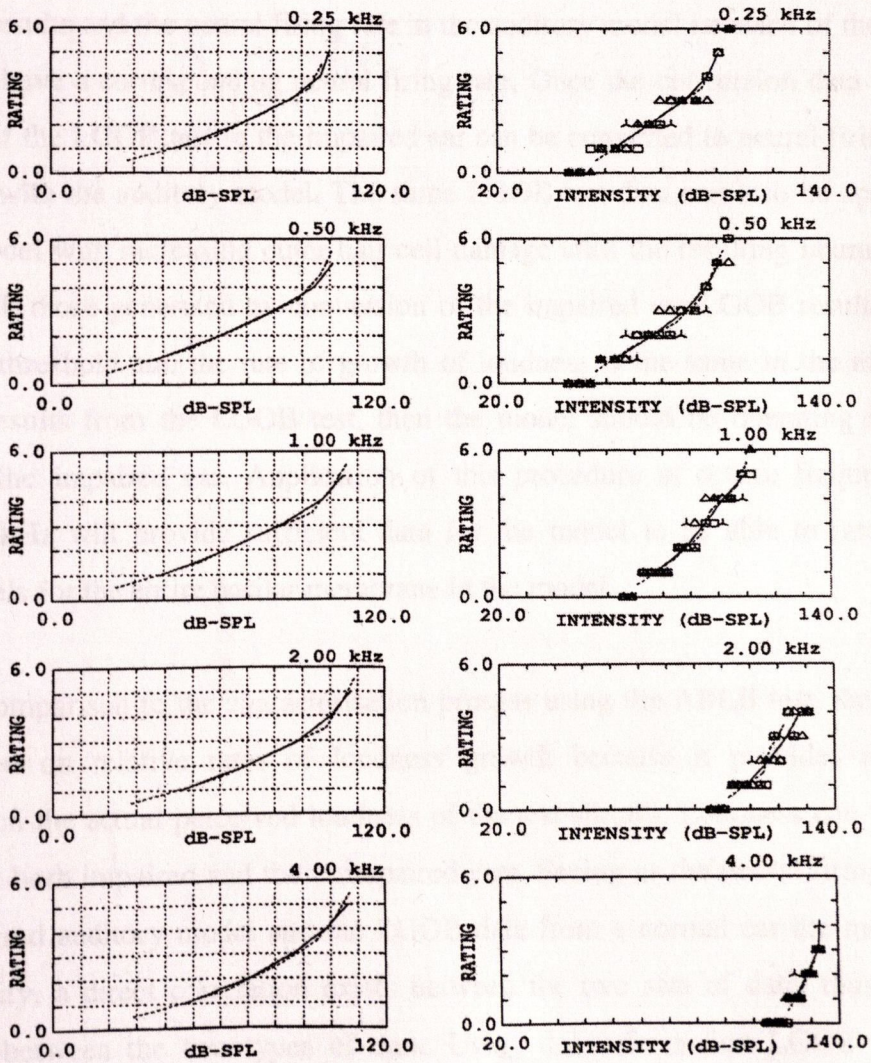


Figure 8.2 Typical results from an LGOB test. LHS normal. RHS Recruited.

Examining Figure 8.2, the left hand column of plots shows the average results of the LGOB test for a set of unimpaired ears and the right hand column shows the results of the same test for an ear suffering from recruitment. Each of the graphs plot the rating in terms of an assigned number, from 0.0 (No Response) up to 6.0 (Too Loud), for each intensity tested. The upper right hand number on each plot indicates the centre frequency associated with the set of results plotted. It is very evident comparing the two columns of graphs that the threshold in the impaired ear is increased and that the rate of growth of loudness is greater in the recruited ear. It just remains to specify how this data can be used to characterise the impaired auditory model.

Applying stimuli with the same frequency and intensity as those used in the LGOB test to the unimpaired auditory model will generate data relating neural firing rate with input intensity. Given that the stimuli applied both in the LGOB test and to the auditory model are the same, the resulting loudness measurements for the unimpaired ear and the unimpaired auditory model should be directly related. This data allows conversion between the LGOB results and the neural firing rate in the auditory model i.e. each of the subjective ratings will have a corresponding neural firing rate. Once the conversion data is available the results of the LGOB test in the impaired ear can be converted to neural firing rate data compatible with the auditory model. The same LGOB test data can also be applied to the auditory model with increasing outer hair cell damage until the resulting neural firing rate curves match those generated by conversion of the impaired ear LGOB results. Once the increase in threshold and the rate of growth of loudness is the same in the model as the converted results from the LGOB test, then the model should be operating in a similar manner to the impaired ear. Application of this procedure at octave frequencies from 250Hz to 8kHz will provide sufficient data for the model to be able to interpolate the damage levels for the entire basilar membrane in the model.

In comparison to the characterisation process using the ABLB test, the LGOB test is not based on relative rates of loudness growth because it provides a subjective judgement on the actual perceived loudness of the test stimuli. Loudness can therefore be assessed for both impaired and the unimpaired ears. Seeing as the neural firing rates from the unimpaired auditory model and the LGOB data from a normal ear are measuring the same quantity, a direct correlation exists between the two sets of data, thus allowing a conversion between the two types of data. Using this information LGOB data can be directly converted to neural firing rate data that the auditory model can be configured to

generate. In this regard the process of matching the neural firing rate curves (from the impaired model) and the converted impaired LGOB curves is a direct process. The impaired auditory model can therefore be configured to more closely emulate the impaired auditory systems characterised by the LGOB test results.

There are several factors associated with this test that must also be noted. Clearly the LGOB test would have to be expanded to cover the range of 250Hz to 8kHz to provide sufficient coverage for the interpolation process in the auditory model thus extending the time taken for the test beyond 30 minutes, which is quite long for a clinical procedure. This may raise important issues such as patient fatigue, however, this is easily overcome by performing multiple test sessions on sub-sets of the desired frequencies. Another more important factor to be aware of is that only a limited set of descriptions is available to the patient, which brings in the whole issue of quantisation error. Strategies for handling this issue are laid out in the specifications for the testing procedure [183] but it must be acknowledged when considering using LGOB rating curves.

In general this procedure provides a more direct method by which the impaired auditory model may be characterised, allowing more accurate mapping of the impaired auditory model to the test results. The LGOB test is quite specialised, requiring special equipment and is quite time consuming. However, it does provide a wealth of information in a form directly applicable to the auditory model. Once the quantisation issues are adequately catered for, this test is very suitable for characterising the impaired auditory model. Given that this is a non-standard test it may require extra training on the part of an audiologist to be able to perform this test.

8.3.3 Distortion Product Otoacoustic Emissions

Otoacoustic emissions have recently received significant attention in the exploration of their use as an auditory diagnostic tool. In particular, distortion product otoacoustic emissions appear to be promising for the acquisition of information regarding outer hair cell function [109], [184]. Chapter 5, section 5.4 has dealt with distortion products, their generation and the effect outer hair cell damage has on the combination tones, this section focuses on how these distortion products and the resulting DPOAEs may be quantified and used to assess outer hair cell damage at specific locations along the basilar membrane with a view towards characterising the recruited auditory model.

When two tones of differing frequency (f_1, f_2 typically $f_1 \langle f_2$) are applied to the auditory system they nonlinearly interact on the basilar membrane to generate combination tones. Of particular interest are the third order distortion products generated at frequency $2f_1 - f_2$. These combination tones can be heard by a patient and can also be detected in the outer ear, as the auditory system reflects the combination tones back out through the middle and outer ears, emerging as distortion product otoacoustic emissions. Given that distortion products are a result of the nonlinearity of the basilar membrane and that the outer hair cells are responsible for this nonlinearity, the outer hair cells are also responsible for the generation of distortion products (and thus for the DPOAEs). Considering this fact, distortion products and DPOAES are an objective measure of the outer hair cell nonlinear compression [61], [109]. Reduction in the outer hair cell function leads to a reduction in the nonlinear compression on the basilar membrane, which in turn alters the distortion products and hence the DPOAEs. Comparing the distortion products and DPOAEs for normal ears with those of an impaired ear will identify reductions in outer hair cell function. It is believed the site of generation of the distortion products is near the frequency f_2, therefore a reduction in distortion product or DPOAE will identify a reduction in outer hair cell function in the vicinity of frequency f_2. Considering distortion products only exist within the auditory system and DPOAEs only exist outside the auditory system, two different methods are required for their quantification.

The first quantification method is the psychoacoustical *method of cancellation*. At its core this method seeks to modify the level and phase of an external tone at frequency $2f_1 - f_2$ in order to cancel either the distortion product within the auditory system or the resulting DPOAEs outside the auditory system. Regardless of the form of cancellation performed, the auditory system is stimulated as indicated above. However, the forms differ in that the listener modifies the phase and level of the external stimulus until the internal distortion product is inaudible whereas the examiner manipulates the level and phase of an external signal to cancel the DPOAEs. Both forms of this method result in phase and level measurements of the external cancelling stimulus that directly relate to the level and phase of the internal distortion product and the external DPOAEs. It is important to note that cancellation occurs when the signal being cancelled is reduced to the noise floor of the system. Once the signal levels reach the noise floor, distortion products (and DPOAEs) cannot be distinguished from background noise. The specific procedures for this method are detailed in [149]. Figure 8.3 [149] illustrates an example of the spectra derived for a

single set of primary test tones and intensities. There are three plots in Figure 8.3 offset from one another by 20dB for clarity. The top plot shows the two primary tones at $f_1 = 1620Hz$, $f_2 = 1851Hz$ and the external stimulus needed to cancel the distortion product at $2f_1 - f_2 = 1389Hz$. The middle plot shows the two primary tones and the resulting DPOAE at $2f_1 - f_2 = 1389Hz$. The bottom plot shows the two primary tones and cancellation of the DPOAE by an examiner. These two forms of cancellation would have to be performed several times at various levels for each set of primary tones to get a good estimate of the various parameters involved.

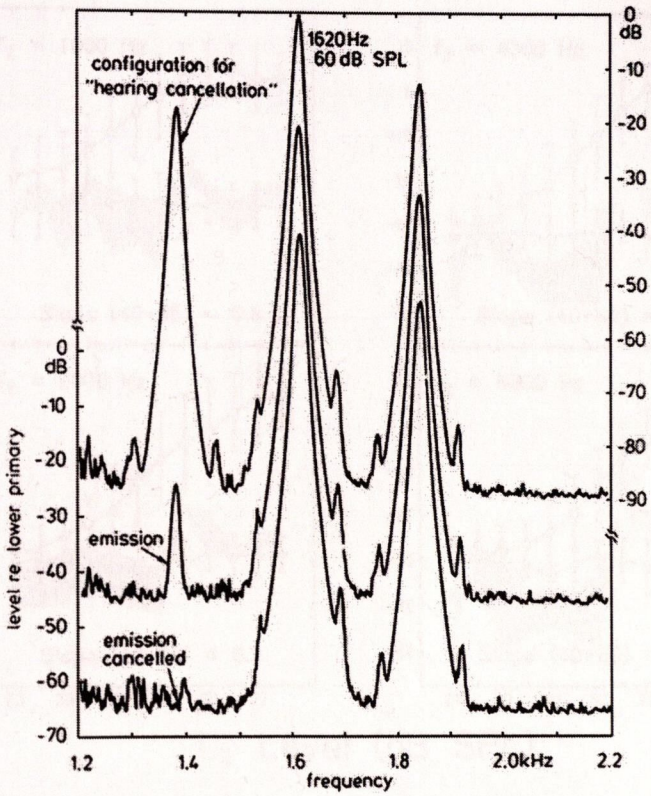


Figure 8.3 Spectra for primary tones, DPOAE, external stimulus used to cancel internal distortion product and external cancellation of DPOAE in the method of cancellation.

The second quantification method is the *method of emissions*. This method focuses on detection of the $2f_1 - f_2$ distortion product as an otoacoustic emission by placing a sensitive microphone in the closed ear canal. The ear is stimulated as indicated above and the otoacoustic emission is measured. One very nice feature of this method is that it does not require any judgements or comparisons by the listener. The specific procedures involved in performing these measurements are detailed in [150]. Figure 8.4 [150] illustrates the average results for normal ears over a range of f_2 frequencies. Each plot

contains three sets of data. The curves joining circles represents the mean DPOAE measurements, the curve joining the triangles represents the noise measurements during the test and the greyed out areas represent mean estimates of the system distortion ± 1 standard deviation. The error bars represent ± 1 standard deviation. Each set of results was measured for various intensity levels of the primary tones such that $L_1/L_2 = 10dB$ (L_1 and L_2 are the levels of the f_1 and f_2 tones respectively) and the ratio of the frequency of the primary tones was maintained at $f_2/f_1 = 1.2$. The DPOAE measurements for each combination of primary tones and intensity levels were taken 10 times for accuracy.

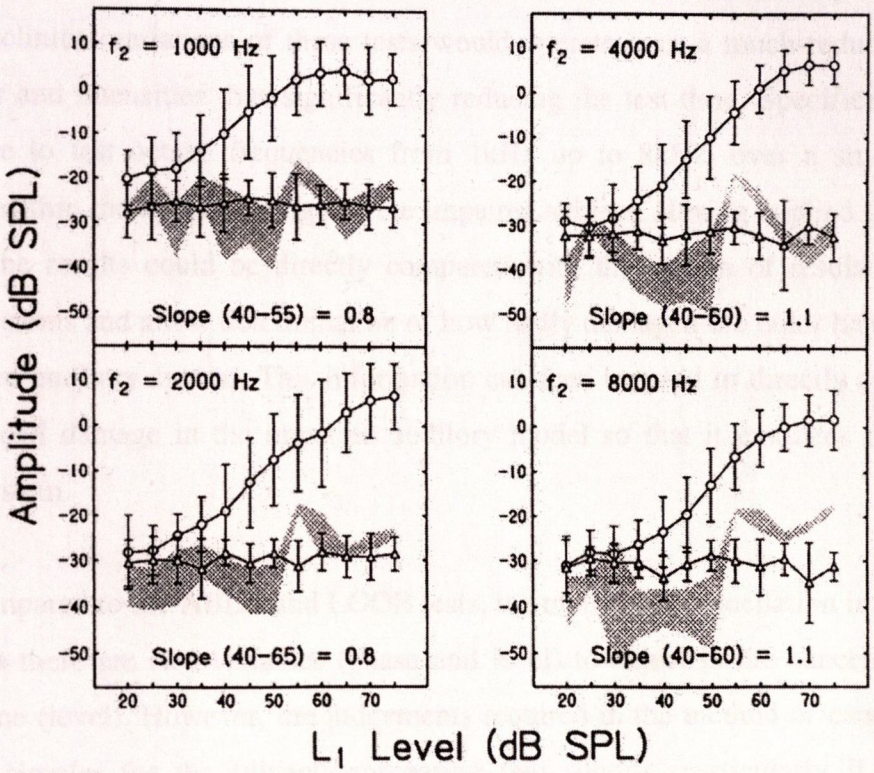


Figure 8.4 DPOAE measurements, noise measurements and system distortion estimates in the method of emissions.

Various studies have shown that distortion products and DPOAEs are highly reproducible for f_2 frequencies ranging from 1kHz to 8kHz over a wide range of intensities [150]. It has also been shown that for frequencies below 1kHz it becomes much more difficult to observe DPOAEs. Therefore in terms of characterising the auditory model, alternate methods would have to be used for lower frequencies.

Considering the procedures described in [150] and [149], they provide exhaustive tests across wide ranges of frequencies and intensities; [149] cites a 2-hour test time for

completion of the method of cancellation for a single combination of frequencies over the range of intensities used. The method of emissions is more time efficient requiring around 7 minutes for a pair of primary tone frequencies each at a particular level, however it still takes over 2 hours for 18 pairs of intensities. Clearly a 2-hour test time is not feasible. However, it must be noted that these procedures are experimental and designed with a different goal in mind. They are required to generate responses for wide ranges of intensities and frequency combinations, providing valuable reference information for the unimpaired auditory system.

In terms of acquiring sufficient information to characterise the impaired auditory model the clinical variations of these tests would operate over a much-reduced range of frequencies and intensities thus significantly reducing the test time. Specifically the tests would have to test octave frequencies from 1kHz up to 8kHz, over a small range of intensities within the dynamic range of the impaired subject. Having applied these tests to a subject the results could be directly compared with the wealth of results for normal auditory systems and allow determination of how badly damaged the outer hair cells are in the impaired auditory system. This information can then be used to directly configure the outer hair cell damage in the impaired auditory model so that it emulates the impaired auditory system.

Compared to the ABLB and LGOB tests, the method of cancellation is a little more complex as there are two variables (phase and level) to adjust in the cancellation rather than just one (level). However, the judgements required in the method of cancellation are somewhat simpler for the subject; comparing two sounds (particularly if at different frequencies) is more difficult than judging when a sound is present of not. Considering the task of the examiner in the method of cancellation, it is significantly more involved than for either the ABLB or LGOB tests but is not overly complex and the skills could be easily acquired with training. Of all the tests, the method of emissions is the most straightforward for both the subject and the examiner. This test does not require the patient to do anything as it probes the ear and records the responses. All the examiner has to do is present the tones to the patient and let the apparatus record the emissions. Regardless of the method used the levels of the primary test tones must be determined for each test; considering the levels must be within the dynamic range of the subject, an audiogram prior to this type of testing could easily determine the limits.

Considering these types of tests are not standard, specialised equipment and training would be required to be able to perform such tests. It has, however, been shown that for frequencies above 1kHz such tests can easily distinguish between unimpaired and impaired ears [150]. Given the wealth of information on healthy ears it seems the logical step to use these tests to acquire the information needed for characterising the auditory model in its impaired form. There is still, however, significant exploration required into the distortion products and DPOAEs generated by impaired ears. This topic will be revisited in the final chapter.

8.4 Characterisation of the Frequency Smeared Auditory Model

Frequency smearing in the auditory system is caused by damage to the outer and inner hair cells. This damage has various effects in the auditory system depending on the extent of the inner and outer hair cell damage. Outer hair cell damage predominantly causes an increase in the auditory threshold and loudness recruitment as indicated in Chapter 5. However, it also causes broadening of the tuning of the basilar membrane and shifting of the tuning of the basilar membrane towards lower frequency thus altering the selectivity of the auditory system. Inner hair cell damage predominantly causes a reduction in the sensitivity of the auditory system. The combination of the effects of outer and inner hair cell damage result in frequency smearing in the auditory system and are the sources of frequency smearing in the impaired auditory model. In a similar manner to the specification of outer hair cell damage in the recruited auditory model, inner hair cell damage in the impaired auditory model is specified as the percentage inner hair cell damage along the modelled basilar membrane at particular frequencies in the auditory range. The damage at all other frequencies is interpolated from the specified damage. In order that the auditory model emulates an auditory system suffering from frequency smearing, both inner and outer hair cell damage must be assessed and specified in the impaired model. Section 8.3 has dealt with methods for identification of outer hair cell damage; therefore this section will focus on assessment of inner hair cell damage.

The task of quantifying inner hair cell damage depends on assessing the subjective sensitivity of the auditory system. Considering the effects of inner hair cell damage as laid out in Chapter 6, a method is required to identify the auditory systems ability to discriminate a stimulus at a specific centre frequency within a broad band signal centred on that frequency. The intensity at which the auditory system can discriminate between the

stimulus at the centre frequency and the broadband stimulus, is dependent on the inner hair cell function in the region of the centre frequency. Comparison of the intensity at which detection occurs in the unimpaired and impaired auditory systems allows assessment of the inner hair cell damage.

This type of analysis is of the same form as that in the notched-noise masking procedure for the measurement of auditory filters. The procedure involves presentation of a tone at the centre frequency to be analysed and masking it with broadband noise with a sharp-edged notch centred on the tone. The tone threshold is then measured as a function of notch width. In practice, the notched-noise procedure provides excessive information as it examines multiple notch widths per centre frequency. However, this procedure identifies the optimal range of *notch width to centre frequency ratios* over which the required tone threshold measurements should occur; between 0.3 and 0.4 [185].

In order to acquire sufficient information to characterise the impaired auditory model a modification of the notched-noise masking procedure as specified in [185], [186], [187] would have to be performed on each ear. This procedure would perform tone threshold measurements at an optimal notch-width across the octave frequencies between 250Hz and 8kHz for several noise levels. The threshold measurement typically takes the form of an adaptive, two-interval, force-choice, maximum-likelihood procedure [188], [189]. This information could then be compared to similar data for the unimpaired auditory system allowing assessment of the inner hair cell damage. Considering the application of the notched-noise masking procedure at 3 centre frequencies with 8 different notch-widths in [185] takes approximately 1 hour per ear, the modified version would take approximately 45 minutes based on 3 different noise level presentations.

Figure 8.5 illustrates threshold curves generated from the notched noise masking procedure in [185]. The plots show threshold curves for four listeners at three signal frequencies, 0.5kHz, 2kHz and 4kHz. The left ear measurements are shown using filled symbols with solid lines and the right ear using empty symbols with broken lines. The ages of the listener are given with each set of curves and the origin is displaced to the right, 0.1 times the ratio of the notch width to the centre frequency, for each successive listener. These particular results arise form an experiment designed to analyse the deterioration of hearing with age.

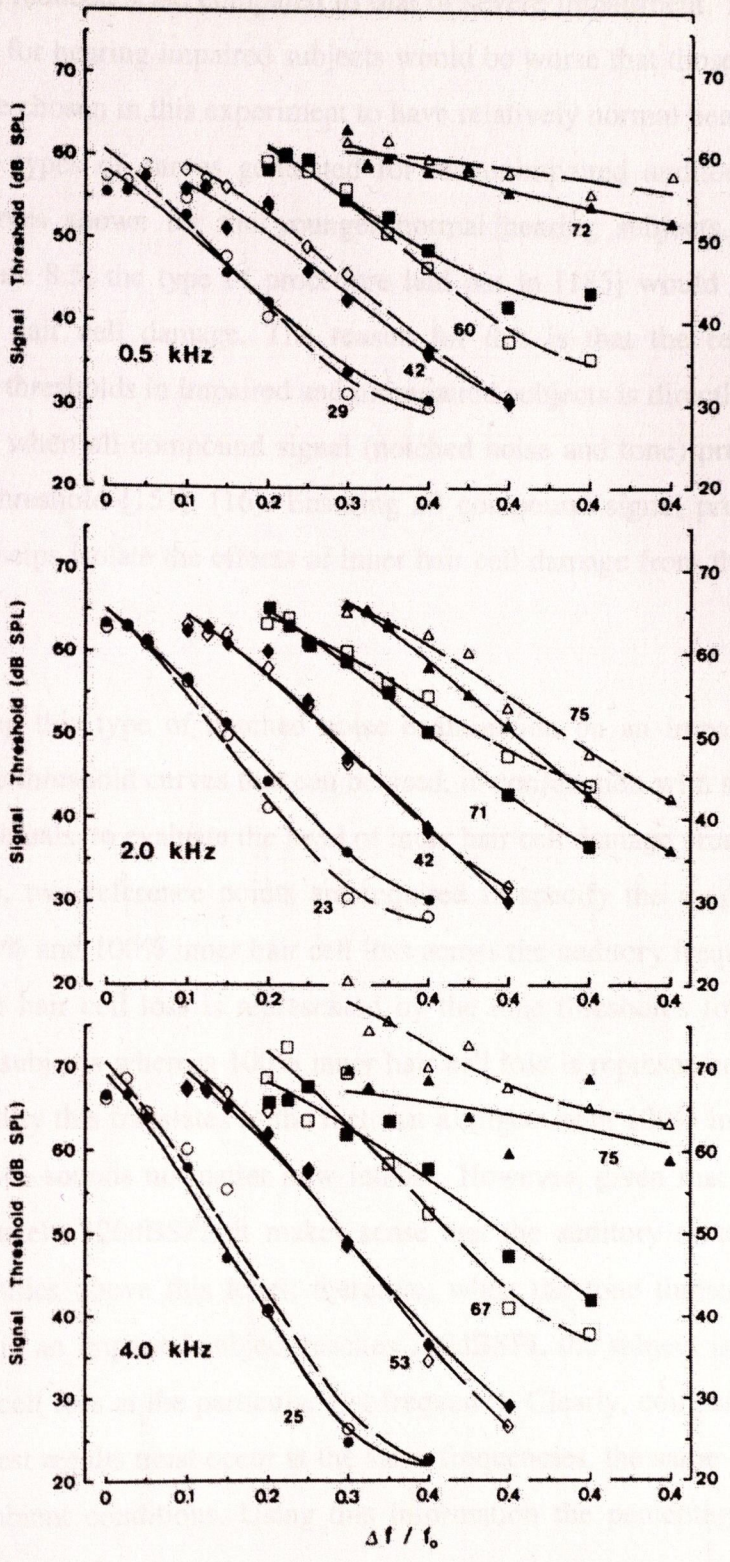


Figure 8.5 Threshold curves measured using the notched noise masking procedure.

Considering the various plots in Figure 8.5, a comparison of the curves for younger subjects with older subjects shows the tone threshold increasing with age. This can be attributed to a reduction in the effectiveness of the hair cells in the auditory system. Over time the hair cells experience a reduction in their ability to function, something akin to hair

cell damage on a reduced scale compared to that of severe impairment. Thus the types of curves generated for hearing impaired subjects would be worse that those shown for older people (who were chosen in this experiment to have relatively normal hearing for their age bracket) and the types of curves generated for the unimpaired auditory system would follow those curves shown for the younger normal hearing subjects. Given the data produced in Figure 8.5, the type of procedure laid out in [185] would also be ideal for evaluating inner hair cell damage. The reason for this is that the relative difference between the *tone* thresholds in impaired and unimpaired subjects is directly related to inner hair cell damage when all compound signal (notched noise and tone) presentations occur above absolute threshold [151], [16]. Ensuring all compound signal presentations occur above threshold helps isolate the effects of inner hair cell damage from those of outer hair cell damage.

Performing this type of notched noise examination on an impaired subject will provide their tone threshold curves that can be used, in conjunction with similar curves for unimpaired individuals, to evaluate the level of inner hair cell damage present. In order that this can be done, two reference points are required to specify the ranges of thresholds associated with 0% and 100% inner hair cell loss across the auditory frequency range. The case of 0% inner hair cell loss is represented by the tone thresholds for an otologically normal group of subjects whereas 100% inner hair cell loss is represented by infinite tone thresholds. In reality this translates to the fact that a subject with 100% inner hair cell loss will not experience sounds no matter how intense. However, given that the threshold of pain is approximately 120dBSPL it makes sense that the auditory system should never experience intensities above this level; therefore, when the tone thresholds required to elicit a response in an impaired subject reaches 120dBSPL the subject is deemed to have 100% inner hair cell loss at the particular test frequency. Clearly, comparing the impaired and unimpaired test results must occur at the same frequencies, the same notch widths and under similar ambient conditions. Using this information the percentage inner hair cell damage can be calculated as

$$D = \frac{IT - UT}{120 - UT} \times 100 \qquad (8.1)$$

where D is the percent damage, UT is the unimpaired threshold and IT is the impaired threshold. Applying this analysis to the threshold curves generated at octave frequencies

between 250Hz and 8kHz would generate sufficient data to configure the inner hair cell damage in the auditory model.

In general this procedure provides a direct method for characterising the inner hair cell damage in the auditory model. Combining the inner hair cell damage data and outer hair cell damage data allows configuration of the auditory model so it emulates an auditory system suffering from frequency smearing. It must, however, be noted that the notched noise masking procedure is an elaborate one requiring special equipment and training to perform the tests successfully.

8.5 Characterisation of the Recruitment Compensation System

The recruitment compensation system is designed to manipulate the stimuli being presented to the impaired auditory model in an attempt to restore the neural firing patterns in the recruited model back to the same levels as those in the unimpaired model when it is stimulated with the same stimuli without compensation. In this manner, the recruitment compensation system must be characterised based on neural information from the recruited auditory model, as the necessary information cannot be measured directly from the impaired auditory system. Provided the impaired model emulates the recruited ear sufficiently accurately then the information elicited from the recruited model should be a valid representation of similar information that could have been acquired from the impaired auditory system (were it available). The reason for not acquiring the information from the impaired ear directly is due to the nature of the data required. It is neither ethical nor feasible to measure the neural firing of the inner hair cells within the human auditory system due to the highly destructive nature of the surgery that would have to be performed in order to measure the data. However, provided the model accurately emulates the functions taking place in the auditory system there is no problem extracting the information from the impaired auditory model.

Specifically the recruitment compensation system requires the impaired neural firing data across the complete dynamic range at the band edges and centre frequencies of all the analysis filters in the analysis/synthesis system. Clearly the number of firing curves is $2M+1$ given the number of channels is M. Having calculated this information a representative curve is generated for each band based on an average of the three curves for each band. Similar data is extracted from the unimpaired model and the two sets of data are

used to determine the compression characteristics and threshold gain for each band as indicated in Chapter 7. Figure 8.6 illustrates the typical types of unimpaired and impaired data extracted from the model at 2kHz and 4kHz respectively. The neural response is shown relative to the driven response at 1kHz (see Figure 5.5) to allow comparison between the responses, therefore the resulting apparent response below 0dBSPL actually indicates the system is more responsive at 2kHz and 4kHz as is expected given the transfer characteristics of the outer and middle ears. A complete explanation of these types of plots can be found in Chapter 5; however, it suffices to say that the thresholds are increased and the rate of growth of neural firing is approximately 1.6 times greater with outer hair cell damage than without.

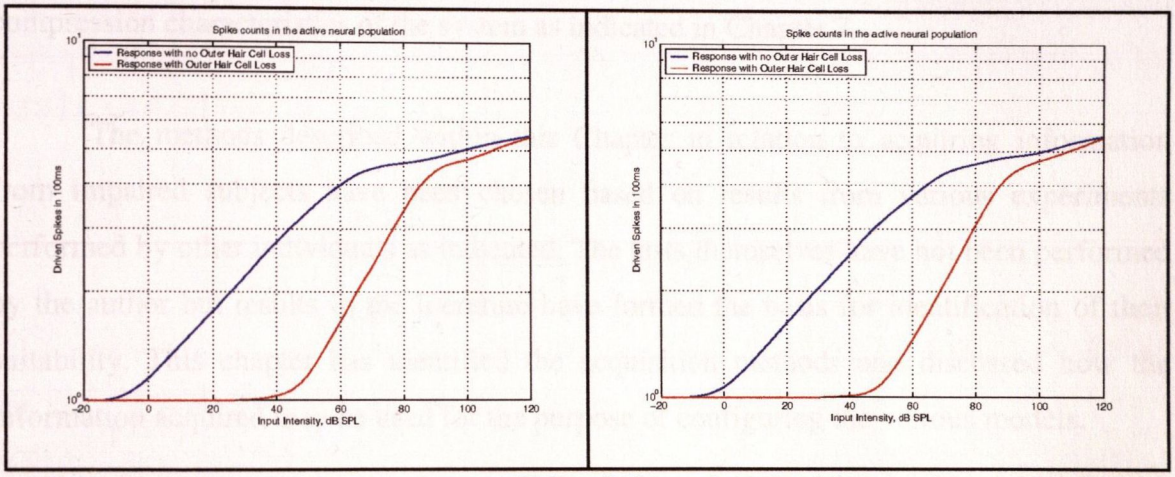


Figure 8.6 Neural response data at 2kHz (left) and 4kHz (right) with and without outer hair cell damage.

8.6 Summary and Conclusions

This chapter has focused on identifying methods for acquiring the information necessary to characterise the models in this work. The methods can be divided into two sections, the first involves characterising the auditory model so it emulates an impaired auditory system and the second involves characterising the compensation system to counter the effects of recruitment.

Characterising the impaired auditory model requires specific information about the causes of the impairment. In the case of recruitment, outer hair cell damage must be quantified to characterise the model. Three tests are identified to quantify outer hair cell damage; the alternate binaural loudness balance test, the loudness growth in ½-octave bands test and the distortion product otoacoustic emission test. Each method has been

described with suitable references to testing protocols and their merits discussed with regard to each other. In the case of frequency smearing, both inner and outer hair cell damage must be assessed to perform the necessary characterisation. The notched noise masking procedure is identified to quantify inner hair cell damage. Outer hair cell damage can be identified with the methods described for the case of recruitment.

Characterisation of the compensation mechanism for recruitment requires specific information from the recruited auditory model because the information cannot be acquired directly from a subject as explained. Specifically the compensation system requires information about the neural response of the auditory model to stimuli across the dynamic range of hearing. This information is used to characterise the threshold gain and the compression characteristics of the system as indicated in Chapter 7.

The methods described within this Chapter in relation to acquiring information from impaired subjects have been chosen based on results from various experiments performed by other individuals as indicated. The tests themselves have not been performed by the author but results in the literature have formed the basis for identification of their suitability. This chapter has identified the acquisition methods and discussed how the information acquired may be used for the purpose of configuring the various models.

Chapter 9 – Summary and Conclusions

9.1 Introduction

This final Chapter provides a summary of the models presented in this work in conjunction with a discussion of the results generated by each of the models to various forms of stimulation. Considering the significant body of work presented for each of the models, unimpaired auditory model, recruited auditory model, frequency smeared auditory model and recruitment compensated recruited auditory model, each will be treated individually to highlight their significant features. Some concluding remarks will be passed on each of the models, followed by a critical evaluation of the work as a whole. This Chapter will then close with some suggestions for future additions and expansions to the current material.

9.2 Summary of Models and Results

Each of the models developed in this work provides a valuable source of information regarding the particular states of the auditory system, be it unimpaired, impaired or impaired with compensation. Since these models are computational, there are no limits to the measurements that can be extracted from them, thus providing access to the types of information that cannot be easily measured in the biological system. This information is of vital importance in determining the complex interactions taking place within the auditory system and in the development of a better understanding of this system. As can be seen from this work, the modelling efforts of the past provide significant stepping-stones in the development of more accurate models today; in a similar manner the models of today will provide important insights for future developments. The following four sub-sections summarise the various models and results presented in this work.

9.2.1 The Unimpaired Auditory Model

The unimpaired auditory model consists of three primary sections modelling the outer, middle and inner ears, with particular emphasis on the inner ear to facilitate sensorineural impairment modelling. The outer and middle ear models are designed to superimpose the typical threshold of audibility curve on stimuli presented to the model,

specifically providing attenuation for very low and high frequencies, thus setting the frequency range of audibility for the model. Both the outer and middle ear models are static and configured for normal levels of stimulation. Importantly, the contributions of the middle ear model are specified to account for moderate acoustic reflex associated with normal levels of stimulation. The inner ear model consists of several sub-models emulating the mechanical motion of the basilar membrane, the resonance of motion between the basilar and tectorial membranes, the mechanical to neural transduction of the inner hair cells and feedback within the inner ear. Of particular importance is the tuning of the inner ear model arising from adaptive level dependent pseudoresonance, which is built up through the cascade of filter sections. The adaptive nature of the cascade depends on the stimulus level, the stimulus frequency content and the outer hair cell activity within the model. The modelled outer hair cell activity also facilitates incorporation of dynamic level dependent compression into the model. The net result of all these interactions is to incorporate highly selective, level dependent tuning into the inner ear model comparable to that found in the auditory system.

The results of various simulations presented in this section were specifically chosen to both demonstrate the characteristics of the complete unimpaired auditory model and to facilitate comparison of the capabilities of this model with other auditory models and with the responses of the auditory system. In particular, the following responses are shown.

1. **The response of the outer and middle ear models.** These illustrate the characteristics to be superimposed on stimuli entering the inner ear model. Notably, attenuation of high and low frequencies in accordance with the typical threshold of audibility measurements for an otologically normal population incorporating high sensitivity for mid-frequency stimuli.

2. **Tuning of the inner ear model.** The tuning curves extracted show the broad tuning of the basilar membrane alone, followed by the enhancement of its selectivity due to the interaction of the tectorial membrane and the adaptive level dependent feedback system. The tuning curves are compared to similar tuning curve measurements of other important models, to the experimentally measured tuning curves in other species and to the psychophysically measured tuning curves of the human auditory system. The tuning curves in the current model compare very favourably to those measured

experimentally (both in humans and in other species) and generally show greater correlation with experimental measurements than other models.

3. **Tone burst response.** Extracting the tone burst response of the auditory model is designed to show the temporal characteristics of the inner hair cell adaptation and the steady state response. In particular, it illustrates the transition from rapid to short term adaptation and the reduction in latency with stimulus level. The response is shown for two different approaches to modelling the inner hair cells both producing results in very good agreement with experimental data. In comparison to other models using similar approaches, the current model generates results in much greater agreement with experimental data.

4. **Compression in the auditory model.** The compression characteristics of the model are primarily extracted for comparison with the compression characteristics of the auditory model suffering from recruitment. However, a comparison of the results for the current model with other models shows the compression characteristics of the current model to be in much closer agreement with the known compression occurring in the auditory system, typically being within 4% of experimentally measured levels.

5. **Ramped and damped response.** Examination of the ramped and damped responses of the auditory model has the objective of showing in the auditory spectra that the width of the spectral peak due to a damped sinusoid is narrower than that due to a ramped sinusoid for various half-lives. This permits differentiation between these stimuli, a task the auditory system can do. Models based on Fourier energy spectra, however, cannot do this because the energy spectrum of these stimuli are identical. The results of simulations involving such stimuli clearly show that the current model has the ability to discriminate between damped and ramped sinusoids.

6. **Response to synthetic speech.** The stimulation of the auditory model with a synthetic syllable is primarily designed to show that the model is capable of responding to the complex content of such stimuli. In particular, the neural response should contain the formant content and transitions occurring in the stimuli being presented. The results presented in this work illustrate that the current model is very capable of neurally coding the complex content of synthetic syllables and that it performs very favourably compared to the response of other models.

9.2.2 The Recruited Auditory Model

The design of the unimpaired auditory model was specifically developed to facilitate the modelling of sensorineural hearing impairment. The first impairment to be modelled was recruitment, which causes a nonlinear loss of dynamic range compression resulting in increased auditory thresholds and an abnormally rapid growth in loudness. It is known that outer hair cell damage is the primary cause of this impairment in the auditory system; therefore incorporation of outer hair cell damage into the auditory model should produce the same effects in the model as occur in the auditory system. Given the form of the model, specification of outer hair cell damage translates into a reduction in the level of outer hair cell performance in the auditory model. As the damage level increases, the level of outer hair cell operation reduces. If the model is an accurate representation of the operations taking place in the auditory system, this should cause a loss in dynamic range compression resulting in the effects of recruitment appearing in the responses of the impaired model.

The results of various simulations presented in this section were chosen to examine the effects of outer hair cell loss on the characteristics of the auditory model and to show that outer hair cell loss in the auditory model would introduce a loss of dynamic range compression in the model. The simulations were also designed to examine if the effects of recruitment would be reproduced by the impaired model. The following responses are shown.

1. **Tuning of the inner ear.** Considering the contributions of the outer hair cells to the sensitivity and selectivity of the tuning of the basilar membrane, recruitment causes a reduction in both of these characteristics in the auditory system. Extraction of the tuning curves from the recruited auditory model, with increasing levels of outer hair cell loss, reveals tuning curves with increasing peak levels and increasing tuning curve bandwidths. As the peak level rises, this means the sensitivity is reducing and as the tuning curve bandwidth increases, the selectivity is reducing, as required.

2. **Response to rising stimulus level.** The purpose of stimulating the auditory model with rising stimulus levels is to examine how loudness grows in the model. As the level of a stimulus presented to the auditory system rises above threshold, the loudness experienced increases. When recruitment is present, the threshold at which the stimulus

is first experienced should increase and once this elevated threshold is breached, the growth of loudness should be abnormally rapid. It is known that loudness correlates with the driven spike count in the active neural population. Stimulating the recruited auditory model with stimuli of rising level and measuring the driven spike count in the active neural population, it has been shown that with increasing levels of outer hair cell loss, the threshold of audibility increases and once this threshold is breached the growth of driven spike count in the active neural population is abnormally rapid, showing excellent agreement with the available published data on recruitment.

3. **Tone burst response.** The effects of recruitment on the tone burst response of the auditory system are to reduce its sensitivity to low intensity signals and cause a loss of compression for high intensity signals. Stimulating the recruited auditory model with tone bursts of low intensity yields significantly reduced neural firing and when the tone burst intensity is high, the steady state neural firing is significantly increased, indicating a loss of compression, as expected.

4. **Response to synthetic speech.** The effect of outer hair cell loss on the auditory system is to reduce the system's ability to accurately code the complex content of acoustic stimuli in the neural information going to the brain. Examining the neural output of the recruited auditory model to a synthetic syllable shows that as the level of outer hair cell loss rises, the ability of the model to reproduce the formant information in the neural output reduces. As the threshold increases, much of the low level content is lost and in particular as the tuning becomes increasingly broad the distinction between formants reduces and formant transitions become increasingly difficult to distinguish.

5. **Compression in the recruited auditory model.** When the auditory system suffers from recruitment it experiences a loss of dynamic range compression. Measurement of the dynamic range compression in the recruited auditory model and comparing it to the compression in the unimpaired model, it can clearly be seen that with increasing outer hair cell loss the level of compression reduces.

9.2.3 The Frequency Smeared Auditory Model

The second impairment to be modelled was frequency smearing, which causes the auditory system to be unable to appropriately respond to the frequency domain resonances

within speech. It is known that inner and outer hair cell damage is the primary cause of this phenomenon. Increasing levels of damage to either or both types of hair cells cause increased levels of frequency smearing resulting in decreased definition in the neural information going to the brain, resulting in a reduction of intelligibility. The worst possible case occurs when both types of hair cells are damaged. Provided the auditory model is a suitable representation of the auditory system, the introduction of damage to inner and outer hair cells should cause the model to experience the same effects and product the same responses as the auditory system suffering from frequency smearing. Given the form of the model, specification of both inner and outer hair cell damage translates into a reduction in the level of inner and outer hair cell performance in the auditory model. As the damage level increases, the level of hair cell activity reduces and the model should generate the effects of frequency smearing in its output, namely reductions in the neural activity and broadening of the neural responses.

The results of various simulations presented in this section were chosen to examine the effects of different types of hair cell loss on the neural response of the auditory model and to show that hair cell loss in the auditory model would introduce reductions and broadening of the neural activity in the output of the model. The following responses are shown.

1. **Tone burst response.** The tone burst response is used to examine the neural activity in and around the centre frequency of the stimulus. The results are shown for varying degrees of inner and outer hair cell loss both individually and combined. Inner hair cell damage alone is shown to cause a reduction in firing at the affected centre frequency. As the level of loss increases the firing at the centre frequency become indistinguishable from the surrounding neural activity. Total inner hair cell loss stops all neural activity in the affected region. Outer hair cell damage alone is shown to cause a reduction in the neural activity and broadening of the neural activity around the stimulating frequency. As the outer hair cell loss increases the finely tuned neural activity becomes significantly broadened and the activity at the centre frequency becomes indistinguishable from the surrounding activity. The worst case occurs when both outer and inner hair cell loss occur together. Combination of damage to both types of hair cell compound the effects produced by both types of damage, generating neural outputs in which the tone burst frequency content becomes completely indistinguishable from the surrounding activity.

2. **Response to synthetic speech.** The response of the model to a synthetic syllable is used to show the effect of hair cell damage on the reproduction of the complex stimulus content, specifically the formant content and formant transitions. The neural output is examined for various levels of inner and outer hair cell damage both individually and combined. Increasing levels of inner hair cell damage reduces the formant definition in the regions of activity; when total inner hair cell damage occurs the formants cease to be reproduced around the affected region. Increasing levels of outer hair cell damage can be seen to cause significant broadening of the formants and can prevent the distinct reproduction of formant transitions. Complete outer hair cell loss does not however stop neural firing but leads to broad, poorly defined formant reproduction. Combined inner and outer hair cell damage generates broad formants with poor centre frequency definition and loss of transition definition. The overall effects reproduced by the model are those of a system suffering from frequency smearing.

9.2.4 The Recruitment Compensated Recruited Auditory Model

The final model presented in this work details the development of a compensation model for recruitment. The recruitment compensation model is designed to show how the unimpaired auditory model and the recruited auditory model can be used in the development of a compensation mechanism to alleviate a modelled auditory impairment. The mechanism operates to manipulate the acoustic stimuli going to the auditory model so as to cause an interaction of the stimuli and the impaired region that better approximates the response that would have occurred in the region if it were not damaged. To perform this function the compensation system has two main sections, an analysis-synthesis system and a system for introducing compensation. The analysis-synthesis system divides the stimuli into multiple band-limited signals to allow access to the various frequency bands and then recombines these signals to generate a compensated signal, free from alias, phase and amplitude distortion. Each band has its own compensation system, which introduces the necessary compressive gain into each band depending on the complete signal level and the sub-band signal level. The compensation systems are configured based on a compression function generated by comparing the neural response to rising stimulus level in the unimpaired model and the recruited model. The compensation system is therefore able to cater for the increased auditory threshold and the loss of dynamic range compression resulting from outer hair cell loss.

The results of various simulations presented in this section were chosen to examine if the compensation model could successfully alleviate the effects of recruitment in the auditory model; overcoming the increased threshold and the abnormal growth of loudness. The following responses are shown.

1. **Response to rising stimulus level.** The response of the compensated recruited auditory model to rising stimulus level facilitates examination of the loudness growth in the compensated model and evaluation of the resulting effective threshold of audibility of the compensated model. Comparing the unimpaired responses and the compensated responses shows that the compensation model restores the loudness growth in the recruited model to within normal levels and generates threshold of audibility levels very close to the unimpaired levels.

2. **Tone burst response.** Generation of the tone burst response facilitates examination of the inner hair cell activity in the model with which to evaluate how far the compensation system goes towards restoring inner hair cell activity to normal. When the compensation system is activated the tone burst response shows that the inner hair cell activity is significantly improved and well within normal limits.

3. **Compression in the compensated model.** Measurement of the dynamic range compression in the compensated model facilitates examination of how well the compensation model restores the levels of compression within the impaired model. Comparing the compression in the unimpaired system and the compensated recruited system shows that the compensation model is able to restore the dynamic range compression very close to the unimpaired levels.

4. **Response to synthetic speech.** The real test of the compensation model is to see how well it can compensate for recruitment in the auditory model when a complex signal like speech is the stimulus. Stimulating the compensated recruited model with a synthetic syllable and comparing the resulting neural activity with that of the recruited auditory model, it can be seen that the compensation model significantly enhances the formant reproduction in the neural activity and as a result assists in enhancing the definition of the formant transitions. The complex stimulus content is clearly visible in the compensated neural activity compared to that of the recruited model. However, compared to the activity in the unimpaired model, the compensation system tends to

cause increased high frequency activity when severe outer hair cell damage is present. This occurs because of an upward spread of excitation caused by the enhancements provided to compensate for lower frequency damage. This can be seen in the fact that the activity is phase locked to the low frequency signal content around the damaged sites; it can, however, be minimised through optimisation of the attack and release times used in the compensation model. Another limiting feature of the compensated model is that it focuses on generating optimum activity for primary stimulus content and as a result very low-level signal content tends to be poorly catered for; this, however, is a trade off that must be made to avoid excessive loudness growth.

9.3 Overall Conclusions

The Unimpaired Auditory Model

Examining the responses of the unimpaired auditory model, it shows good overall correlation with the known responses of the unimpaired auditory system. The model has been designed to incorporate as much of the current data on the auditory system as possible and given the model has been configured to function as closely as possible to the operations taking place in the auditory system, it is satisfying to see that the responses are so well aligned. The model is not, however, without its limitations and compromises. During the development of the model the limitations of the currently accessible technology must be accepted and certain compromises must be made. Primary among these have been

- The auditory model models the responses and interactions of the elements of the auditory system rather than the physiology of the elements. Given the uses of the model this is an acceptable compromise and directs the focus of the model on the auditory responses, which are of greater significance in this work. This primarily reduces the computational burden.

- The auditory model used to generate the responses in this work does not cater for reflections within the auditory system. An experimental reflection path was developed and the results showed that the reflections do not have a significant effect on the forward travelling waves or the inner hair cell activity; however, activation of the reflection path in the model sees a significant increase in the computational load making the model excessively unwieldy. Considering the

objectives of the current work, this is not a severe limitation but it does prevent evaluation of the various types of otoacoustic emissions generated by the auditory system which could potentially be used for configuration purposes when specifying outer hair cell damage.

Overall the unimpaired auditory model provides a realistic model of the auditory system and compares very favourably with other current models in the literature. In many cases the model in this work shows much greater correlation with the known responses of the auditory system than other published works.

The Impaired Auditory Models

Considering the impaired auditory model suffering from recruitment and/or frequency smearing, their success derives from the success of the auditory model itself.

- The introduction of outer hair cell damage into the auditory model can be shown to cause the model to suffer from the effects of recruitment and generate responses equivalent to those that arise in the recruited auditory system. Specifically, outer hair cell damage in the auditory model causes a reduction in the dynamic range compression resulting in increased thresholds and abnormally rapid growths in loudness. The current recruited auditory model conforms to all of the known data on recruitment and generates a realistic model of the auditory system suffering from recruitment.

- The introduction of inner and outer hair cell damage into the auditory model can be shown to cause the model to suffer from frequency smearing and generate responses equivalent to those that arise in the frequency smeared auditory system. Specifically, inner and outer hair cell damage causes the neural activity in the affected regions to reduce and broaden, having detrimental effects on the reproduction of important stimulus content in the neural activity going to the brain. The effects of this impairment are to reduce the ability to distinguish information content in acoustic stimuli due to the spectral content being indistinguishable in the smeared neural activity.

Overall, the impaired auditory models provide realistic models of the auditory system suffering from the particular impairments. The recruitment model in particular has seen the greatest evaluation, comparing very favourably with other forms of recruitment models and experimental results on the recruited auditory system. The frequency smeared model compares very favourably with experimental results and provides a much more realistic simulation of hearing impairment than other models.

The Compensation Model

The development of a compensation model for recruitment is mainly to explore the possibility of using auditory models, both unimpaired and impaired, to facilitate the construction of a mechanism for the alleviation of the effects of an impairment in an impaired auditory model. In the case of recruitment, the compensation model needs to have access to the effects the impairment has on the neural activity generated in auditory model, data that is readily accessible in a modelling environment. The results before and after insertion of the compensation model are quite remarkable; the compensation model can be shown to lower the effective threshold of audibility to around normal levels and allows loudness to grow within normal limits, effectively compensation for the effects of impairment within the auditory model. The success of the compensation model indicates that these types of models can prove very useful for counteracting impairments in computer models and should prove useful in the development of compensation systems to alleviate impairments in real auditory systems.

9.4 Critical Evaluation

There is no doubt that the advent of digital technology has lead to remarkable far reaching revolutions in how things are done today and has played a pivotal role in the evolution of mankind into the so called "*information age.*" At the heart of this digital technology is digital signal processing which specifies the types of processes carried out and how these processes are performed. It is only within the last decade or so that this technology has seen application in the development of hearing aid strategies for the provision of relief for the hard of hearing. The main reasons of the delay in the uptake of this technology have been the lack of developmental signal processing models coupled with the strict limitations on size and power consumption associated with hearing aids. Technological progress in ASIC design/fabrication and power supply technologies have

permitted the manufacture of hardware capable of operating within the limits imposed by hearing aids and have significantly increased the availability of these devices. However, they are very expensive (IR£3500+), as with all leading edge technologies. Aside from the physical production of the hearing aids and the internal hardware, the development of the signal processing strategies within these devices is extremely complicated and hence very costly. Considering these factors, the devices themselves have to be as generic as possible to address as wide an audience as possible in an attempt to recoup the research and development costs. This is not without its problems as it limits the possibility of developing strategies to tackle specific impairments and leads to very complex fitting strategies when it comes to optimally matching a device to an end user, strategies that typically depend on the results of subjective tests designed to characterise a subject's impairment. How effective a device is at providing relief is generally assessed using speech reception tests with a view towards providing optimum speech intelligibility and sound reception. This is all very well if the subject knows what they should be hearing and have an understanding of how normal sounds should sound. They will be able to provide suitable feedback as to the performance of a device and its configuration. However, what if the subject is a child or a person of limited/reduced mental capacity or is born with an auditory impairment? They often lack the auditory experience to be able to judge when sounds are normal and what information is being carried by the sounds. They have limited internal references with which to compare the sounds they are hearing and thus cannot provide relevant feedback to judge the performance of compensatory measures, significantly complicating the fitting process.

Digital signal processing technology has the potential to revolutionise how auditory impairment is alleviated. However, the complexity involved in its development and deployment necessitate high costs per unit, placing this technology beyond the means of many people. Consequently, the generic nature of these devices has resulted in complex fitting procedures paralleled with a retardation of research and development into the alleviation of specific impairments. In order that compensatory measures better alleviate the effects of impairment, the focus of research and development must move towards applying greater emphasis on specific impairments and the solutions required for their alleviation. However, the simple fact of the matter is that *economics* drive corporate research and development efforts. Until the core costs involved in the research and development of signal processing strategies reduces, the potential benefits of this technology in alleviating auditory impairment will remain largely untapped.

It is within this environment that one may begin to appreciate the potential value of the work presented in this thesis; when considering how best to harness the benefits of this technology, making it more *reliable*, more *effective* and more *affordable*.

The auditory models in this work provide a framework in which experimentation with the auditory system can be performed quite easily and cheaply. Any number of test signals can be presented to the auditory model both in its unimpaired and impaired states permitting objective examination of the effects an impaired auditory system has on the neural stimulation of the brain. This potentially provides researchers with an unlimited source of information on the auditory system for the development of compensatory strategies. It also makes it possible to examine much greater sets of impaired scenarios in conjunction with unlimited arrays of corresponding compensation strategies without the typically high evaluation and auditory survey costs (in both time and money) associated with subjective analysis. These types of simulations can easily emulate broad ranges of impairments and compensation strategies providing objective evaluations of vast sets of data, all of which can be automated thus reducing the time and expense of such operations. The potential to evaluate and experiment extensively with many compensatory strategies greatly increases the likelihood of finding very efficient and effective strategies for dealing with auditory impairment and at a fraction of the cost. Obviously this approach cannot totally replace subjective testing when it comes to evaluation but it would significantly reduce the quantity of unfruitful testing. It is clear that the potential for these types of models and this type of approach could be quite significant.

There is however a very important consideration to using models in this way and applying this type of approach. The models themselves, both unimpaired and impaired, must emulate the auditory system sufficiently accurately for this approach to be of any real benefit. Even under these conditions, ultimately the following questions must be faced; "*is any model capable of completely replicating the function of such a complex biological system?*" and "*can any model realistically be a valid substitute for human subjects?*" Clearly the answer must *currently* be no; neither sufficient technology nor information is available with which to perform such an enormous task. However, that is not to say the effort is futile, the known functions and interactions of the auditory system can be modelled and have been in this work. It would be unrealistic to think that the models presented in this work are perfect operational models. Apart from the deliberate compromises that have been mentioned during the development, there may be

imperfections in the current knowledge and further testing may bring undesirable interactions to light generating unrealistic responses. Although significant testing has been performed to evaluate the most significant aspects of the auditory model, limits must be drawn to the extents to which testing may be performed within the scope of this current work. However, the testing suggested for future work in Section 9.5 should bring to light any unforeseen limitations of the current model. Generally speaking though, the work in this thesis can be considered an evaluation of whether or not the unimpaired and impaired auditory system can be accurately modelled in signal processing terms (within the bounds of the current knowledge) and if within this environment it is possible to develop a compensatory mechanism to alleviate the effects of a modelled impairment? Given the models and the simulation results presented in this work, it is the opinion of the author that the answer must be an emphatic yes.

The unimpaired auditory model replicates the known responses of the auditory system evaluated against reported human cadaver experimentation, psychoacoustic experimentation, experimentation with species having similar auditory systems and the responses of accepted models in the literature. The impairment of this auditory model with hair cell damage generates responses in the auditory model consistent with the reported effects of similar damage in the impaired auditory system, generating the sensorineural impairments of recruitment and frequency smearing. Having developed models that replicate the responses of the unimpaired auditory system and the recruited auditory system, a compensation model has been developed, using information from these auditory models, which provides significant alleviation of the effects of recruitment in the impaired model. Considering the auditory models generate responses in accordance with the known responses of the auditory system, the compensation model would be deemed a suitable candidate for further evaluation into the compensation of recruitment in impaired individuals. This work not only presents a real opportunity to reduce the developmental costs of compensation strategies for digital hearing aids but also the possibility of developing better hearing aids.

The discussion so far has highlighted the possible developmental benefits of the current work to the hard of hearing; however, there are other benefits, particularly for those for whom it is particularly difficult to provide compensation. Generally speaking this work has the potential to assist in finding optimal compensation solutions which benefit all users. However, it also hopes to provide badly needed assistance for those least capable of

getting help; namely the very young and those of limited/reduced mental capacity. The principle difficulty when dealing with these people is that most of the time it is neither reasonable nor possible to perform many of the more advanced subjective acoustic tests, thus greatly limiting the available impairment information (and the ability to determine how well particular compensatory measures perform). Considering the case of recruitment, it should be possible using the distortion product test to identify the level of outer hair cell damage in a subject and configure a recruited auditory model to operate in a similar fashion to the impaired subject's auditory system. Using this impaired model in conjunction with the unimpaired auditory model, it should be possible to configure an optimal compensation system for the impairment, thus providing the best configuration for the impaired individual. Clearly this too falls under the requirements for sufficiently accurate models but it has the potential to be extremely beneficial to all sufferers.

Generally speaking, the models presented in this work and the approach for developing a recruitment compensation model, as pieces of work in themselves are of significant academic value both for teaching and research into the auditory system and its impairments. However, the more important point this work brings to light is that it is feasible to use engineering models for the development of signal-processing strategies for alleviating auditory impairment and that such models should reduce the cost of researching and developing these strategies. The net result should be more efficient and effective compensation strategies that cost significantly less than their present day counterparts.

9.5 Suggestions for Future Work

The suggestions for future work are loosely divide into three sections model testing, model enhancement/expansion and model applications.

Model Testing

All of the tests performed on the models in this work to date have been theoretical challenges of the models. What is meant by this is that the test signals have all been artificially generated and the impairments modelled have all been based on artificial impairments. Having developed an auditory model whose characteristic responses to audiometric and characteristic evaluations are closely correlated with those typical of the unimpaired human auditory system, it is necessary at this stage to apply the

characterisation methods suggested to impaired subjects, to gather the relevant live data to configure impaired models so they can emulate the auditory systems of impaired subjects. The objective of such procedures would be to facilitate comparative analysis of the responses generated by the impaired models and the impaired subjects, to determine how well the models are capable of emulating the impaired processes taking place. This would allow further evolution of the impairment models, assist in highlighting any weaknesses there may be in the models and provide a valuable tool for objectively analysing particular cases of auditory impairment. Although a relatively straightforward data gathering task it would involve gaining access to suitable equipment, suitable facilities, highly skilled audiologists and most of all cooperative subjects with the time available to spend at this endeavour. The most difficult part of these procedures would be finding subjects with sufficient available time to spend performing test-retest scenarios to compare the various responses required to validate the impaired models operation.

Having performed these configuration processes, in the case of recruitment, the compensation model should then be configured and optimised to alleviate the effects of recruitment in a recruited auditory model corresponding to an impaired subject. Considering that the recruited auditory model would be emulating a live subject's impairment, the corresponding compensation model should be of benefit in alleviating their impairment. Sets of test compensated data should therefore be generated for presentation to the subject to evaluate the capabilities of the compensation model and evaluate its ability to provide alleviation. This would provide important information necessary for validating and optimising the compensation process suggested. All of the tests involving the models would occur *offline*, as the model does not operate in real time and detaining subjects for long periods of time makes such a process very unattractive. Although it would be very time consuming to generate sets of compensated test data using the current implementation of the models, it would provide very important information. Considering the primary effects of auditory impairment is a reduction in the ability of the sufferer to comprehend the information being conveyed acoustically, it would be extremely beneficial to perform some of the standard speech recognition tests on subjects with and without the compensation in place to evaluate the real world benefit such schemes may provide.

Another set of tests that should be performed would involve characterising the compensation model to alleviate the effects introduced by recruitment models designed to

elicit in normal subjects the effects of impairment. Presenting sets of compensated test data to such recruitment models and presenting their output to unimpaired subjects would provide additional checks on the capabilities of the compensation system suggested.

A further area of study that would be very useful would be to perform an evaluation identifying how recruitment and frequency smearing affect the intelligibility of acoustic stimuli. Examinations of this type should be able to isolate at what levels these impairments become a significant problem and the models in this work would permit detailed examination of the resulting neural activity under these conditions. This would assist in isolating the features most important for comprehension and permit a more focused approach to impairment compensation taking into account the perceptual consequences of impairment.

In general the suggested future testing would involve significant endeavour and would require someone with a large degree of knowledge in the area, access to audiology clinics and have the special abilities required to cater for the needs of impaired subjects. This type of testing is the vital next step in the evolution of these models and should instil even greater confidence in the capabilities of these models.

Model Enhancement/Expansion

Testing of the models as suggested above will provide valuable information to both validate the models and optimise their operation. However, the current implementation is a very complex combination of a compiled numerical processing engine and a custom interface. It would be fair to say that the models as they stand are quite complex and require an in depth knowledge of their operation to alter their configuration. This is not too big an issue to resolve but the execution time of the models must be addressed. In general the stimulation of the auditory model with 100ms of raw data takes approximately 20 seconds to generate the complete response on a Pentium Pro 200MHz computer; although significantly lower than the several minutes required for pure Matlab code, it is still too long for manipulation of real speech which may be several seconds in length and thus require significant processing time. In order to remedy this, the rate at which the numerical operations can be performed must be accelerated using additional hardware. There are two possible methods of doing this; the more traditional approach would be to port the numerical engine to a DSP processing board, which typically would have much higher

numerical operation throughput than standard digital computers. A more novel approach would be to write a network operable version of the numeric processing engine using the distributed common object model (DCOM) paradigm permitting distribution of the processing load over several networked machines. Considering it is the numerical processing taking place in each of the inner ear section that occupies most of the processor time, assigning several sections per network machine would dramatically increase the throughput. Preliminary load balancing tests using a network of Pentium II 300MHz machines has shown that 20 machines would be quite capable of operating the distributed model in real time. With the recent advent of Pentium III 1GHz processors, better than real time operation could be achieved with 4-5 machines. Although a novel approach, the parallelisation of the model would be the optimum progression path as a single DSP processor, although potentially capable of operating the model in real time, would leave little room for expansion of the model given the current technology. Expansion of the model would require the cooperation of several DSP processors operating in parallel, which would lend itself very neatly to porting of the distributed model to a parallel DSP processor bank.

The development of faster variants of the models would allow further development and expansion of the auditory models. In particular the activation of the reflection pathway would be possible allowing the generation and examination of the various forms of otoacoustic emissions generated by the auditory system. This in turn would allow expansion of the DPOAE characterisation method for outer hair cell damage determination. Having determined the level of damage and implementing this in the impaired auditory model, the measured DPOAEs from the model and the subject should be the same. This would provide yet another test of the model's capabilities. Greater processing capabilities would also mean that the responses of the middle ear could be made dynamic, permitting active manipulation of the middle ear model's responses, more in line with the active acoustic reflex in the auditory system.

In general, enhancing the operational speed of the models would make it much easier to generate test signals and reduce the dead time experienced by users while waiting for results. In terms of the testing procedures mentioned above, a real time model would allow the tests take place online permitting greater flexibility in the testing procedure as test signals could be generated to suit each situation as it developed. Expanding the model to include a reflection path would enhance the realism of the model and also facilitate

experimentation with otoacoustic emissions. This would be extremely beneficial given the current interest in using the various emissions of the auditory system as a diagnostic tool for auditory impairment. The inclusion of an active acoustic reflex in the model would further enhance its response particularly when dealing with real stimuli having large randomly fluctuating energy content, varying across the entire dynamic range.

Model Applications

Academically speaking the models are gratifying mathematical representation of the processes taking place in the auditory system and as such will provide a valuable educational tool and source of information. However, their real benefit would be in the development and evaluation of hearing aid strategies for the alleviation of real impairments. This work shows that it is theoretically possible, using the current models, to provide alleviation of artificial impairment. Subject to further testing, the compensatory model should be tested with impaired subjects to evaluate its performance and potential use as a hearing aid strategy for alleviating recruitment in live subjects. The development of a hearing aid strategy based on the proposed approaches would be a real justification of the practicality of this work.

Given the complexity involved in developing hearing aid strategies, the auditory models themselves could be used in the evaluation of other test compensatory strategies. The models facilitate measurement of all the responses of the auditory system permitting an objective evaluation of how near different strategies approach their desired goal, thus providing a real benefit to their designers.

Finally, considering the auditory model generates the neural data going to the brain, the coding of the acoustic stimuli into neural activity will incorporate all the specific features of the stimulus, including the unique voice patterns generated by human speakers. The auditory model therefore could provide a very high-resolution front end for automatic voice recognition systems, a much sought after goal in the computer science community.

In general, the models presented in this work are primarily focused on the auditory system and its impairment with a view towards encapsulating the essence of this very complex system in an engineering model, permitting objective analysis of this system and its impairments. The title of this work promotes an examine into the *application of digital*

signal processing to the modelling and alleviation of auditory impairment and has shown that digital signal processing is very suitable for modelling the auditory system and its impairments and has the potential to provide significant assistance in the development of compensatory strategies for the alleviation of auditory impairment.

References

[1] Gearge Taylor and Juliet Bishop, *Being Deaf: The Experience of Deafness*, Pinter, London, 1991.

[2] Ronald A. van Buuren, Joost M. Festen and Tammo Houtgast, "Compression and expansion of the temporal envelope: Evaluation of speech intelligibility and sound quality," *Journal of the Acoustical Society of America*, vol. 105, no. 5, May 1999, pp. 2903-2913.

[3] Brian C.J. Moore, Robert W. Peters and Michael A. Stone, "Benefits of linear amplification and multichannel compression for speech comprehension in backgrounds with spectral and temporal dips," *Journal of the Acoustical Society of America*, vol. 105, no. 1, January 1999, pp. 400-411.

[4] Laura A. Drake, Janet C. Rutledge and Jonathan Cohen, "Wavelet Analysis in Recruitment of Loudness Compensation," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, December 1993, pp. 3306-3312.

[5] Todd Schneider and Robert Brennan, "A multichannel compression strategy for a digital hearing aid," *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 1, 1997, pp. 411-414.

[6] Brian Strope and Abeer Alwan, "A novel structure to compensate for frequency-dependent loudness recruitment of sensorineural hearing loss," *International Conference on Acoustics, Speech and Signal Processing, Conference Proceedings IEEE*, vol. 5, 1995, pp.3539-3542.

[7] Janet C. Rutledge and Mark A. Clements, "Compensation for Recruitment of Loudness in Sensorineural Hearing Impairments using a Sinusoidal Model of Speech," *International Conference on Acoustics, Speech and Signal Processing IEEE*, vol. 5, 1991, pp. 3641-3644.

[8] Harvey Dillon, "Tutorial Compression? Yes. But for Low or High Frequencies. For Low or High Intensities, and with What Response Times?," *Ear & Hearing*, by Williams & Wilkins, vol. 17, no. 4, 1996, pp.287-307.

[9] Siemens Audiology Group, "What is the benefit of digital hearing instruments?" http://sat.eskatoo.de/cgi-bin/faq.pl?e,pop_gefr,6, 1998.

[10] Manfred R. Schroeder, "Models of Hearing", *Proceedings of the IEEE*, vol. 63, no. 9, September 1975, pp. 1332-1350.

[11] Jont B. Allen, "A hair-cell model of neural response," in *Mechanics of Hearing*, E. DeBoer and M.A. Viergever, Eds. Martinus Nijhoff, 1983.

[12] H. Davis, "A mechanoelectrical theory of cochlear action," in *Annals of Otorhinolaryngology*, vol. 67, 1956, pp. 789-801.

[13] Walt Jesteadt, "Modelling Sensorineural Hearing Loss," Ed. Walt Jesteadt, 1997, pp. 1-4.

[13.1] C.A.P. Rogers, "Pinna Transformations and Sound Reproductions," Journal of the Audio Engineering Society, vol. 29, 1981, pp. 226-234.

[14] G. Von Békèsy, "Experiments in Hearing," McGraw Hill, 1960.

[15] Maryanne Tate, *Principles of Hearing Aid Audiology*, Chapman & Hall, 1994.

[16] S.T. Neely and J.B. Allen, "Relation between the rate of growth of loudness and intensity difference limen," in Walt Jesteadt (Ed), *Modelling Sensorineural Hearing Loss*, Lawrence Erlbaum Associates, 1997, pp. 213-222.

[17] Christian Giguère, Philip C. Woodland , "A computational model of the auditory periphery for speech and hearing research. II. Descending paths," *Journal of the Acoustical Society of America*, vol. 95, no. 1, 1994, pp. 331-342.

[18] John A. Seikel, Douglas W. King, David G. Drumright, *Anatomy and Physiology for Speech, Language, and Hearing*, (2nd Ed.), Singular Publishing Group, Inc., 1997.

[19] George T. Mencher, Sanford E. Gerber and Andrew McCombe, *Audiology and Auditory Dysfunction*, Allyn & Bacon, 1996.

[20] J.B. Allen and S.T. Neely, "Micromechanical models of the cochlea," *Physics Today*, July 1992, pp. 40-47.

[21] Egbert de Boer, "A method for forward and inverse solutions of a three-dimensional model of the cochlea," *Journal of the Acoustical Society of America*, vol. 103, no. 6, June 1998, pp. 3725-3728.

[22] Egbert de Boer and Alfred L. Nuttall, "The "inverse problem" solved for a three-dimensional model of the cochlea. III. Brushing-up the solution method," *Journal of the Acoustical Society of America*, vol. 105, no. 6, June 1999, pp. 3410-3420.

[23] E. deBoer, "Auditory physics. Physical principles in hearing theory. III," *Physics Reports*, Netherlands, vol. 203, no. 3, May 1991, pp. 125-231.

[24] E. deBoer, "Mechanics of the cochlea: modelling efforts, " in *The Cochlea*, edited by P. Dallas, A. N. Popper, and R.R. Fay, Springer-Verlag, New York, 1996, pp. 258-317.

[25] E. deBoer, "Cochlear models and minimum phase," *Journal of the Acoustical Society of America*, vol. 102, no. 6, December 1997, pp. 3810-3813.

[26] E. deBoer, "Classical and non-classical models of the cochlea," *Journal of the Acoustical Society of America*, vol. 101, no. 4, April 1997, pp. 2148-2150.

[27] J.B. Allen and M. M. Sondi, "Cochlear macromechanics - time-domain solutions," *Journal of the Acoustical Society of America*, vol. 66, 1979, pp. 123-132.

[28] M. A. Viergever, "Cochlear macromechanics-a review," in *Peripheral Auditory Mechanics*, edited by J.B. Allen, J.L. Hall, A. Hubbard, S.T. Neely and A. Tubis, Springer-Verlang, Berlin, 1986, pp. 63-72.

[29] E. deBoer and A.L Nuttall, "On cochlear cross-correlation functions: connecting nonlinearity and 'activity'," in *Diversity in Auditory Mechanisms*, edited by E.R. Lewis, G.R. Long, R.F. Lyon, P.M. Narins, C.R. Steele and E. Hecht-Poinar, World Scientific Singapore, 1997, pp. 291-297.

[30] A.E. Hubbard, "A travelling wave-amplifier model of the cochlea," *Science*, vol. 259, 1993, pp. 68-71.

[31] C.R. Steele, G. Baker, Jr. Tolomeo and D. Zetes, "Electromechanical models of the outer hair cell," in *Biophysics of Hair-Cell Sensory Systems*, edited by H. Duifhuis, J.W. Horst, P. Van Dijk and S.M. Van Netten, World Scientific, Singapore, 1993, pp. 207-214.

[32] C.D. Geisler and C. Sang, "A cochlear model using feed-forward outer-hair-cell forces," *Hearing Research*, vol. 86, 1995, pp. 132-146.

[33] M.H. Holmes and J.D. Cole, "Cochlear Mechanics: Analysis for a pure tone," *Journal of the Acoustical Society of America*, vol. 76, no. 3, September 1984, pp. 767-778.

[34] A. Recio, S.S. Narayan and M.A. Ruggero, "Wiener-kernel analysis of basilar-membrane responses to white noise," in *Diversity in Auditory Mechanisms*, edited by E.R. Lewis, G.R. Long, R.F. Lyon, P.M. Narins, C.R. Steele and E. Hecht-Poinar, World Scientific Singapore, 1997, pp. 325-331.

[35] Mark. H. Holmes, "Frequency discrimination in the mammalian cochlea: Theory versus experiment," *Journal of the Acoustical Society of America*, vol. 81, no. 1, January 1987, pp. 103-114.

[36] J.B. Allen, "Cochlear micromechanics – A physical model of transduction," *Journal of the Acoustical Society of America*, vol. 68, pp. 1660-1670, 1980.

[37] D.O. Kim, S.T. Neely, C.E. Molnar, J.W. Matthews, "An active cochlear model with negative damping in the cochlear partition. Comparison with Rhode's ante- and post-mortem results," in *Psychological, Physiological and Behavioural Studies in Hearing*, edited by G. van den Brink, F.A. Bilsen, Delft U. P., Delft, The Netherlands, 1980, pp. 7-14.

[38] G. Zweig, "Finding the impedance of the organ of Corti," *Journal of the Acoustical Society of America*, vol. 89, no. 3, March 1991, pp. 1229-1254.

[39] S.T. Neely and D.O. Kim, "A model for active elements in cochlear biomechanics," *Journal of the Acoustical Society of America*, vol. 79, no. 5, May 1986, pp. 1472-1480.

[40] S.T. Neely, "A model of cochlear mechanics with outer hair cell motility," *Journal of the Acoustical Society of America*, vol. 94, no. 1, July 1993, pp. 137-146.

[41] J.B. Allen and P.F. Fahey, "Using distortion products to measure the cochlear amplifier gain on the basilar membrane," *Journal of the Acoustical Society of America*," vol. 92, no. 1, July 1992, pp. 178-188.

[42] Christian Giguère and Philip C. Woodland, "A computational model of the auditory periphery for speech and hearing research. I. Ascending paths," *Journal of the Acoustical Society of America*, vol. 95, no. 1, January 1994, pp. 343-349.

[43] P.A.Garbes and Jont B. Allen, "Deriving human neural tuning curves from a mechanical model of the cochlea," *Journal of the Acoustic Society of America*, vol. 95 no. 5, May 1994, pp. 3006.

[44] L. Deng and C.D. Geisler, "A composite auditory model for processing speech sounds," *Journal of the Acoustical Society of America*, vol. 82, no. 6, 1987, pp. 2001-2012.

[45] E. Zwicker, "A hardware cochlear nonlinear preprocessing model with active feedback," *Journal of the Acoustical Society of America*, vol. 80, no. 1, 1986, pp. 146-153.

[46] L.C. Peterson and B.P. Bogert, "A dynamical theory of the cochlea," *Journal of the Acoustical Society of America*, vol. 22, no. 3, 1950, pp. 369-381.

[47] S.T. Neely, "Mathematical modeling of cochlear mechanics," *Journal of the Acoustical Society of America*, vol. 78, no. 1, pt. 2, July 1985, pp. 345-352.

[48] C.D. Geisler, "A cochlear model using feedback from motile outer hair cells," *Hearing Research*, vol. 54, no. 1, 1991, pp. 105-117.

[49] J.M. Kates, "Accurate tuning curves in a cochlear model," *IEEE Transactions on Speech and Audio Processing*, vol.1, no. 4, October 1993, pp. 453-462.

[50] M. R. Schroeder, "An integrable model for the basilar membrane," *Journal of the Acoustical Society of America*, vol. 53, no. 2, February 1973, pp. 429-434.

[51] M. Holmes and J.D. Cole, "Pseudoresonance in the cochlea," in E. deBoer and M.A. Viergever, (Eds.), *Mechanics of Hearing*, Hague, The Netherlands: Martinus Nijoff, 1983.

[52] J.M. Kates, "A Time-Domain Digital Cochlear Model," *IEEE Transactions on Signal Processing*, vol. 39, no. 12, December 1991, pp. 2573-2592.

[53] K. Wang, S. Shamma, "Self-Normailzation and Noise-Robustness in Early Auditory Representations," *IEEE Transactions on Speech and Audio Processing,* vol. 2, no. 3, 1994, pp. 421-435.

[54] M. Pflueger, R. Hoeldrich and W. Riedler, "A nonlinear model of the peripheral auditory system," *Audio Engineering Society*, 103rd Convention 1997.

[55] R. Patterson and M. Allerhand, "Time-domain modelling of peripheral auditory processing: A modular architecture and a software platform," *Journal of the Acoustical Society of America*, vol.98, no. 4, October 1995, pp. 1890-1894.

[56] Sunil Puria and Jont B. Allen, "Measurements and model of the cat middle ear: Evidence of tympanic membrane acoustic delay," *Journal of the Acoustical Society of America*, vol. 104, no. 6, December 1998, pp. 3463-3481.

[57] J.J. Zwislocki, "Analysis of some auditory characteristics," in *Handbook of Mathematical Physiology*, edited by R.D. Luce, R.R. Bush and E. Galanter (Wiley, New York), vol. III, 1965, pp. 1-97.

[58] M.B. Gardner and M.S. Hawley, "Network representations of the external ear," *Journal of the Acoustical Society of America*, vol. 52, 1972, pp. 1620-1628.

[59] Jérôme Pascal, Antoine Bourgeade, Michel Lagier, Claude Legros, "Linear and nonlinear models of the human middle ear," *Journal of the Acoustical Society of America*, vol. 104, no. 3, September 1998, pp. 1509-1516.

[60] W. Rosenblith, "Sensory performance of organisms," *Reviews of Modern Physics*, vol. 31, 1959, pp. 485-491.

[61] J.B. Allen, "Derecruitment by multiband compression in hearing aids," in Walt Jesteadt (Ed.), *Modelling Sensorineural Hearing Loss*, Lawrence Erlbaum Associates, 1997, pp. 99-112.

[62] W.F. Carver, "Loudness balance procedures," *Handbook of Clinical Audiology*, J. Katz ed., Baltimore: Williams & Wilkins, 2nd edition, 1978, pp. 164-178.

[63] M. Liberman, "Single-neuron labelling and chronic cochlear pathology III: Stereocilia damage and alterations of tuning curve thresholds," *Hearing Research*, vol. 16, 1984, pp. 55-74.

[64] M. Liberman and N. Kiang, "Acoustic trauma in cats," *Acta Otolaryngologica Supplement*, 358, 1978, pp. 1-63.

[65] J.B. Allen and P.F. Fahey, "Using acoustic distortion products to measure the cochlear amplifier gain on the basilar membrane," *Journal of the Acoustical Society of America*, vol. 1, 1992, pp. 178-188.

[66] J. Pickles, *An introduction to the physiology of hearing*, London: Academic Press, 1982.

[67] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and models*, New York: Springer-Verlag, 1990.

[68] E. Zwicker and B. Scharf," A model of loudness summation," *Psychological Review*, vol. 72, 1965, pp. 3-26.

[69] Mary Florentine and E. Zwicker, "A model of loudness summation applied to noise induced hearing loss," *Hearing Research*, vol. 1, 1979, pp. 121-132.

[70] Mary Florentine, Søren Buus and Rhona P. Hellman, "A Model of Loudness Summation Applied to High-Frequency Hearing Loss," in Walt Jesteadt (Ed.), *Modelling Sensorineural Hearing Loss*, Lawrence Erlbaum Associates, 1997, pp. 187-198.

[71] L.E. Humes and Walt Jesteadt, "Models of effects of threshold on loudness growth and summation," *Journal of the Acoustical Society of America*, vol. 90, 1991, pp. 1933-1943.

[72] B.C.J. Moore, *Perceptual consequences of cochlea damage*, London: Oxford University Press, 1995.

[73] Stefan Launer, Volker Hohmann and Birger Kollmeier," Modelling Loudness Growth and Loudness Summation in Hearing-Impaired Listeners," in Walt Jesteadt (Ed.), *Modelling Sensorineural Hearing Loss*, Lawrence Erlbaum Associates, 1997, pp. 175-185.

[74] Eric Javel, "Cochlear Excitation Patterns in Sensorineural Hearing Loss," in Walt Jesteadt (Ed.), *Modelling Sensorineural Hearing Loss*, Lawrence Erlbaum Associates, 1997, pp. 9-34.

[75] P.A. Leake and G.T. Hradek, "Cochlear pathology of long term neomycin induced deafness in cats," *Hearing Research*, vol. 33, 1988, pp. 11-34.

[76] H. Fletcher and W. Munson, "Loudness, its definition, measurement, and calculation," *Journal of the Acoustical Society of America*, vol. 5, 1933, pp. 82-108.

[77] Jont B. Allen and Stephen T. Neely, "Modelling the relation between the intensity just-noticeable difference and loudness for pure tones and wideband noise," *Journal of the Acoustical Society of America*, vol. 102, no. 6, December 1997, pp. 3628-3646.

[78] D.B. O'Reilly and J.B. Foley, "Modelling the Human Auditory System suffering from Recruitment," *Irish Signals and Systems Conference*, June 1998, pp. 409-416.

[79] Tai Lin and Julius L. Goldstein, "Implementation of the MBPNL Nonlinear Cochlear I/O Model in the C Programming Language, and Applications for Modelling Impaired Auditory Function," in Walt Jesteadt (Ed.), *Modelling Sensorineural Hearing Loss*, Lawrence Erlbaum Associates, 1997, pp. 67-78.

[80] Julius L. Goldstein, "Relations among compression, suppression, and combination tones in mechanical responses of the basilar membrane: Data and MBPNL model," *Hearing Research*, vol. 89, 1995, pp. 52-68.

[81] Brian C.J. Moore, Brian R. Glasberg and Andrew Simpson, "Evaluation of a method of simulating reduced frequency selectivity," *Journal of the Acoustical Society of America*, vol. 91, no. 6, June 1992, pp. 3402–3423.

[82] M.R. Leek, M.F. Dorman and Q. Summerfield, "Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners," *Journal of the Acoustical Society of America*, vol. 81, pp. 148-154, 1987.

[83] David S. Lum and Louis D. Braida, "DSP Implementation of a Real-Time Hearing Loss Simulator Based on Dynamic Expansion," in Walt Jesteadt (Ed.), *Modelling Sensorineural Hearing Loss*, Lawrence Erlbaum Associates, 1997, pp. 113-130.

[84] P. Duchnowski and P.M. Zurek, "Villchur revisited: Another look at agc simulation of recruiting hearing loss," *Journal of the Acoustical Society of America*, vol. 98, no. 6, December 1995, pp. 3170-3181.

[85] Yoshito Nejime and Brian C.J. Moore, "Simulation of the effect of threshold elevation and loudness recruitment combined with reduced frequency selectivity on the intelligibility of speech in noise," *Journal of the Acoustical Society of America*, vol. 102, no. 1, July 1997, pp. 603-615.

[86] E. William Yund and Thomas R. Crain, "Voiced Stop Consonant Discrimination with Multichannel Expansion Hearing Loss Simulations," in Walt Jesteadt (Ed.), *Modelling Sensorineural Hearing Loss*, Lawrence Erlbaum Associates, 1997, pp. 149-167.

[87] B.C.J Moore, B.R. Glasberg and D.A. Vickers, "Simulation of the effects of loudness recruitment on the intelligibility of speech in noise," *Br. J. Audiol.* vol. 29, 1995, pp. 131-143.

[88] B.C.J Moore and B.R. Glasberg, "Simulation of the effects of loudness recruitment and threshold elevation on the intelligibility of speech in quiet and in a background of speech," *Journal of the Acoustical Society of America*, vol. 94, 1993, pp. 2050-2062.

[89] E. Villchur, "Simulation of the effect of recruitment on loudness relationships in speech," *Journal of the Acoustical Society of America*, vol. 56, 1974, pp. 1601-1611.

[90] Steven V. De Gennaro and Louis D. Braida, "Effects of Linear Amplification and Amplitude Compression on Sentence Reception by Listeners With Hearing Loss Simulated by Masking Noise," in Walt Jesteadt (Ed.), *Modelling Sensorineural Hearing Loss*, Lawrence Erlbaum Associates, 1997, pp. 131-148.

[91] D.A. Fabry and D. Van Tasell, "Masked and filtered simulation of hearing loss: Effects on consonant recognition," *Journal of Speech and Hearing Research*, vol. 29, 1986, pp. 170-178.

[92] L.E. Humes, D.D. Dirks, T.S. Bell and G.E. Kincaid, "Recognition of nonsense syllables by hearing-impaired listeners and by noise-masked normal hearers," *Journal of the Acoustical Society of America*, vol. 81, 1987, pp. 765-773.

[93] P.M. Zurek and L.A. Delhorne, "Consonant reception in noise by listeners with mild and moderate sensorineural hearing impairment," *Journal of the Acoustical Society of America*, vol. 82, 1987, pp. 1548-1559.

[94] J.P. Gagné and N.P. Erber, "Simulation of sensorineural hearing impairment," *Ear and Hearing*, vol. 8, 1987, pp. 232-243.

[95] E. Villchur, "Electronic models to simulate the effect of sensory distortions on speech perception by the deaf," *Journal of the Acoustical Society of America*, vol. 62, 1977, pp. 665-674.

[96] R.D. Celmer and G.R. Bienvenue, "Critical bands in the perception of speech signals by normal and sensorineural hearing loss listeners." in *The Psychophysics of Speech Perception*, edited by M. E. H. Schouten (Nijhoff, Dordrecht, The Netherlands), 1987.

[97] M. ter Keurs, J.M. Festen and R. Plomp, "Effect of spectral envelope smearing on speech reception. I," *Journal of the Acoustical Society of America*, vol. 91, 1992, pp. 2872-2880.

[98] T. Baer and B.C.J. Moore, "Effects of spectral smearing on the intelligibility of sentences in the presence of noise," *Journal of the Acoustical Society of America*, vol. 94, 1993, pp. 1229-1241.

[99] I.R. Summers, "Electronically simulated hearing loss and the perception of degraded speech," in *Bioinstrumentation and Biosensors*, edited by D.L. Wise (Marcel Dekker, New York), 1991.

[100] I.R. Summers and A.D. AI-Dabbagh, "Simulated loss of frequency selectivity and it's effects on speech reception," *Acoustical Letters*, vol. 5, 1982, pp. 129-132.

[101] T. Baer and B.C.J. Moore, "Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech," *Journal of the Acoustical Society of America*, vol. 95, 1994, pp. 2277-2280.

[102] Arthur Boothroyd, Bethany Mulhearn, Juan Gong and Jodi Ostroff, "Simulation of Sensorineural Hearing Loss: Reducing Frequency Resolution by Uniform Spectral Smearing," in Walt Jesteadt (Ed.), *Modelling Sensorineural Hearing Loss*, Lawrence Erlbaum Associates, 1997, pp. 313-327.

[103] P.A. Howard-Jones and I.R. Summers, "Temporal features in spectrally degraded speech," *Acoustical Letters*, vol. 15, 1992, pp. 159-163.

[104] Denis Byrne and Harvey Dillon, "The National Acoustics Laboratories' (NAL) New Procedure for Selecting the Gain and Frequency Response of a Hearing Aid," *Ear and Hearing*, Williams & Wilkins Co., vol. 7, no. 4, 1986, pp. 257-265.

[105] Volker Hohmann and Birger Kollmeier, "Digital Hearing Aid Techniques employing a Loudness Model for Recruitment Compensation," *International Conference on Spoken Language Processing*, 1994, pp. 2111-2114.

[106] A. Kaelin, A. Lindgren and S. Wyrsch, "A digital frequency-domain implementation of a very high gain hearing aid with compensation for recruitment of loudness and acoustic echo cancellation," *Elsevier Signal Processing*, vol. 64, no. 1, 1998, pp. 71-85.

[107] S. Wyrsch and A. Kaelin, "Subband Signal Processing for Hearing Aids," *Proceedings of the IEEE International Symposium on Circuits and Systems*, vol. 3, 1999, pp. 29-32.

[108] S. Wyrsch and A. Kaelin, "Adaptive cancelling in subbands for hearing aids," *IEEE Conference on Acoustic, Speech and Signal Processing*, vol. 2, 1999, pp 921-924.

[109] J.B. Allen, "Recruitment Compensation as a Hearing Aid Signal Processing Strategy," *Proceedings of the IEEE International Symposium on Circuits and Systems*, 1998, pp. 565-568.

[110] A.T. Schneider, *An Adaptive Digital Dynamic Range Controller*, Masters thesis, University of Waterloo, Dept. of Electrical Engineering, Oct. 1991.

[111] A. Ramasubramanian, K.L. Payton and A.H. Costa, "A comparison between Conventional and Wavelet based Amplitude Compression," *Proceedings of the IEEE-SP International Symposium on Time-frequency and Time-scale Analysis*, 1998, pp.161-164.

[112] M.C. Killion, "Hearing aids: Past, present and future: Moving toward normal conversation in noise," *Br. J. Audiol.*, vol. 31, 1997, pp. 141-148.

[113] R. Meddis, "Computer models of the auditory periphery," *Journal of the Acoustical Society of America*, vol. 105, no. 2, Pt. 2, February 1999, pp. 963.

[114] R. Meddis, "Simulation of auditory-neural transduction: Further studies," *Journal of the Acoustical Society of America*, vol. 83, no. 3, 1988, pp. 1056-1063.
[115] *Hearing Impairment, Auditory perception and Language disability*, J. Bamford, E. Saunders, Eds. 2nd Edition, Whurr Publishers Ltd., 1991.

[116] A. Rimell, *Reduction of loudspeaker polar response aberrations through the application of psychoacoustic error concealment*, PhD thesis, Centre for Audio Research and Engineering, University of Essex, Colchester, UK, 1996.

[117] Y. Suzuki, T. Sone, "Frequency characteristics of loudness perception: Principles and Applications," in *Sixth Oldenburg Symposium on Psychological Acoustics*, 1993, pp. 193-221.

[118] T.J. Lynch, V. Nedzelnitsky, W.T. Peake, "Input impedance of the cochlea in the cat," *Journal of the Acoustical Society of America*, vol. 72, no. 1, 1982, pp. 108-130.

[119] W.S. Rhode, "Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique," *Journal of the Acoustical Society of America*, vol. 49, no. 4, 1971, pp. 1218-1231.

[120] J. B. Allen, "Cochlear micormechanics – A physical model of transduction," *Journal of the Acoustical Society of America*, vol. 68, no. 6, 1980, pp. 1660-1670.

[121] M. C. Liberman, "The cochlear frequency map for the cat: Labeling auditory nerve fibers of known characteristic frequency," *Journal of the Acoustical Society of America*, vol. 72, 1982, pp. 1441-1449.

[122] D. D. Greenwood, "A cochlear frequency-position function for several species – 29 years later," *Journal of the Acoustical Society of America*, vol. 87, no. 6, 1990, pp. 2592-2605.

[123] R.F. Lyon and C. Mead, "An analog electronic cochlea," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, no. 7, 1988, pp. 1119-1134.

[124] Carrick L. Talmadge, Arnold Tubis, Glenis R. Long and Pawel Piskorski, "Modeling otoacoustic emission and hearing threshold fine structures," *Journal of the Acoustical Society of America*, vol. 104, no. 3, 1998, pp. 1517-1543.

[125] J. Zwislocki, "Theory of the acoustical action of the cochlea," *Journal of the Acoustical Society of America*, vol. 22, no. 6, 1950, pp. 778-784.

[126] U. Eysholdt and V. Mellert, "A simulation of the motion of the basilar membrane using digital filters," *Journal of the Acoustical Society of America*, vol. 33, 1975, pp. 116-121.

[127] E. Ambikairajah, N.D. Black and R.Linggard, "Digital filter simulation of the basilar membrane," *Comp. Speech and Lang.*, vol. 3, 1989, pp. 105-118.

[128] N.Y. Kiang and E.C. Moxon, "Tails of tuning curves of auditory-nerve fibers," *Journal of the Acoustical Society of America*, vol. 55, no. 3, 1974, pp. 620-630.

[129] A.R. Moller, "Frequency selectivity of single auditory-nerve fibers in response to broad-band noise stimuli," *Journal of the Acoustical Society of America*, vol. 62, no. 1, 1977, pp. 135-142.

[130] L. Robles, M.A. Ruggero, N.O. Rich, "Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input-output functions, tuning curves, and response phases," *Journal of the Acoustical Society of America*, vol. 80, no. 5, 1986, pp. 1364-1374.

[131] E. Javel, "Cochlear Excitation Patterns in Sensorineural Hearing Loss," in Walt Jesteadt (Ed.), *Modelling Sensorineural Hearing Loss*, Lawrence Erlbaum Associates, 1997, pp. 9-34.

[132] L.A. Westerman, R.L. Smith, "Conservation of adapting components in auditory-nerve responses," *Journal of the Acoustical Society of America*, vol. 81, no. 3, 1987, pp. 680-691.

[133] R. Meddis, "Simulation of Mechanical to Neural Transduction in the Auditory Receptor," *Journal of the Acoustical Society of America*, vol. 79, 1986, pp. 702-711.

[134] L.A. Westerman, *Adaptation and Recovery of Auditory-Nerve Responses*, Ph. D. thesis, Syracuse University, Syracuse, NY.

[135] "Acoustics – Normal Equal-Loudness Level Contours," in *ISO 226 (E)*, 1987, pp. 20-27.

[136] D.A. Nelson, T.W. Fortune, "High-level psychophysical tuning curves: Simultaneous masking by pure tones and 100-Hz-wide noise bands," *Journal of Speech and Hearing Research*, vol. 34, 1991, pp. 360-373.

[137] Bertrand Delgutte, "Representation of speech-like sounds in the discharge patterns of auditory fibers," *Journal of the Acoustical Society of America*, vol. 68, no. 3, September 1980, pp. 843-857.

[138] C.E. Molnar, P.R. Pfeiffer, "Interpretation of spontaneous spike discharge patterns of neurons in the cochlear nucleus," *Proceedings of the IEEE*, vol. 56, 1968, pp. 993-1004.

[139] L.A. Westerman, R.L. Smith, "Rapid and Short Term Adaptation in Auditory-Nerve Responses," *Hearing Research*, vol. 15, 1984, pp 249-260.

[140] P.M. Sellick, R. Patuzzi, B.M. Johnstone, "Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique," *Journal of the Acoustical Society of America*, vol. 72, no. 1, 1982, pp. 131-141.

[141] R.P. Hellman and J. Zwislocki, "Monaural loudness function of a 1000-cps tone and interaural summation," *Journal of the Acoustical Society of America*, vol. 35, 1963, pp. 856-865.

[142] B.M. Johnstone, R. Patuzzi, G.K. Yates, "Basilar membrane measurements and the travelling wave," *Hearing Research*, vol. 22, 1986, pp. 147-153.

[143] R.D. Patterson, "The sound of a sinusoid: Spectral models," *Journal of the Acoustical Society of America*, vol. 96, no. 3, September 1994, pp. 1409-1418.

[144] R.D. Patterson, "The sound of a sinusoid: Time-interval models," *Journal of the Acoustical Society of America*, vol. 93, no. 3, September 1994, pp. 1419-1428.

[145] D.H. Klatt, "Software for a cascade/parallel formant synthesizer," *Journal of the Acoustical Society of America*, vol. 67, no. 2, 1983, pp. 502-517.

[146] S.A. Shamma, "Speech processing in the auditory system I: The representation of speech sounds in the responses of the auditory nerve," *Journal of the Acoustical Society of America*, vol. 78, no. 5, 1985, pp. 1612-1621.

[147] J.B. Allen, " OHCs shift the excitation pattern via BM tension," in E.R. Lewis, G.R. Long, R.F. Lyon, P.M. Narins, C.R. Steel (Eds.), *Diversity in Auditory Mechanics*, Singapore: World Scientific Press.

[148] P.F. Fahey and J.B. Allen, "Measurement of distortion product phase in the ear cannel of the cat," *Journal of the Acoustical Society of America*, vol. 102, no. 5, 1997, pp. 2880-2891.

[149] E. Zwicker and F.P. Harris, "Psychoacoustical and ear canal cancellation of $(2f_{-1} - f_2)$ distortion products," *Journal of the Acoustical Society of America*, vol. 87, no. 6, 1990, pp. 2583-2591.

[150] M.P. Gorga, S.T. Neely, B.M. Bergman, K.L. Beauchaine, J.R. Kaminski and Z. Liu, "Towards understanding the limits of distortion product otoacoustic emission measurements," *Journal of the Acoustical Society of America*, vol. 96, no. 3, September 1994, pp. 1494-1500.

[151] P. Trautwein, P. Hofstetter, J. Wang, R. Salvi and A. Nostrant, "Selective inner hair cell loss does not alter distortion product otoacoustic emissions," *Hearing Research*, Elsevier Science, vol. 96, 1986, pp. 71-82.

[152] R.P. Hellman, "Growth of loudness in sensorineural impairment: Experimental results and modeling implications," in Walt Jesteadt (Ed.), *Modelling Sensorineural Hearing Loss*, Lawrence Erlbaum Associates, 1997, pp. 199-212.

[153] S. Launer, V. Hohmann, B. Kollmeier, "Modeling loudness growth and loudness summation in hearing-impaired listeners," in Walt Jesteadt (Ed.), *Modelling Sensorineural Hearing Loss*, Lawrence Erlbaum Associates, 1997, pp. 175-185.

[154] E. Zwicker, H. Fastl, *Psychoacoustics – facts and models*, Berlin: Springer.

[155] E. Zwicker, B. Scharf, "A model of loudness summation," *Psychological Review*, vol. 72, 1965, pp. 3-26.

[156] M.C. Killion, "The K-Amp Hearing Aid: An attempt to present high fidelity for persons with impaired hearing," *American Journal of Audiology*, 1993, pp. 52-74.

[157] P.P. Vaidyanathan, *Multirate Systems and Filter Banks*, Brendan M. Stewart, Ed., P T R Prentice Hall Inc., 1993, pp 188.

[158] H.S. Malvar, "Extended lapped transforms: fast algorithms and applications," *Proc. IEEE Int. Conf. On Acoustics, Speech and Signal Processing*, May 1991, pp. 1797-1800.

[159] T.A. Ramstad, "Cosine modulated analysis-synthesis filter bank with critical sampling and perfect reconstruction," *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, May 1991, pp. 1789-1792.

[160] R.D. Koilpillai, P.P. Vaidyanathan, "Cosine-modulated FIR filter banks satisfying perfect reconstruction," *IEEE Transactions on Signal Processing*, vol. SP-40, April 1992, pp. 770-783.

[161] P.P. Vaidyanathan, *Multirate Systems and Filter Banks*, Brendan M. Stewart, Ed., P T R Prentice Hall Inc., 1993, pp. 353-391.

[162] P.P. Vaidyanathan, *Multirate Systems and Filter Banks*, Brendan M. Stewart, Ed., P T R Prentice Hall Inc., 1993, pp. 125-126.

[163] R.E. Crochiere and L.R. Rabiner, *Multirate Digital Signal Processing*, Barbara Cassel, Ed.,Prentice Hall Inc., 1983, pp. 289-396.

[164] H.S. Malvar, "Modulated QMF filter banks with perfect reconstruction," *Electronics Letters*, vol. 26, June 1990, pp. 906-907.

[165] P.P. Vaidyanathan, "Theory and design of M-channel maximally decimated quadrature mirror filters with arbitrary M, having the perfect reconstruction property," *IEEE Transactions on Signal Processing*, vol. 35, April 1987, pp. 476-492.

[166] P.P. Vaidyanathan, T.Q. Nguyen, Z. Doganata and T. Saramaki, "Improved technique for design of perfect reconstruction FIR QMF banks with lossless polyphase matrices," *IEEE Transactions on Acoustics, Speech ad Signal Processing*, vol. 37, no. 7, July 1989, pp. 1042-1056.

[167] T.Q. Nguyen, "Two-channel perfect-reconstruction FIR QMF structures which yield linear-phase analysis and synthesis filters," *IEEE Transactions on Acoustics, Speech ad Signal Processing*, vol. 37, no. 5, May 1989, pp. 676-690.

[168] Z. Doğanata, P.P. Vaidyanathan and T.Q. Nguyen, "General synthesis procedures for FIR lossless transfer matrices, for perfect-reconstruction multirate filter bank applications," *IEEE Transactions on Acoustics, Speech ad Signal Processing*, vol. 36, no. 10, October, 1988, pp. 1561-1574.

[169] K. Swaminathan and P.P. Vaidyanathan, "Theory and design of uniform DFT, parallel, quadrature mirror filter banks," *IEEE Transactions on Circuits and Systems*, vol. Cas-33, no. 12, December 1986, pp. 1170-1191.

[170] P.P. Vaidyanathan, "Passive cascaded-lattice structures for low-sensitivity FIR filter design, with applications to filter banks," *IEEE Transactions on Circuits and Systems*, vol. 33, November 1986, pp. 1045-1064.

[171] P.P. Vaidyanathan, *Multirate Systems and Filter Banks*, Brendan M. Stewart, Ed., P T R Prentice Hall Inc., 1993, pp. 337-352.

[172] Numerical Algorithms Group, *NAG Fortran Library – Mark 13*, Downers Grover, July 1988.

[173] The Mathworks Inc., *Signal Processing Toolbox User's Guide*, December 1996.

[174] J.C. Schmidt, J.C. Rutledge, "Multichannel dynamic range compression for music signals," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, IEEE, May 1996, pp. 1013-1016.

[175] R. P. Lippmann, L. D. Braida, and N. I. Durlach, "Study of multi-channel amplitude compression and linear amplification for persons with sensorineural hearing loss," *Journal of the Acoustical Society of America*, vol. 69, Feb. 1981, pp. 524-534.

[176] D. Byrne and H. Dillon, "The National Acoustic Laboratories' (NAL) New Procedure for Selecting the Gain and Frequency Response of a Hearing Aid," *Ear and Hearing*, vol. 7, no. 4, 1986.

[177] E. Villchur, "Signal processing to improve speech intelligibility in perceptive deafness," *Journal of the Acoustical Society of America*, vol. 53, no. 6, 1973, pp. 1646-1657.

[178] G.W. Walker, D. Byrne and H. Dillon, "The effects of multichannel compression/expansion amplification on the intelligibility of nonsense syllables in noise," *Journal of the Acoustical Society of America*, vol. 76, no. 3, September 1984, pp. 747-757.

[179] E.W. Yund and K.M. Buckles, "Enhanced speech perception at low signal-to-noise ratios for multichannel compression hearing aids," *Journal of the Acoustical Society of America*, vol. 97, no. 2, February 1995, pp. 1224-1240.

[180] E.W. Yund and K.M. Buckles, "Multichannel hearing aids: Effect of number of channels on speech discrimination in noise," *Journal of the Acoustical Society of America*, vol. 97, no. 2, February 1995, pp. 1206-1223.

[181] S. Arlinger, *Manual of Practical Audiometry*, Volume II, Whurr Publishers, London, 1991.

[182] J. Katz, *Loudness Balance Procedures*, Baltimore, p167, 2nd edition, 1978.

[183] J.B. Allen, J.L. Hall and P.S. Jeng, "Loudness growth in ½-octave bands (LGOB) – A procedure for the assessment of loudness," *Journal of the Acoustical Society of America*, vol. 88, no. 2, August 1990, pp. 745-753.

[184] L.J. Kanis and E. de Boer, "Frequency dependence of otoacoustic distortion products in a locally active model of the cochlea," *Journal of the Acoustical Society of America*, vol. 101, no. 3, March 1997, pp. 1527-1531.

[185] R.D. Patterson, I. Nimmo-Smith, D.L. Weber and R. Milroy, "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold," *Journal of the Acoustical Society of America*, vol. 72, no. 6, December 1982, pp. 1788-1803.

[186] B.R. Glasberg and B.J. Moore, "Derivation of auditory filter shapes from notched-noise data," *Hearing Research*, Elsevier Science Publishers B.V., vol. 47, 1990, pp. 103-138.

[187] M.R. Leek and V. Summers, "Timbre Discrimination by Hearing-Impaired Listeners," in Walt Jesteadt (Ed.), *Modelling Sensorineural Hearing Loss*, Lawrence Erlbaum Associates, 1997, pp. 343-357.

[188] D.M. Green, "A maximum-likelihood method for estimating thresholds in a yes-no task, " *Journal of the Acoustical Society of America*, vol. 93, no. 4, April 1993, pp. 2096-2105.

[189] X. Gu and D.M. Green, "Further studies of a maximum-likelihood yes-no procedure," *Journal of the Acoustical Society of America*, vol. 06, no. 1, July 1994, pp. 93-101.

[190] P.P. Vaidyanathan, *Multirate Systems and Filter Banks*, Brendan M. Stewart, Ed., P T R Prentice Hall Inc., 1993, pp. 119-120.

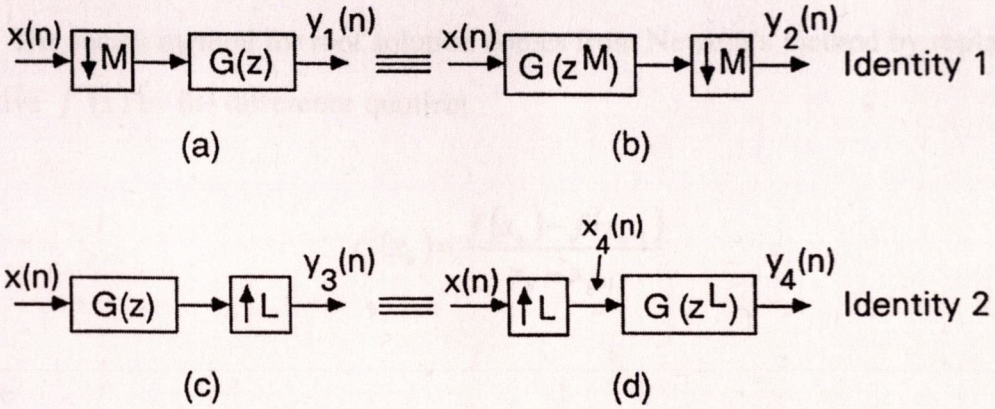
[191] P.P. Vaidyanathan, *Multirate Systems and Filter Banks*, Brendan M. Stewart, Ed., P T R Prentice Hall Inc., 1993, pp. 296-297.

[192] Erwin Kreysig, *Advanced Engineering Mathematics*, John Wiley & Sons, Inc., 7th edition, 1993.

Appendix

A1 The Noble Identities

When decimators and/or interpolators are cascaded with linear time-invariant systems the noble identities are very useful in implementation of these systems [190]. These arrangements frequently arise using polyphase representations of multirate systems.



In identity 1, the output of systems (a) and (b) are equal.

In identity 2, the output of systems (c) and (d) are equal.

The noble identities are guaranteed to work only if G(z) is rational.

A2 Power Complementary Functions

A set of M transfer functions are said to be power complementary if

$$\sum_{k=0}^{M-1}\left|H_k\left(e^{j\omega}\right)\right|^2 = c, \forall \omega$$

where $c > 0$ and is constant [191].

A3 The Method of Bisections and the Secant Method for Root Solution

The method of bisections is a slowly convergent method for finding a solution of $f(x)=0$ with continuous f, based on the intermediate value theorem, which states that if a continuous function f has opposite signs at some $x=a$ and $x=b(\rangle a)$, that is either $f(a)<0, f(b)>0$ or $f(a)>0, f(b)<0$, then f must be 0 somewhere on [a, b]. The solution is found by repeated bisection of the interval and in each iteration picking that half which also satisfies that sign condition [192].

The secant method for root solution comes from Newton's method by replacing the derivative $f'(x)$ by the difference quotient

$$f'(x_n)\approx\frac{f(x_n)-f(x_{n-1})}{x_n-x_{n-1}}$$

To give

$$x_{n+1}=x_n-f(x_n)\frac{x_n-x_{n-1}}{f(x_n)-f(x_{n-1})}$$

The solution is arrived at in a similar manner to Newton's method for root solution, the secant method, however, does not require access to the derivative of the function being solved [192].