UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

MODELLING PSYCHOACOUSTIC EXPERIMENT OUTCOMES FROM SPACE-TIME NEURAL FIRING PATTERNS FOR ACOUSTIC AND ELECTRICAL HEARING

by

Stuart Joseph Cecil Smith

Submitted in partial fulfilment of the requirements of the degree
Master of Engineering (Bioengineering)
in the
Faculty of Engineering, the Built Environment and Information Technology
UNIVERSITY OF PRETORIA

16 February 2011

SUMMARY

Modelling psychoacoustic experiment outcomes from space-time neural firing patterns for acoustic and electrical hearing

by

Stuart Joseph Cecil Smith

Supervisor : Prof JJ Hanekom
Department : Electrical, Electronic and Computer Engineering
Degree : Master of Engineering (Bioengineering)

Models predicting the outcomes of psychoacoustic experiments can assist in increasing the pace of research in the field of cochlear implants. In addition, they can provide insight into the functioning of complex biological systems. By utilising modelled space-time patterns of action potentials, a model of the central auditory nervous system was developed that predicts the outcomes of psychoacoustic experiments. The action potentials were decoded by means of an optimal central estimator used in conjunction with a state estimation technique (either a Viterbi decoder or a centre-of-gravity decoder). The approach used allowed the central auditory nervous system model (or central estimator) to use normal and cochlear implant hearing interchangeably as input. Modelled action potential patterns elicited by acoustic stimulation as well as those generated by the electrical pulse trains from a Nucleus 24 cochlear implant were used as input to the interchangeable central estimator. The latter was used to simulate frequency discrimination and intensity discrimination experimental outcomes, previously obtained in a sound field environment. The model predictions were similar in trend and magnitude to experimental results for a limited range of frequencies and intensities when using the Viterbi decoder. The centre-of-gravity decoder provided better frequency discrimination trends over a wider range of frequencies than the Viterbi decoder. However, the magnitudes predicted by the centre-of-gravity decoder were closer to optimal detection model results and therefore far superior to experimentally obtained results. Limits in the performance of the model are due to the tuning functions and the limited range of auditory channels used in frequency and intensity discrimination respectively. The results

suggest that the central auditory nervous system uses a likelihood-based decoding technique in frequency and intensity discrimination tasks, as has also been observed in visual motion detection. This reaffirms the need to consider the processing mechanisms employed by the central auditory nervous system when designing speech processing strategies for cochlear implants. The proposed central mechanism provides a solid basis for the development of future central processing models of this kind.

KEY WORDS

Cochlear implant, frequency discrimination, intensity discrimination, sound field, optimal estimation, Viterbi decoder, centre-of-gravity decoder

OPSOMMING

Modellering van uitkomste van psigoakoestiese eksperimente met kogleêre inplantings uit ruimte-tyd neurale aktiveringspatrone

deur

Stuart Joseph Cecil Smith

Studieleier	:	Prof JJ Hanekom
Departement	:	Elektriese, Elektroniese en Rekenaaringenieurswese
Graad	:	Magister in Ingeniersewese (Bio-ingeniersewese)

Modelle wat die uitkoms van psigoakoestiese eksperimente in kogleêre inplantinggehoor voorspel, kan die pas van navorsing in hierdie veld aanhelp en insig ten opsigte van die werking van ingewikkelde biologiese stelsels bevorder. Met die gebruik van ruimte-tyd aksiepotensiale kon ’n model van die sentrale gehoorstelsel ontwikkel word wat die uitkoms van psigoakoestiese eksperimente bevredigend kan voorspel. Aksiepotensiale is met behulp van ’n optimale sentrale skatter en toestandvoorspeller (òf Viterbi- òf swaartepuntdekodeerder) gedekodeer. Met hierdie benadering kon die sentrale gehoorstelselmodel (of sentrale skatter) gebruik word met òf normale òf kogleêre inplantinggehoor as inset. Gemodelleerde aksiepotensiaalpatrone soos opgewek deur akoestiese stimulasie, sowel as dié deur elektriese pulsreekse met ’n Nucleus-24 kogleêre inplanting gegenereer, was die inset tot die uitruilbare sentrale skatter. Laasgenoemde is gebruik om bestaande frekwensie- en intensiteitsdiskriminasieresultate soos in ’n vrye veld klankomgewing bepaal, te simuleer. Binne beperkte frekwensie- en intensiteitsbereike kom die voorspellings met die Viterbidekodeerder ooreen met eksperimentele data in tendens en grootte. Beter frekwensiediskriminasietendense oor ’n wyer frekwensiebereik is met die swaartepuntdekodeerder as met die Viterbidekodeerder verkry. Die grootte van voorspelde resultate met die swaartepuntdekodeerder is egter nader aan dié van ’n optimale waarnemermodel en daarom veel beter as eksperimentele resultate. Funksionele beperkings van die model kan toegeskryf word aan senuweeinstemkurwes en die beperkte aantal ouditiewe kanale beskikbaar vir onderskeideik frekwensie- en intensiteitsdiskriminasie. Die resultate dui daarop dat die sentrale gehoorstelsel moont-

lik ’n waarskynlikheid-gebaseerde dekoderingstegniek gebruik tydens frekwensie- en intensiteitsdiskriminasie, soortgelyk aan dié vir visuele bewegingsdeteksie. Die bevindinge beklemtoon opnuut dat prosesseringsmeganismes in ag geneem moet word tydens die ontwerp van spraakverwerkingstrategieë vir kogleêre inplantings. Die voorgestelde sentrale meganisme kan as basis dien vir toekomstige ontwikkeling van sentrale verwerkingsmodelle.

SLEUTELWOORDE

Kogleêre inplanting, frekwensiediskriminasie, intensiteitsdiskriminasie, vrye veld, optimale skatter, Viterbidekodeerder, swaartepuntdekodeerder

LIST OF ABBREVIATIONS

ACE	:	Advanced Combinational Encoder	(p. 17)
AP	:	Action Potential	(p. 4)
CANS	:	Central Auditory Nervous System	(p. 4)
CF	:	Characteristic Frequency	(p. 25)
CI	:	Cochlear Implant	(p. 2)
CID	:	Central Institute for the Deaf	(p. 17)
CIS	:	Continuous Interleaved Sampling	(p. 2)
COG	:	Centre-Of-Gravity	(p. 64)
ERB	:	Equivalent Rectangular Bandwidth	(p. 23)
FFT	:	Fast Fourier Transform	(p. 70)
HMM	:	Hidden Markov Model	(p. 65)
ISI	:	Interspike Interval	(p. 24)
IRE	:	Instantaneous Rate Estimator	(p. 99)
JND	:	Just-Noticeable Difference	(p. 30)
LGF	:	Loudness Growth Function	(p. 76)
NH	:	Normal Hearing	(p. 3)
RI	:	Rate-Intenstiy	(p. 14)
RS	:	Relative Spread	(p. 16)
SDT	:	Signal Detection Theory	(p. 13)
SPEAK	:	Spectral Peak	(p. 2)
SPL	:	Sound Pressure Level	(p. 40)
2AFC	:	Two-Alternative Forced Choice	(p. 21)

Table of Contents

CHAPTER 1

INTRODUCTION

1.1 PROBLEM STATEMENT

1.1.1 Context of the problem

The mechanism of hearing is complex. Sound waves from the external environment travel down the ear canal, where they cause the tympanic membrane to vibrate, which in turn translates into the mechanical motion of the ossicles of the middle ear. This motion causes movement of the oval window at the opening of the inner ear and this creates pressure waves that travel through the fluid contained within the cochlea. The pressure variations in this fluid result in the displacement of the basilar membrane. The vibration of this membrane acts as an acoustic filter because maximal vibrations as a result of high frequencies occur close to the round window, while maximal vibrations as a result of low frequencies occur at the cochlea apex (Clark, 2003). As the basilar membrane moves, the hair cells of the organ of Corti shift and deform. This process results in neural firing, which is interpreted by the brain as the sensation of hearing.

The hearing process can be irreparably damaged if one component does not function correctly. Wilson and Dorman (2008a) state that the principal cause of hearing loss is damage to, or the complete destruction of, the sensory hair cells. These structures are rather vulnerable to damage, which may be caused by genetic defects, diseases (e.g. rubella and menin-

gitis), extremely loud sounds, specific chemical compounds and age. When these hair cells are damaged or destroyed, the link between the outside world and the auditory processes of the brain is severed. The purpose of a cochlear implant (CI) is to bypass the hair cells and restore the link by electrically stimulating the surviving neurons in the cochlea.

The general setup of a CI is as follows: a directional microphone detects and converts sound waves into electrical voltages, which are then transmitted to a speech processor that is usually worn behind the ear or carried at the waist. Examples include the SPrint, ESPrit-3G, and Clarion S speech processors, the former two were designed by Cochlear Ltd and the latter one was designed by Advanced Bionics Corp. (Clark, 2003). The speech processor performs the necessary signal processing on the signal received from the microphone, depending on the speech processing strategy (otherwise known as a speech processing algorithm) that has been implemented on the speech processor. Examples of these strategies include the continuous interleaved sampling (CIS) and spectral peak (SPEAK) strategies (Loizou, 2006). When mention is made of an implant, such as the Nucleus 24[1] CI, it usually refers to the receiver-stimulator, but can refer to the entire device including the receiver-stimulator, the speech processor and the microphone.

The speech processor implements a coding scheme to represent the sound at each instant using the available stimulus parameters: these include site of stimulation (encoding frequency) and stimulus level (encoding signal intensity) for most commercial implants. This code is transmitted via a round aerial to the receiver-stimulator implanted just beneath the skin. The signal is decoded by the receiver-stimulator into a series of pulsatile current waveforms, which are transmitted to the electrode array implanted in the scala tympani of the cochlea. The current waveforms stimulate the auditory nerve fibres that can in turn transmit the information to the brain. This stimulation is intended to provide a meaningful representation of speech and environmental sounds.

Research in the field of CIs has been underway for approximately three decades, but the progress in this time has been remarkable. Many implantees are now able to obtain 100% correct scores on sentence recognition tests in quiet (Wilson and Dorman, 2008a).

An important aspect in the research and development of any technology is modelling, as it provides a means of repeating experiments tirelessly in order to provide one with an improved understanding of the modelled system. Although models are not perfect represen-

[1] A product of Cochlear Ltd, Lane Cove, Australia.

tations of their real-world counterparts, they can offer adequate representations of systems and supply useful information. In the context of CI research, modelling can be applied to a number of the research areas in the field (Hanekom and Hanekom, 2000).

Specifically, physiological models of the cochlea have been developed to study the current spread caused by electrode stimulation of the cochlea (e.g. Briaire and Frijns, 2000). They have also been used extensively in the study of the perceptual mechanisms in humans (e.g. Xu and Collins, 2007; Hanekom and Krüger, 2001; Bruce, White, Irlicht, O'Leary and Clark, 1999b; Forrest and Formby, 1996; Siebert, 1970).

1.1.2 Research gap

Perceptual models of the cochlea have generally been directed at predicting specific psychoacoustic properties (by predicting the outcomes of perceptual experiments in normal hearing (NH) and cochlear implantee subjects), such as gap detection or discrimination (Forrest and Formby, 1996; Creelman, 1962), frequency discrimination (Hanekom and Krüger, 2001), threshold levels and dynamic range (Xu and Collins, 2007; Xu and Collins, 2005; Xu and Collins, 2004) and intensity discrimination (Bruce *et al.*, 1999b; Xu and Collins, 2005). Relatively few models have been developed to predict multiple perceptual outcomes (e.g. Heinz, Colburn and Carney, 2001a; Jepsen, Ewert and Dau, 2008; Dau, Püschel and Kohlrausch, 1996a) and fewer still have been developed to predict the outcomes of both acoustic and electrical hearing (Hanekom, 2000), in which case only the direct stimulation of nerves by electrical stimuli was investigated, and not the contribution of CI processing. The reason for this is that the link between stimulus and perception is not fully understood, and hence the models describing this link are created with certain restrictive assumptions, making them applicable to only one of the hearing modalities.

The ongoing development in the field of CIs has been aided to a large extent by modelling. Modelling has assisted researchers in understanding the underlying mechanisms of the biophysical and perceptual processes involved in hearing. This understanding has been translated into improvements in the signal processing of hearing prostheses thus far (Nogueira, Büchner, Lenarz and Edler, 2005; Clark, 1996), but owing to the problems presented by perceptual models, there is no clear way forward in terms of how further improvements should be made.

The problem of perceptual modelling of the auditory system was tackled in this study. This was done based on the reasoning that if a good perceptual model for NH listeners and cochlear implantees can be developed, and new conclusions about the link between stimulus and perception can be drawn, it is possible that further advancements in hearing and stimulus algorithms can be made. The aim of this research was therefore to investigate the possibility of using the same model of the central auditory nervous system[2] (CANS) in predicting perceptual experiment outcomes for both NH and CI listeners. The implications of a model that achieves this (see chapters 4 and 5) are discussed in chapter 6.

The possibility of the development of a model that can be generally applied to multiple perceptual experiments was investigated by examining the coding strategies of the auditory system, as well as existing models of perception (e.g. Hanekom, 2000; Heinz *et al.*, 2001a). The model was developed in such a way that it could be used to predict the outcomes of acoustic and electrical hearing experiments and was used to predict the outcomes of two of the most common perceptual experiments, namely frequency discrimination and intensity discrimination. In order to provide a good idea of how CI processing affects perceptual outcomes, the model considered stimuli presented in a sound field[3] environment instead of at the electrode level. Prior to this study, the modelling of CI discrimination thresholds in a sound field environment had not yet been performed.

1.2 RESEARCH OBJECTIVES AND QUESTIONS

The research gap described in section 1.1.2 provides an introduction to what this study aimed to accomplish. In order to clearly define specific objectives that should be achieved, research questions that needed to be answered were formulated before commencing with the study. These questions directed the course of the study in conjunction with the central hypothesis, and outlined the approach and tasks that were performed. The central hypothesis follows below.

- Using a model of electrically stimulated space-time action potential (AP) patterns, it

[2]The CANS can be thought of as an observer of neural spikes originating from the auditory periphery, that needs to interpret the nature of the auditory environment from the neural spikes alone. The CANS can be thought of as a "central observer" within this paradigm.

[3]Sound field measurements are performed by presenting acoustic signals to a subject through a sound source in a sound room, as opposed to electrode-level stimulation (ISO 8253-2, 2009).

is possible to predict the outcomes of perceptual experiments for NH and CI listeners, using an identical central observer for both modalities.

This central hypothesis can be separated into further questions:

- How do the outcomes of the model compare with actual results obtained in experimental conditions with NH and CI human subjects?
- If the predicted outcomes of the model are similar to experimental outcomes for at least one perceptual experiment, can the model be applied to other perceptual experiments by adjusting a small set of parameters, while maintaining the model structure?
- What can be inferred about the nature of the central processing mechanism in the auditory system from the answers to the above questions?
- Does the approach taken provide a feasible framework in which future models of a similar nature can be developed?

From these research questions, the main tasks that were carried out in the study are outlined below:

- An optimal estimation model was designed in Matlab[4], based on the statistics of space-time neural firing resulting from acoustic and electrical stimulation, to function as the central observer.
- A model of space-time AP firing patterns resulting from acoustic stimulation was implemented and used as input for the central observer to model acoustic hearing.
- A model of space-time AP firing patterns resulting from electrical stimulation was implemented and used as input for the central observer to model electrical hearing.
- A model of the processing in a commercially available CI was generated to act as input for the model of electrical stimulation, so that a reliable representation of the encoding used by CIs was available to the central observer.

[4]Matlab is a product of the Mathworks, available from www.mathworks.com.

- The model had to be able to predict the outcomes of pure-tone frequency discrimination and intensity discrimination experiments. This had to be achieved by minimal adjustment of the parameters of the central estimator.
- The outcomes predicted by the model had to be compared with results obtained in experimental conditions from cochlear implantees, such as those given in Pretorius and Hanekom (2008) and Rogers, Healy and Montgomery (2006), for frequency and intensity discrimination respectively.
- Conclusions had to be drawn from the observed predictions, as well as the comparison of these with actual experimental data, about the nature of the processing mechanisms in the CANS.

1.3 APPROACH

In most studies where an attempt is made to understand the nature of a complex system, a modelling approach to the problem is taken in order to obtain a thorough understanding of the system being dealt with (e.g. a thorough framework for the modelling of neurological systems is provided by Eliasmith and Anderson, 2003). The modelling approach is applied in this study. By using an appropriate model of the system, simulations can be performed to predict experimental outcomes with reasonable accuracy far more rapidly than perceptual experiments can be performed, thus accelerating the pace of the research being performed. However, most models only yield the average performance that is observed in perceptual experiments and usually fail to account for subtle physiological differences that result in discrepancies between the results of human subjects (e.g. Moore, 1973; Pretorius and Hanekom, 2008; Wier, Jesteadt and Green, 1977). A series of articles by Cohen (2009a; 2009b; 2009c; 2009d; 2009e) modelled the neural excitation elicited by CIs in six subjects. The models that Cohen developed were customised to the results of the subjects that were tested.

In previous psychophysical studies, attempts have been made to model the outcomes of various psychoacoustic experiments in both NH listeners (e.g. Dau *et al.*, 1996a; Dau, Püschel and Kohlrausch, 1996b; Forrest and Formby, 1996; Goldstein and Srulovicz, 1977; Hanekom and Krüger, 2001; Jepsen *et al.*, 2008) and CI listeners (e.g. Bruce *et al.*, 1999b; Hanekom, 1999; Xu and Collins, 2004; Xu and Collins, 2005; Xu and Collins, 2007). These studies

have aimed to model either acoustic or electrical hearing. Optimal estimation techniques have been utilised in only a few studies to model NH percepts (such as in the studies by Hanekom, 1999; Hanekom and Krüger, 2001). Optimal estimation provides the central observer with the best possible estimate of a certain parameter (such as the instantaneous firing rate on a certain nerve channel), which the central observer can use to determine the nature of an input stimulus.

For the purposes of this study, and to answer the research questions posed, an optimal estimation model was designed so that parameters could be extracted for every time instant in which space-time neural spike trains are provided. This serves to provide a more accurate model of how the CANS processes incoming neural data, i.e. it processes information as it is received, as opposed to "spike-counting" models, which utilise the average rate (Siebert, 1968). The approach was based on the hypothesis that the CANS functions in a similar manner to an optimal estimator and that the outcomes of psychoacoustic experiments can be predicted by such an estimator (as pioneered by Siebert, 1970).

The first important step in this study was to obtain a reliable model of space-time AP patterns. Such a model was used as input to the central estimator and was designed to provide a close approximation of the neural firing patterns that would normally be obtained from a cochlea as it undergoes either acoustic or electrical stimulation. Optimal estimators, such as the Kalman filter (Kalman, 1960), require a description of the system in order to function. Such an estimator was a possible candidate for the application described herein, however, a Kalman filter was not used in this study.

It was required that the central estimator fulfil certain criteria so that it could be applied in the manner mentioned above. It needed to be applicable to two different hearing modalities and two different perceptual tests, and therefore should not have been specific to the nature of input that it receives, i.e. it had to be able to process the AP patterns generated either acoustically or electrically. In order to draw reasonable conclusions from the model, it was necessary that the nature of the central estimator was not substantially changed when switching between the two hearing modes, so that any differences in the predicted outcomes could be attributed to the mode of hearing rather than parameter changes made in the model. By judging the validity of the perceptual predictions made by the model, conclusions were drawn as to how the CANS functions based on the structure of the model.

Processing of the parameter estimates was set to be dependent on the predefined perceptual

task, the outcomes of which were estimated by the central observer. The underlying principle of these predictions was based on that originally proposed by Siebert (1965; 1968; 1970), where the standard deviation of the optimal estimator was used to predict the outcomes of perceptual experiments. Ultimately, a central observer was developed and used to predict the outcomes of certain perceptual experiments. Results from the simulations were compared to actual experimental data in order to answer the research questions posed in section 1.2, as well as to test the central hypothesis. More specifically, the central estimator's predictions were compared to published experimental data of frequency discrimination experiments (e.g. Pretorius and Hanekom, 2008; Sek and Moore, 1995) and intensity discrimination experiments (e.g. Florentine, Buus and Mason, 1987; Rogers *et al.*, 2006). Modelled data from literature (such as that by Heinz *et al.*, 2001a; Jepsen *et al.*, 2008) was also used in the comparisons. Figure 1.1 shows an outline of the iterative procedure that was followed in this study.

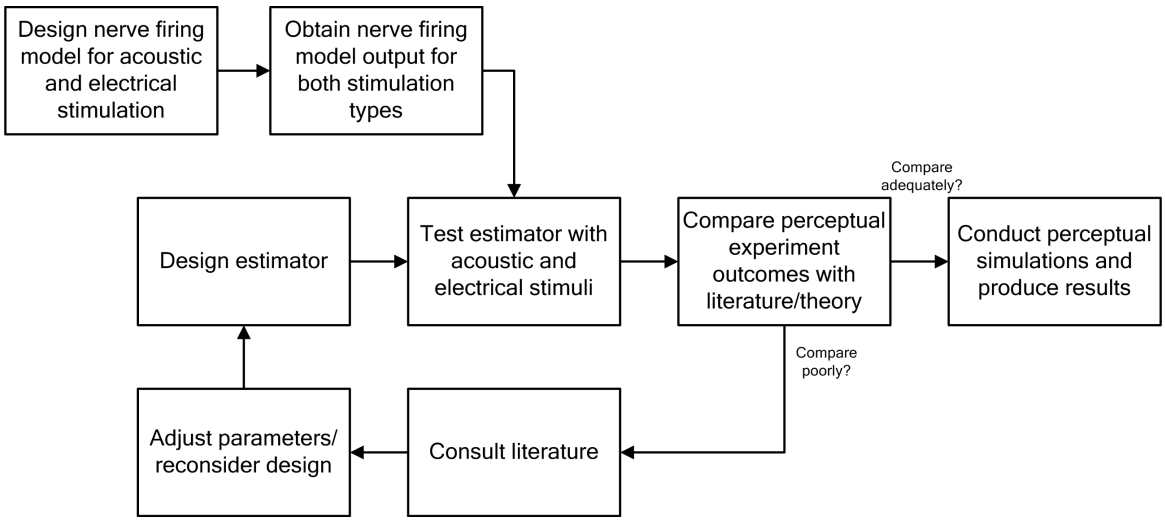

Figure 1.1.
Graphical representation of the approach that was used in this study.

1.4 RESEARCH CONTRIBUTION

At present, development in CIs that is related to perception is based on attempts to translate the outcomes of psychoacoustic experiments into the implementation of speech processors. Owing to the fact that these experiments are generally designed to test auditory perceptual thresholds under specific conditions, their application in speech processing is limited and may eventually slow the pace of progress in speech processing.

The study could therefore provide the following contributions:

- The model could offer insight into the impact of the electrically evoked AP patterns on perception and may thus provide possible routes into the future development in CIs.
- The model could supply a method of predicting the outcomes of perceptual experiments.
- The model may provide new insights into the functioning of the CANS, potentially providing a basis for new processing methods in CI signal processors.

1.5 OVERVIEW OF THE STUDY

This section provides an overview of this document, chapter by chapter. The dissertation is divided into the following chapters:

In Chapter 2 the background argument necessary for understanding the research problem is presented. The research work performed on the various knowledge areas that is relevant to the problem is detailed with the necessary literature references. Psychophysical modelling is discussed, presenting the reader with a background on the currently available neural firing models for acoustic and electrical hearing. A background on CIs is given, detailing the signal processing strategies used by these devices, as well as providing information on the general stimulation parameters of these devices that might be applicable to the modelling of electrical hearing. Finally, an overview of the literature on both frequency and intensity discrimination in the context of NH and CI experiments is provided.

In Chapter 3 the methods used in the development of the model are discussed. The model is described in full mathematical detail. This chapter includes detailed descriptions of the space-time AP model of acoustic stimulation, the CI model and the space-time AP model of electrical stimulation, as well as the model of the central estimator. Sample outputs are provided graphically for each stage of the model.

In Chapter 4 the results of the frequency discrimination predictions are provided. The parameters used in the model for both perceptual experiments are listed in this chapter accompanied by justification for the chosen parameters. Frequency discrimination predictions as a function of different parameters are provided for NH and CI listeners, compared to data from literature. Discussions on the new modelled data and its comparison to literature results are presented.

In Chapter 5 the results of the intensity discrimination predictions are provided. Intensity discrimination predictions as a function of different parameters are provided for NH and CI listeners, compared to data from literature. Discussions on the new modelled data and its comparison to literature results are presented.

In Chapter 6 a general discussion exploring the implications of the findings from chapters 4 and 5 is provided. This discussion looks at the findings from a general perspective.

In Chapter 7 the dissertation is concluded with a brief overview of the research describing how the research questions were answered. A presentation of the main findings of the study, as well as directions for future research are provided.

CHAPTER 2

LITERATURE REVIEW

2.1 CHAPTER OBJECTIVES

This chapter provides an overview of the relevant literature. First, various models that have been developed for the purpose of predicting psychoacoustical data are reviewed. Additionally, information on the neural models that often provide a basis for perceptual models is provided. Both acoustically stimulated and electrically stimulated models of this nature are covered. The aim of this section is to provide insight into how the modelling problems dealt with in this study have been approached in the past, thus placing the study into context, and to lay a foundation for the structure of the model and its various modules that are described in chapter 3.

Next, a section on CIs continues from where the description in chapter 1 left off, and provides an overview of some of the processing strategies that are currently in use. This is to provide an idea of the types of CIs and signal processing strategies that are currently available, and why the particular CI model that was used in this study was chosen (see section 3.3.6). A review of electrical stimulation parameters is provided as a basis for the decisions that were made for the model of excitation spread detailed in section 3.3.6.2.

This is followed by a section providing an overview of the study of frequency discrimination in psychophysics for acoustic hearing and electrical hearing. Finally, an overview of intensity discrimination is provided, also for both acoustic and electrical hearing. The reviews

of the psychophysical studies that have been carried out before are intended to provide a background of the mechanisms that the CANS probably uses in the aforementioned perceptual tasks. This places the findings of the study into context (see sections 4.4 and 5.4 and chapter 6) and provides a basis for some of the decisions that were made in the design of the model (see section 3.3).

2.2 PLACING THE PRESENT MODEL INTO CONTEXT

Chapter 1 describes the importance of modelling in the study of CIs. Much attention has paid to the subject of mathematical modelling of perception (e.g. the books of Green and Swets, 1966; Zwicker and Fastl, 1999), and this work continues to play an important role in the study of perceptual modelling. This section provides an overview of how the task of perceptual modelling has been approached before in order to place the model presented in this study into context. A review of neural firing models that are associated with perceptual models is also provided here.

2.2.1 Existing models of auditory perception

AP patterns are generated by the stimulation of neurons in the sense organs and encode information about the nature of the stimulus for interpretation by the brain. The precise method that the brain uses to encode information in spike trains is still not known (Eliasmith and Anderson, 2003), but given the nature of the auditory system, certain hypotheses have been developed in an attempt to explain how this is done. Some examples of neural encoders that have been used include an average rate code (Rieke, Warland, de Ruyter van Steveninck and Bialek, 1997), a timing code (e.g. Goldstein and Srulovicz, 1977; Hanekom and Krüger, 2001; Rose, Brugge, Anderson and Hind, 1967), a population code (e.g. Pouget, Dayan and Zemel, 2000; Sanger, 2003) and a place code (Zwicker, 1970), which has specific application in the auditory system and can be thought of as a type of population code. These encoders do not necessarily need to be used independently: most models make use of two or more of the above-mentioned encoders.

Encoding implies that decoding must be present in order for the signals to be usefully interpreted. The means by which the signals are decoded are numerous. Based on the hypothesis

that the brain uses information provided to it optimally, optimal detectors have been used as decoders. An example of this is the Kalman filter (Kalman, 1960) used by Hanekom and Krüger (2001) to model frequency discrimination.

One of the most well-established optimal methods is signal detection theory (SDT) (Green and Swets, 1966). SDT is useful as a decoder because it provides a means of separating factors that are unrelated to signal discriminability, such as bias, criterion and motivation, from sensory capabilities. Trial-to-trial variation when using the same stimulus can also be accounted for (Moore, 2003). These features make SDT an attractive option, but it often yields results that far exceed experimental data (Gresham and Collins, 1998; Heinz *et al.*, 2001a; Heinz, Colburn and Carney, 2001b; Huettel and Collins, 1999; Huettel and Collins, 2004; Siebert, 1970). Gresham and Collins (1999a) attempted to resolve this discrepancy by including phase uncertainty in the detector, although they still achieved unrealistically superior results. They proposed that some form of internal noise be added to the system in order to achieve more realistic values. Internal noise was added to the masking models by Dau *et al.* (1996a), Dau *et al.* (1996b) and Jepsen *et al.* (2008). In these cases, the internal noise source is difficult to account for biologically, but was used to include a level of information loss in the system. This illustrates one of the gaps in understanding with regard to optimal detection: the difference in the performance in psychophysical tasks predicted by optimal detectors and those actually obtained from humans cannot easily be accounted for.

Another optimal decoding approach was used by Hanekom (1999) and Hanekom (2000). Hanekom made use of an optimal estimator by Rudemo (1972), which estimated the instantaneous rate of a Poisson process. The estimator was originally designed for applications in reliability and replacement theory (which has also been applied to the auditory nervous system by Geldenhuys, 2007). The estimator yields the approximate rate of the Poisson process and is optimal in the least-squares sense. From this representation of a population of nerve fibres, a decoder such as a Viterbi decoder (Forney, 1973; Viterbi, 1967), along with the principles of Siebert (1970), can be used to decode the rate (this is what was used by Hanekom, 1999). The model used by Hanekom (1999; 2000) succeeded in predicting the trends in frequency discrimination to a limited extent.

Carney (1993; 1994) used a spatiotemporal approach to model intensity encoding. This model was dependent on the response patterns of a population of nerve fibre responses. Coincidence detection was used as the decoder. The model used both the place and timing properties of auditory neural activations, instead of just one of the two dimensions. The

model successfully predicted some of the effects that sensorineural hearing loss has on level encoding. Similarly, Patterson, Allerhand and Giguere (1995) developed the idea of the auditory image model. This computational model simulated the space-time descriptions of the outputs of populations' auditory neurons. Many models have utilised this idea and have been used to simulate various percepts and model certain processing aspects in the CANS (Bleeck, Ives and Patterson, 2004; Gresham and Collins, 1998; Gresham and Collins, 1999a; Huettel and Collins, 1999; Patterson, 2000).

Some of the models mentioned in this section have utilised APs to encode information from the auditory periphery (Hanekom, 1999; Hanekom, 2000; Hanekom and Krüger, 2001; Heinz *et al.*, 2001a; Heinz *et al.*, 2001b; Heinz, Colburn and Carney, 2002). This approach provides a number of benefits. The CANS model can extract information about the external environment from every spike, which according to Rieke *et al.* (1997), is an important consideration in perceptual models. Additionally, APs act as a bottleneck between the auditory periphery and the CANS which allows them to transmit only the information that is transferable to a firing rate by the rate-intensity (RI) functions. This means that a good representation of the information that is actually available to the CANS is provided.

In order to test the hypothesis proposed in section 1.2, a single model of the CANS is required for both acoustic and electrical hearing. Although SDT is a well-established method for predicting the outcomes of psychoacoustic experiments, the superior performance compared to that of human data is a persistent problem in SDT models. A spatiotemporal, AP-based model would provide a feasible solution to modelling both acoustic and electrical stimulation, provided that spatiotemporal AP patterns for both stimulation modes could be generated. Using such a model in conjunction with the CANS modelled as an optimal estimator (such as the technique used by Hanekom, 1999), would fulfil the requirement imposed by the hypothesis that a single central observer be used for both hearing modalities.

2.2.2 Neural firing models

When modelling perception, the APs of neurons are generally taken into consideration in some way. They are then processed by computational or analytical means (see section 2.2.1) in order to obtain a perceptual measure. The study of neural firing has seen much attention since Hodgkin and Huxley (1952) first modelled the bio-electrical dynamics of neurons. Their model described the initiation of APs by a set of differential equations that need to

be solved numerically. Variations of and modifications to this model have been proposed many times since it was first introduced (Frankenhaeuser and Huxley, 1964; Schwarz, Reid and Bostock, 1995; Smit, Hanekom and Hanekom, 2009b; Smit, Hanekom and Hanekom, 2009a; Wesselink, Holsheimer and Boom, 1999).

There are simpler alternatives to the complex Hodgkin and Huxley (1952) model, such as the leaky integrate-and-fire model (Eliasmith and Anderson, 2003). Another commonly used alternative is the approximation of neural firing as a Poisson process, which has been applied to modelling of the stimulation of the auditory nerve (e.g. Siebert, 1965; Siebert, 1970). This model of neural firing times has been considered so useful that it has seen much use in modelling the firing behaviour of auditory neurons (Carney, 1993; Colburn, 1973; Geldenhuys, 2007; Hanekom, 1999; Hanekom and Krüger, 2001; Johnson and Swami, 1983; Siebert, 1970; Teich and Lachs, 1979; Young and Barta, 1986).

Models of auditory neural firing as a result of electrical stimulation have also been developed (Bruce, White, Irlicht, O'Leary, Dynes, Javel and Clark, 1999c; Bruce, Irlicht, White, O'Leary, Dynes, Javel and Clark, 1999a; Bruce, Irlicht, White, O'Leary and Clark, 2000; Chen and Zhang, 2007; Imennov and Rubinstein, 2009; Macherey, Carlyon, Van Wieringen and Wouters, 2007; O'Leary, Clark and Tong, 1995; Rubinstein, Wilson, Finley and Abbas, 1999). Indeed, Hodgkin and Huxley's model can also be applied to electrical stimulation and variations of it have been applied to the cochlea (Rattay, Lutter and Felix, 2001b; Rattay, Leao and Felix, 2001a). This is necessary because the firing patterns induced by electrical stimulation are very different from those of acoustic stimulation (Javel, Tong, Shepherd and Clark, 1987).

The stochastic model developed by Bruce *et al.* (1999c) was based on a simple circuit representation of the system and models the dynamics of electrical stimulation relatively well (see also Bruce, 1997). The model incorporated the refractory effects observed in neurons (Plonsey and Barr, 2000), and included absolute refractory time[1] and relative refractory time[2]. It was developed in two parts, namely an analytical model and a computational model. The analytical version (fully described in Bruce *et al.*, 2000), is quite limited in its

[1]The absolute refractory time is the short period after a neural spike has occurred wherein no neural firing can occur. This is a result of the electro-chemical dynamics involved in neural firing (see Plonsey and Barr, 2000; Rattay, 1990).

[2]The relative refractory time is the period after the absolute refractory period wherein an abnormally high stimulus level is required to elicit an AP. Gradually, the stimulus level required decreases until it reaches normal levels. This effect is a result of the electro-chemical dynamics involved in neural firing (see Plonsey and Barr, 2000; Rattay, 1990).

assumptions but simulates much faster. The computational version is less restrictive in the assumptions made, but owing to its computational nature, takes a longer time to simulate. The model was able to predict the mean and the variance of discharge rates in experimental data.

Another simple model was developed by Chen and Zhang (2007) and was based on an integrate-and-fire model. A circuit-based representation was also used for this model, which was slightly more complicated than the one proposed by Bruce *et al.* (1999a; 1999b; 1999c). The study focused on the predictions the model made regarding different electrical input stimuli, and found that there is an optimal pulse width that can be used to represent the temporal structure of the stimulus waveform.

The model proposed by Imennov and Rubinstein (2009) was more computationally complex than the models of both Bruce *et al.* (1999a) and Chen and Zhang (2007). It was a biophysically based stochastic population model that succeeded in emulating detailed aspects of the auditory nerve, namely relative spread[3] (RS), jitter, spike latency, chronaxie and relative refractory period, although they did not include neural adaptation in their model. Macherey *et al.* (2007)'s integrator-resonator model (Macherey *et al.*, 2007) was able to account for both neural behavior and psychoacoustic data (much like the model of Bruce *et al.*, 1999a; Bruce *et al.*, 1999b; Bruce *et al.*, 1999c) such as the effects of pulse widths and pulse rates, but was also more complex than the models of Bruce *et al.* (1999a) and Chen and Zhang (2007).

From the review provided in the paragraphs above, it appears that a Poisson process provides a simple model of acoustic neural stimulation. Such a model would be well suited to simulating the firing patterns of neural populations, as it would not be as computationally intensive as the biologically accurate Hodgkin and Huxley (1952) model. The model by Bruce *et al.* (1999a; 1999b; 1999c) provides an equally simple model, but for electrical neural stimulation. Although there are other relatively simple models of electrical stimulation, the model by Bruce *et al.* can be applied in a similar way to the Poisson process model, owing to the fact that it yields the same type of output – the time of the neural firing. This would be useful in a perceptual model that requires both acoustic and electrical stimulation to be simulated.

[3]Relative spread is a concept introduced by Verveen (1961) that defines the slope of the rate-intensity function resulting from electrical stimuli. The greater the RS, the shallower the slope of the rate-intensity function becomes. RS is variable by the stimulating pulse rate and pulse width; see Bruce *et al.* (1999c).

2.3 COCHLEAR IMPLANTS

Chapter 1 described how CIs function. There are a number of different CIs on the market (Wilson and Dorman, 2008a), some based on early designs (Patrick, Busby and Gibson, 2006; Clark, Tong, Black, Forster, Patrick and Dewhurst, 1977). The three manufacturers of CIs that hold a 99% market share are Cochlear Ltd (Lane Cove, Australia), Advanced Bionics Corp. (Valencia, California, USA) and MED-EL GmbH (Innsbuck, Austria) (Wilson and Dorman, 2008a). Each of these companies has released many different CI models and speech processors – a review of the development history of Cochlear Ltd's Nucleus processor is described in detail in Patrick *et al.* (2006) – but there are only a small number of processing strategies that are widely used today. These include the CIS, *n*-of-*m*, advanced combinational encoder (ACE), SPEAK and HiRes strategies.

2.3.1 Processing strategies

The CIS strategy (Wilson, Finley, Lawson, Wolford, Eddington and Rabinowitz, 1991) is based on a fixed-filter principle. The audio signal from the CI microphone is processed initially by a front end, which normally includes pre-emphasis, automatic gain control and sensitivity control. From here, the signal is passed through six or more band-pass filters. The envelope of the outputs of these signals is then detected through a rectifier and low-pass filter. The analogue signal on each electrode (corresponding to each filter) modulates the current amplitudes of pulse trains, which are fed to the external transmitter and then on to the electrodes via the implant (Wilson and Dorman, 2008a; Loizou, 1999a). The pulses are interleaved so as to reduce channel interaction. Kessler, Loeb and Barker (1995) and Schindler, Kessler and Barker (1995) found that Clarion[4] CIS users had an open-set Central Institute for the Deaf (CID) sentence-recognition score of approximately 60%. The HiRes strategy is similar to the CIS strategy and is sometimes used synonymously. The differences between the two strategies lie in the rates, envelope detector cut-off frequencies and number of processing channels – HiRes uses relatively higher rates and cut-off frequencies and can use up to 16 channels (Wilson and Dorman, 2008a).

The *n*-of-*m*, ACE and SPEAK strategies are based on similar principles to one another (Wilson and Dorman, 2008a). They all rely on finding the spectral maxima of the signal

[4] A product of Advanced Bionics, Valencia, California, USA.

and presenting these most prominent signal components to the implant user. The SPEAK strategy analyses the signal and chooses five to 10 (normally six) spectral maxima of a set of 16 to 22 filters which are modulated onto pulse trains and passed to the external transmitter. The filter banks are programmable and can be changed to suit a user's specific needs (Loizou, 1999b). CI listeners utilising the SPEAK strategy scored about 71% on open-set CID sentence recognition tests (Clark, 2003). The ACE strategy is similar in principle to the SPEAK strategy, except that it allows representation of SPEAK at different rates and the ability to vary the number of channels up to a specific maximum, usually eight (Clark, 2003; Swanson, Van Baelen, Janssens, Goorevich, Nygard and Van Herck, 2008). This was done in an effort to provide additional timing information and reduce redundant information that is provided to the listener. ACE is one of the strategies available for use with the Nucleus 24 implant.

There have been a few studies performed to determine how the three most commonly used strategies, CIS, SPEAK and ACE, compare (Arndt, Staller, Arcaroli, Hines and Ebinger, 1999; Kiefer, Hohl, Stürzebecher, Pfennigdorff and Gstöettner, 2001; Nilson, Soli and Sullivan, 1994; Skinner, Holden, Whitford, Plant, Psarros and Holden, 2002), and in all cases the best performing or preferred strategy, depending on the test, was the ACE strategy. The majority of the tests made use of the Nucleus 24 implant. As an example, in the case of Arndt *et al.* (1999), 12 users were given the three strategies in different orders, adjusted to suit the needs of each user. Of these, 58% preferred the ACE strategy, 25% preferred the SPEAK strategy and 17% preferred the CIS strategy.

There are a few processing strategies that are currently undergoing clinical trials and testing. The Fine Hearing (Arnoldner, Riss, Brunner, Durisin, Baumgartner and Hamzavi, 2007; Hochmair, Nopp, Jolly, Schmidt, Schößer, Garnham and Anderson, 2006), MP3000 (Nogueira *et al.*, 2005; Büchner, Nogueira, Edler, Battmer and Lenarz, 2008) and HiRes 120 (Buechner, Brendel, Krüeger, Frohne-Büchner, Nogueira, Edler and Lenarz, 2008) strategies have shown promise in improving speech recognition capabilities of cochlear implantees. The Fine Hearing strategy uses temporal fine structure to improve the perception abilities of the listener, HiRes 120 utilises current steering, and the MP3000 (or psychoacoustic advanced combinational encoder) uses psychometric functions to improve the choices of spectral peaks made by the ACE strategy.

The paragraphs above have outlined the most common speech processing strategies that are currently available. Although there are a number of promising new strategies that are

undergoing clinical trials, one of the most effective and widely-used strategies is the ACE strategy. Considering the substantial market share held by Cochlear Ltd, a model of the ACE strategy implemented on a SPrint speech processor with a Nucleus 24 implant would provide results that are representative of a number of CI users.

2.3.2 Electrical stimulation parameters

Electrical hearing perception differs from individual to individual. Speech recognition and perceptual capabilities provide an indication of the variability in the performance of different individuals (Dorman, Loizou and Rainey, 1997b; Friesen, Shannon, Baskent and Wang, 2001). This variability is not necessarily due to the type of processor or processing strategy, but a multitude of factors can influence the hearing ability of cochlear implantees. Some factors that can influence performance include: stimulation rate (Galvin and Fu, 2005; Vandali, Whitford, Plant and Clark, 2000), insertion depth of the electrode array (Baskent and Shannon, 2005; Dorman, Loizou and Rainey, 1997a) and the electrode stimulation mode (Boëx, De Balthasar, Kós and Pelizzone, 2003).

In terms of insertion depth, performance can be degraded if the electrode array is not fully inserted into the cochlea as information can be lost from those electrodes that may lie outside of the cochlea, which is compounded by the fact that there may be a mismatch between the placement of the electrodes and the frequencies that they provide information for (Baskent and Shannon, 2005). Dorman *et al.* (1997a) found that modelled insertion depths of 22 mm and 23 mm from the base showed a significant degradation in performance, while a depth of 25 mm from the base showed normal performance. Similarly, Faulkner, Rosen and Stanton (2003) found insertion depths of less than 19 mm from the base to cause significant loss of speech information. Therefore, a CI model of multi-electrode stimulation should simulate an electrode array that is placed approximately 25 mm from the cochlea base. Modelling the electrode array closer to the base could result in predictions that are not representative CI listeners' performance.

The common stimulation modes are common ground, monopolar and bipolar stimulation (Clark, 2003), and these modes are all possible with the Nucleus 24 CI (Vandali *et al.*, 2000). These configurations differ in which electrodes act as the electrical ground and which act as the electrical source. Common ground stimulation uses a single electrode while others in the array are connected electronically. Monopolar stimulation is performed by creating a

potential difference between one active electrode and a distant ground situated outside of the cochlea. Bipolar stimulation is performed by creating a potential difference between any two adjacent electrodes in the array, which allows current to flow between the two. Studies have been performed to determine the effects of the different stimulation modes. The main effect is that the modes produce different rates of current decay and therefore different degrees of excitation spread[5] (Nelson, Donaldson and Kreft, 2008; O'Leary, Black and Clark, 1985; O'Leary *et al.*, 1995; Shannon, 1983b), although the perceptual difference this makes is not thought to be significant (Bingabr, Espinoza-Varas and Loizou, 2008; McKay, McDermott and Clark, 1996). Different measuring techniques have been used to determine the current decay rates. In review, Bingabr *et al.* (2008) provide values between 1 dB/mm and 4 dB/mm for monopolar stimulation, and values between 3.5 dB/mm and 9 dB/mm for bipolar stimulation. Based on the data listed above, it can be said that a decay rate of 4 dB/mm is a representative value of bipolar stimulation, but could also be used to represent the decay rate of monopolar stimulation. Therefore, it is reasonable to use this value for the decay rate in a model of multi-electrode stimulation.

2.4 FREQUENCY DISCRIMINATION

The study of frequency discrimination ability in hearing has been a vibrant topic since the early-mid 20th century, when works such as that of Wever (1949) brought to light the possibility of temporal mechanisms being used by the auditory system instead of simply place cues. Since then, much experimental and modelling work has been performed on NH listeners in order to discern the origin of the human ability to perceive small changes in frequency.

Frequency discrimination experiments have been used for many years in auditory neuroscience as a fundamental perceptual measure to gain insight into the functioning of the cochlea and the CANS (Dye and Hafter, 1980; Henning, 1966; Javel and Mott, 1988; Klinge and Klump, 2009; Moore, 1973; Moore, Glasberg, Flanagan and Adams, 2006b; Novitski, Tervaniemi, Huotilainen and Näätänen, 2004; Rosenblith and Stevens, 1953; Sek and Moore, 1995; Shower and Biddulph, 1931; Wier *et al.*, 1977). Frequency discrimination refers to the ability of the listener to perceive a difference in frequency between two succes-

[5]Excitation spread is an unavoidable phenomenon in multichannel CIs, where neural channels (independent nerve fibres of nerve bundles) are stimulated by stray currents in the process of stimulation of nearby neural channels.

sive sounds i.e. the listener's ability to detect frequency changes over time. This is different from frequency selectivity[6], which is the ability to resolve the frequency components of complex sounds[7] (Moore, 2003). Frequency discrimination is assumed to play a role in the perception of speech as well as complex sounds such as music. There have been numerous paradigms in which frequency discrimination has been tested in order to investigate the functional mechanisms of this perceptual characteristic. These include frequency discrimination tests in quiet (Moore, 1973; Sek and Moore, 1995; Wier *et al.*, 1977), frequency discrimination in noise (Dye and Hafter, 1980), frequency modulation detection (Sek and Moore, 1995) and fundamental frequency discrimination (Moore *et al.*, 2006b; Moore, Glasberg and Hopkins, 2006a).

2.4.1 Normal hearing frequency discrimination

The vast majority of experimental work regarding frequency discrimination has been performed on NH subjects. The first tests of this kind were performed by Shower and Biddulph (1931), using a modulation method that is not as reliable as the two-alternative forced choice (2AFC) method commonly in use in psychophysical experiments (Levitt, 1971). Another effective method that is used is the maximum likelihood method (He, Dubno and Mills, 1998). Moore (1973) measured frequency difference limens as a function of duration and frequency and developed a model of their results. They utilised the more reliable 2AFC method. Their work brought to light the commonly recognised bowl shape in the behaviour of Weber fractions[8] for frequency ($\Delta f/f$) measured against frequency (see figure 2.1). They suggested that the relationship between difference limen and frequency is due to a temporal mechanism acting at lower frequencies, while a spectral or place[9] mechanism acts at higher frequencies, with the change occurring somewhere between 4 and 5 kHz. It has been shown by studies such as those of Moore (1973) and Moore (1974) that the human frequency discrimination ability is quite good when compared to other animals (e.g. gerbils: see Klinge and Klump, 2009) and can be as low as 0.2% at 1000 Hz. Hienz, Sachs and Aleszczyk (1993) found that cats are similar to humans in that they probably use a place coding scheme at higher frequencies (also supported by the model of Kim and Parham, 1991), but probably use a different scheme (rate coding) at lower frequencies.

[6] See appendix A.
[7] A complex sound is composed of multiple sinusoidal tones at different frequencies and/or magnitudes.
[8] See appendix A.
[9] See appendix A.

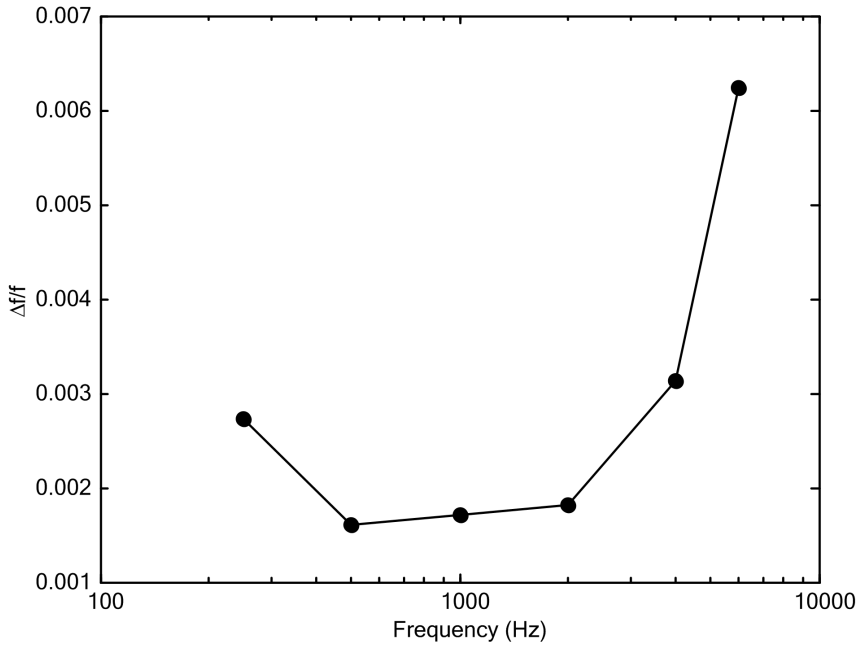

Figure 2.1.
Example showing the bowl shape that can be observed in graphs of $\Delta f/f$ plotted against frequency. It is clear that the discrimination improves until about 500 Hz, where it flattens until 2 kHz, after which it becomes progressively worse. The data shown is taken from Moore (1973).

The work of Wier *et al.* (1977) tested frequency discrimination as a function of frequency and level for a large range of values, in order to test the validity of the theoretical models of frequency discrimination existing at that time, such as the place-code models of Zwicker (1970). Instead of representing the data as a bowl shape as Moore (1973) had done, they fitted their data to straight lines on a log Δf vs $\sqrt{f}$ scale, in an attempt to linearise the data. Although the data could be fitted to linear graphs using that scale, there were nonmonotonicities in their data that made the bowl shape trend difficult to observe. They concluded that the available models did not sufficiently explain the data; specifically, the interaction between frequency discrimination, intensity and duration.

The possibility of a temporal coding mechanism was contradictory to a popular model for frequency modulation discrimination[10] at the time (Zwicker, 1970), which was based on excitation patterns in the cochlea. Frequency modulation discrimination differs from frequency

[10]See appendix A.

discrimination in that for the former task, a difference in two low-rate modulated pure tones must be detected instead of just two pure tones. Zwicker's model proposed that frequency modulation discrimination is based on the slight changes in excitation pattern elicited by the slight changes in frequency (a place model). Although this model is useful for predicting some facets of human frequency modulation discrimination ability, when it is applied to frequency discrimination the model does not account for the fact that frequency difference limens vary more with frequency than equivalent rectangular bands[11] (ERB) (Moore, 2003). The finding that frequency discrimination and bandwidth are not related in this way has been verified numerous times (Moore, 1974; Moore and Glasberg, 1986; Moore and Glasberg, 1989).

Moore and Glasberg (1989) went about investigating the findings of Zwicker's model (Zwicker, 1970) and found that the frequency difference limen/ERB relation varied by as much as a factor of 4 over the range of 500 Hz to 6.5 kHz. They also found that the frequency difference limens cannot only be based on the information from the low-frequency side of the excitation pattern, as implied by Zwicker's model. They verified that frequency modulation difference limens could be predicted using only the low-frequency side of the excitation pattern, but not in the presence of band-pass noise. In this condition, neither a single-band nor a multi-band excitation pattern model was able to account for the behaviour of frequency modulation difference limens. Furthermore, Moore and Glasberg (1989) found evidence that the phase locking could be used in frequency discrimination tasks below 4 kHz.

Sek and Moore (1995) experimented further with the frequency discrimination abilities of human subjects. They performed three different types of frequency difference limen experiments, including frequency discrimination, frequency discrimination of change[12] and frequency modulation discrimination. The frequency discrimination data and frequency discrimination of change data were found to be very similar. Sek and Moore went on to suggest that phase locking is the dominant coding mechanism at frequencies below 4 kHz, with place coding being dominant at frequencies above this.

At the time of Sek and Moore's study, the idea that the temporal aspects of acoustic signals play an important role in frequency discrimination was well established. The pioneering modelling work performed by Siebert (1965; 1970) queried, by use of an analytical model based on SDT, whether frequency discrimination was caused by a place mechanism or by a

[11] See appendix A.
[12] See appendix A

temporal mechanism. Siebert found that by utilising the interspike intervals (ISIs) optimally, frequency discrimination performance superior to human performance could be yielded. He managed to obtain performance comparable to experimental human data by using an average rate-place model, but in both cases the model could only be applied to lower frequencies.

This problem was later addressed by Goldstein and Srulovicz (1977), who included a roll-off in synchronisation of the auditory neurons normally observed between 2 and 3 kHz (Johnson, 1980). Their model predicted the bowl-shaped curve observed in the frequency discrimination data, as well as the magnitudes of frequency discrimination. This suggested that the roll-off to phase locking is responsible for the decrease in frequency discrimination performance seen at higher frequencies. Wakefield and Nelson (1985) extended this model to account for the intensity effects on frequency discrimination, by including the dependence of phase locking on intensity and providing additional constraints in the model. This reinforced the case for temporal coding at lower frequencies.

Hanekom and Krüger (2001) and Hanekom (2000) developed a model that also used the roll-off to phase-locking principle, but applied it using a Kalman filter (Kalman, 1960). The Kalman filter model relies on an internal representation of the signal to be estimated. This optimal approach yielded similar results to those of Sek and Moore (1995) and performed well in predicting the intensity dependence (Wier *et al.*, 1977) and the duration dependence (Moore, 1973) of frequency discrimination. They suggested that the successful implementation of the model could mean that the CANS makes use of a volley principle and a recursive computational mechanism. This processing could be accommodated by feedback through cortico-thalamic loops.

The results of models and experiments such as those mentioned above (and others, such as those of Dye and Hafter (1980), Javel and Mott (1988) and reviews, such as that of Moore (2003)), are consistent with the notion that some sort of temporal mechanism is responsible for frequency discrimination tasks below 4 to 5 kHz, and a place mechanism is responsible for frequencies above this. There are a few studies that have utilised the notion that although a temporal mechanism may be dominant at lower frequencies, place coding could still possibly play a role in the task of frequency discrimination.

Evans (1978) investigated the coding mechanisms of the peripheral auditory system, and based on his data, came to the conclusion that the use of both place and temporal cues cannot be excluded as a possibility in the extraction of frequency information in the cochlea.

Based on physiological evidence, he argued that it is reasonable that both temporal and place information play a role over the frequency between 500 Hz and 5 kHz. He agreed that it is unlikely that temporal mechanisms are responsible for frequency discrimination above 5 kHz.

Recently, frequency discrimination has been measured in a different way from the 2AFC method. Novitski *et al.* (2004) measured the percept electrophysiologically and behaviourally, i.e. using electroencephalograms to measure subjects' neural responses (called event-related potentials) to changing frequencies, in addition to change detection tasks. Their results suggested that the place and temporal coding mechanisms could transition from the one being dominant at lower frequencies to the place mechanism being dominant from approximately 2 kHz.

Srulovicz and Goldstein (1983) developed a model utilising an approach that combined both temporal information and place information. They named this model the central spectrum model, based on the principles of matched filtering. The ISI histogram from each fibre is filtered by a bank of matched filters (for each characteristic frequency (CF) of the fibre), the output of which produces a spectral representation of the input signal. Through probabilistic analysis of the spectral characteristics, the frequency discrimination thresholds could be obtained. The model provided a good prediction of experimental data. Cohen, Grossberg and Wyse (1995) built on the central spectrum idea by developing a model based on neural networks, to successfully model various aspects of pitch perception.

In an attempt to reproduce similar conditions to those in the human auditory system, Heinz *et al.* (2001a; 2001b) developed a comprehensive model of human frequency discrimination based on SDT. They distinguished this model from other available models as an "all-information" model, compared to average rate-place models. Their model took many biological factors into account, including neural adaptation and the roll-off to phase locking at 2 kHz, and it was implemented in a computational, instead of an analytical, manner. This allowed for a time-based representation of each individual processing step in the model, and for a more realistic representation of the signal, as it is unlikely that a single, time-averaged section of APs is used to perform psychophysical tasks (Viemeister and Wakefield, 1991). They also made use of the entire neural representation of the cochlea, instead of breaking the information down into critical bands or smaller units (such as individual neurons) as is done by other models (Goldstein and Srulovicz, 1977; Hanekom, 1999). The model performed better than the human system owing to the optimal nature of the model. In comparison

to Siebert's model (Siebert, 1970), the magnitudes were similar but the trends were better predicted, and the bowl shape was clearly visible.

The review of frequency discrimination provided by this section illustrates that the large body of work surrounding the subject has led to much speculation regarding the mechanisms that cause the results observed in frequency discrimination tests. However, the exact nature of frequency discrimination perception is still unknown. The common idea that timing information plays an important role in frequency discrimination at lower frequencies, while place information plays an important role at higher frequencies is not definitive and the degree to which these mechanisms contribute to the percept is not clear. This ambiguity leaves a gap for a different type of model to be introduced. Considering that Heinz *et al.*'s model (Heinz *et al.*, 2001a; Heinz *et al.*, 2001b) predicted experimental trends well, a similar model that yields less information to the central processor may succeed in predicting the correct magnitudes and trends observed in psychoacoustic data. Such a model could utilise an optimal estimation technique, similar to that which was used by Hanekom (1999) over a limited neural population to measure instantaneous rate. This follows from the suggestion in section 2.2 that an AP-based model be used, as well as from the success of other internal-model based approaches such as those applied by Hanekom and Krüger (2001) and Srulovicz and Goldstein (1983).

2.4.2 Cochlear implant frequency discrimination

Not as much work has been performed on electrical hearing frequency discrimination as on NH frequency discrimination as the former field of study is a newer and can be more difficult to carry out. In CIs, frequency discrimination can be carried out in different paradigms. Discrimination between sinusoidal electrical stimuli (Bilger, Black and Hopkinson, 1977; Helmerich and Edgerton, 1982; Pfingst, 1988) or discrimination between pulsatile electrical stimuli (Bilger *et al.*, 1977; Barretto and Pfingst, 1992) on one or more electrodes can be performed. Signals can be presented directly to the signal processor (Dorman, Smith, Smith and Parkin, 1996; Nobbe, Schleich, Zierhofer and Nopp, 2007) or, less commonly, in sound field conditions (Pretorius and Hanekom, 2008).

Shannon (1983a) carried out basic psychophysical tests on multichannel CI subjects, including tests of threshold and dynamic range, loudness and intensity discrimination, temporal integration and adaptation, frequency resolution and pitch perception. The experiments

were performed using sinusoidal electrical currents instead of the pulsatile stimulation in use today. Shannon (1983a) found that pitch was related monotonically to the position of the electrodes and was also a function of stimulus periodicity below 300 Hz and of stimulus level. Although frequency discrimination and pitch discrimination are not the same, Shannon showed some of the difficulties associated with the discrimination of tones for implantees.

Zeng (2002) investigated the pitch coding mechanisms of the auditory system using cochlear implantees. By using CIs, it was possible to adjust the temporal code independently of the rate code. The findings were that the temporal code is only effective up to about 300 Hz when isolated from the place mechanisms of the cochlea. This suggested that the place code was necessary for pitch perception, along with the temporal code. The implications of this for implants is that more independent electrodes should be used.

Pfingst (1988) studied the similarities between human and macaque monkey frequency discrimination data. The data were compared with acoustic results and it was found that the frequency discrimination magnitudes for sinusoidal electrical stimulation were nearly two orders of magnitude lower than those for acoustic stimulation. Other studies have shown the quantified frequency discrimination abilities of cochlear implantees (e.g. Nobbe *et al.*, 2007; Pretorius and Hanekom, 2008), and when these are compared to NH data, for example by Moore (1973) and Sek and Moore (1995), the performance of implantees is substantially poorer, which is likely due to a lack of numerous independant channels (Zeng, 2002). Barretto and Pfingst (1992) tested the effect of pulse width on monkey subjects and found that when testing at levels where the subjects obtained the minimum percentage correct responses, the discrimination thresholds became larger when increasing the pulse width. They suggested that this was probably a result of loudness cues being eliminated by the change in frequency, but suggested that it could also be due to changes in pitch associated with the change in level or due to the decrease in frequency discrimination performance that normally accompanies a reduction in level.

Frequency discrimination tests in cochlear implantees can provide an indication of speech recognition abilities (Dorman *et al.*, 1996). More recently, frequency discrimination tests have been performed on CI users in order to determine whether “virtual channels” are possible in implants (Nobbe *et al.*, 2007; Pretorius and Hanekom, 2008; Vondrášek, Tichý and Sovka, 2006; McDermott and McKay, 1994; Wilson, Lawson, Zerbi and Finley, 1994), so that existing implants might be better utilised by next-generation signal processors. Virtual channels are created by electric field interactions between electrodes during simultaneous or

non-simultaneous stimulation, and these activate different populations of neurons to those that are activated by individual electrodes.

Testing in sound field conditions can provide slightly more realistic listening conditions than tests supplying signals directly to the electrodes or to the processor. This is because in direct-stimulation experiments, the contribution of the CI processor to perception is bypassed, and removes any potentially complex interactions that the processor may contribute to the perception task (Rogers *et al.*, 2006). A sound field experiment by Rogers *et al.* with CI listeners tested intensity discrimination and formant frequency discrimination. They found that intensity discrimination thresholds in their subjects where higher (some by a factor of 2.4) than those of NH listeners. This is in contrast to intensity discrimination tests performed at the electrode level that yield very fine intensity discrimination abilities in CI listeners (e.g. Nelson, Schmitz, Donaldson, Viemeister and Javel, 1996). This emphasised the difference that could be present when performing psychoacoustic experiments with direct electrode activation, when compared with normal listening conditions that cochlear implantees experience in everyday life. Pretorius and Hanekom (2008) studied the frequency discrimination abilities of CI users in the sound field. They found that under these conditions, CI users where able to detect a finer pitch than expected by near-simultaneous activation of electrodes. They suggested that this finding might lead to improved music processing strategies, if processing strategies could be improved to take advantage of the phenomenon.

Although there are a multitude of models for NH frequency discrimination, there appears to be few models of electrical psychoacoustics, and electrical hearing frequency discrimination in particular. Hanekom (2000) attempted to model sinusoidal electrical stimulation using a Kalman filter and a temporal code (utilising ISI times). The model predicted that CI frequency discrimination thresholds should be larger than NH thresholds, but it underestimated the values and slope of the decrease in performance above 600 Hz. A pulsatile electrical stimulation-based model was developed in a series of articles by Xu and Collins (2004; 2005; 2007), in which frequency modulation discrimination thresholds were predicted, and the trends obtained were similar to those found in literature.

A model based on the principles outlined in sections 2.2 and 2.4.1, that is able to predict the outcomes detailed in this section, could provide insight into what information is needed by CI listeners to perceive speech and music better. Thus, improvements to the design of CIs could be identified.

2.5 INTENSITY DISCRIMINATION

Intensity discrimination is a measure of the ability to tell the difference between the loudness of two sounds, that is, the smallest detectable change in intensity. Along with frequency discrimination, it is one of the fundamental perceptual measures. It is important as it can provide information about the direction from which a sound originates, the speed that a sound is travelling and the direction in which it is travelling, and in the case of speech, it can provide prosodic cues associated with tone, which is particularly important for tonal languages (Drennan and Pfingst, 2005).

2.5.1 Normal hearing intensity discrimination

Although intensity discrimination can be measured in a few different ways, namely by: amplitude modulation detection (Riesz, 1928), increment detection (Miller, 1947; Viemeister and Bacon, 1988), discrimination of consecutive pulses of stimuli (Florentine, 1983; Florentine, 1986; Florentine *et al.*, 1987; Jesteadt, Wier and Green, 1977; Viemeister, 1972) or maximum-likelihood procedures (He *et al.*, 1998)[13], the general trend observed in the results is the same.

When intensity discrimination is measured with wide-band noise or band-pass filtered noise, the trend follows Weber's law[14] approximately. This is not so when it is measured using pure tones (Florentine, 1983; Florentine *et al.*, 1987; Viemeister, 1972), where the performance tends to improve at higher frequencies. If the intensity difference limen is plotted against intensity (both scales in dB), the gradient of the resulting plot is approximately 0.9, while Weber's law predicts that this should be approximately 1.0 (Moore, 2003). The inability of Weber's law to account for the trends in intensity discrimination data is know as the "near miss" to Weber's law and continues up to about 100 dB SPL. Depending on sensation level, intensity discrimination can be between 0.3 dB and 1.5 dB. In comparison to other species, humans perform relatively well in intensity discrimination tasks (Sinnott, Brown and Brown, 1992).

Miller (1947) tested intensity discrimination in order to investigate its relation to loudness

[13] See also appendix A.
[14] See appendix A

and masking. He found that a just-noticeable difference (JND) at a low intensity produced a much smaller change in apparent loudness than a JND at a high intensity. As he tested this using wide-band noise, his data were consistent with Weber's law.

In order to illustrate the near miss to Weber's law, Moore and Raab (1974) used noise specifically designed to disrupt aural harmonics. They wished to test that the trend found by studies such as that of Miller (1947), i.e. that intensity discrimination is better at higher levels, was due to the fact that the auditory system made use of harmonics (suggested by the modelling study by Viemeister, 1972). They found that this was not the case, and favoured a pattern of excitation explanation similar to that used by the model of Zwicker (1970), who described the phenomenon in his investigation of modulation detection. Zwicker suggested that changes in the excitation pattern could be detected and used as a cue for intensity changes. As stimulus intensity increases, sinusoidally induced excitation patterns grow nonlinearly close to the centre, but more linearly on the high frequency side of the pattern, and thus the rate of the growth at the high frequency side is greater (Moore, 2003).

Another contributing factor to the near miss was proposed in a model by Florentine and Buus (1981). They suggested that as the stimulus increases, the excitation pattern and therefore the number of active channels grows. The auditory system utilises these newly available channels, providing more information to the CANS; thus, the listener can detect smaller changes in intensity. Buus (1990) tested intensity discrimination with narrow-band and frozen noise[15], and found that such stimuli are also subject to the near miss to Weber's law and therefore favoured the explanation provided by Florentine and Buus (1981). Results from the series of articles by Florentine and colleagues also agreed with much of the data from that multi-band excitation pattern model (Florentine, 1983; Florentine, 1986; Florentine *et al.*, 1987). A simple neural counting model was proposed by Teich and Lachs (1979), which utilised the sum of the neural counts on a population of neurons as an input to a decision device that compared the total number of neural spikes in the population for two stimuli of different intensities and decided which had more spikes. The model managed to account for the near miss. This evidence pointed to a simpler, rate-based explanation for intensity coding.

Recently, there have been models that place more emphasis on the central mechanisms involved with the task of intensity discrimination and assign new levels of complexity to the

[15]This refers to the fact that the same sample of noise was used in all of the experimental trials that were performed.

mechanism that performs this task. Heinz, Colburn and Carney (2001c) used monaural cross-frequency coincidence detection. Their model managed to predict the near miss to Weber's law, as well as a nonmonotonic irregularity at high frequencies. They concluded that monaural coincidence detection is a physiologically realistic mechanism owing to its ability to predict intensity discrimination data over the human dynamic range.

Jepsen *et al.* (2008) constructed a comprehensive model to be applied to many psychophysical experiments, mainly involving masking, but intensity discrimination was one of the psychophysical measures that they investigated. The model featured many aspects of the auditory system, such as nonlinear basilar membrane processing, hair cell transduction, adaptation and a band-pass modulation filterbank (from the original model by Dau *et al.*, 1996a). This version of the model did not manage to predict the results as well as the intensity discrimination data generated by the model of Dau, Kollmeier and Kohlrausch (1997a; 1997b), which managed to predict the near miss to Webers law and the approximate magnitudes measured by Houtsma, Durlach and Braida (1980). Jepsen *et al.* (2008) attributed this to the inclusion of basilar membrane compression in their model.

SDT, in combination with an auditory image model, was used by Huettel and Collins (2002) to simulate the effects of hearing loss on intensity discrimination. They tested the hypothesis that near-normal performance obtained by hearing-impaired listeners would be related to either a normal spread of excitation (Schroder, Viemeister and Nelson, 1994) or a greater level of intensity discrimination. They concluded that the near-normal performance was due to a normal spread of excitation.

Considering the literature cited above, it seems that intensity discrimination is largely dependent on the excitation pattern, although there are many physiological factors that contribute to this percept. A model that is to predict intensity discrimination thresholds should therefore provide a reasonable representation of the excitation pattern to the central observer. A neural population model could provide the necessary information.

2.5.2 Cochlear implant intensity discrimination

Intensity discrimination as measured in electrical hearing is quite different from that of acoustic hearing as it involves differentiating between the intensities of electrical stimuli. In the case of pulsatile electrical stimuli, this implies that the intensity can be changed in

two different ways, namely by changing the current magnitude or the width (duration) of the electrical pulse. In either case, the quantity of charge that stimulates the nerves is changed, and the smallest noticeable difference in charge would be considered as the intensity discrimination threshold in electrical hearing (Drennan and Pfingst, 2005).

CI listeners have a much finer intensity discrimination ability than NH listeners, but owing to their dynamic range being much smaller, they still have fewer discrimination steps than NH listeners (Shannon, 1983a). According to an early study by Shannon (1983a), cochlear implantees were only able to achieve approximately 20 discriminable steps, as opposed to the 60 to 100 discriminable steps of NH listeners. Other studies on current intensity discrimination have been performed by Dillier, Spillmann and Guntensperger (1983), Drennan and Pfingst, (2005; 2006), House and Edgerton (1982) and Nelson *et al.* (1996). In some subjects in these studies, a difference limen of intensity of less than 1% was observed, while a difference limen of up to 16% was observed in others that did not perform quite as well.

Nelson *et al.* (1996) found that when the decibel representation of the Weber fractions of current intensity was measured against the sensation level, the result was a power function. This was shown to be the case for acoustic hearing as well. A comparison between the average exponents of the two types of power function showed that the exponents for electrical hearing were one order of magnitude higher than those for acoustic hearing. The difference was attributed to the lack of compressive preprocessing of pathological cochleae.

Bipolar and monopolar stimulation were studied by Drennan and Pfingst (2005) in order to determine if either stimulation mode provided an advantage in intensity discrimination. They did not find any significant difference in the discrimination thresholds of the two stimulation modes, and similarly to Nelson *et al.* (1996), they found that Weber fractions decreased with increasing current level. They did suggest, however, that nerve survival contributed to the subjects' improved performance at low levels in a specific case when bipolar stimulation was used. They proposed that the cochlear array was closer to a surviving nerve bundle that experienced a greater rate of spike count increase and nerve fibre recruitment, resulting in smaller increases in current producing a perceptual difference. The follow-up study (Drennan and Pfingst, 2006) investigated more parameters, namely the number of stimulated electrodes, the stimulation rate and the distance between electrodes. In stimulating multiple electrodes, they found that discrimination sensitivity to intensity on a single electrode decreased as more electrodes were added. This appeared to be due to the reduction in overall loudness contribution by the single electrode as more electrodes were added. The results of testing stimulation

rate and electrode distance showed that most of the subjects had the greatest sensitivity at the same pulse rate that they used for everyday use in their implant. It was unclear whether this was because they were used to the pulse rate or because they had chosen a pulse rate during clinical testing that was inherently optimal.

Galvin and Fu (2005) showed that a high stimulation rate in cochlear implantees improved their dynamic range. But later, Galvin and Fu (2009) also tested the effects of high rates of stimulation on a listener's performance in intensity discrimination and modulation detection tasks. The results showed that the high rates had no effect on intensity discrimination ability, as the number of discriminable steps did not increase. The higher rates negatively affected modulation detection. They proposed that these effects were a result of the slower loudness growth associated with higher stimulation rates.

Other than testing the effects that different electrical parameters have on intensity discrimination, the electrically stimulated ear can be used to learn about the psychophysical properties of the auditory system (Zeng and Shannon, 1999). This can be done without the effects of the peripheral hearing playing a role, as this is bypassed with a CI. Zeng and Shannon (1999) used intensity discrimination data to investigate the roles played by the peripheral and CANS in the relationship between stimulus and perceived stimulus magnitude. They concluded that Weber's law in the stimulus domain is due to logarithmic compression at the periphery, but in the perception domain it is due to an exponential expansion performed in the CANS, the periphery being bypassed by electrical hearing. They developed a simple model based on these principles, which managed to predict JNDs in intensity adequately.

Bruce *et al.* (1999b) used their model of electrical neural firing in combination with SDT to test whether the inclusion of stochastic activity would improve the predictions made by deterministic models of electrical intensity discrimination. They found that the stochastic model provided far better predictions of all the psychoacoustical percepts that were tested, including intensity discrimination. Xu and Collins (2005) expanded on this particular model to test whether the addition of noise to such a model would improve performance in certain psychophysical tasks, as has been suggested by studies such as that by Zeng, Fu and Morse (2000). Xu and Collins (2005) found that noise did significantly improve the performance in the tasks, and in fact degraded intensity discrimination performance. These results have further been validated by Shepherd and Hautus (2009), who suggest that stochastic resonance[16] in psychophysical tasks may be based on decision processes rather than on sensory

[16]The phenomenon whereby signal detection in a nonlinear system is improved by the addition of noise.

enhancement.

A model of CI intensity discrimination at the electrode level would be expected to provide intensity difference limen predictions that are superior to NH listeners. However, the work by Rogers *et al.* (2006) mentioned in section 2.4.2, and the work by Drennan and Pfingst (2006) suggest that a model that simulates sound field conditions might yield results that are inferior to NH results. It appears that an AP-based model would be suitable for predicting the outcomes of intensity discrimination experiments, as similar experiments have been carried out before by Bruce *et al.* (1999b) and Xu and Collins (2005). Additionally, it seems that low stimulation rates and bipolar stimulation should not affect the results drastically.

2.6 SUMMARY

In this chapter, a background on previous studies relating to this work was presented. Based on the existing models and the principles presented in this chapter, a psychoacoustic model of frequency and intensity discrimination is developed in the following chapter. This model provides a contribution in that it can predict the outcomes of both NH and CI percepts, and the modelled outcomes will be compared in later chapters with data presented in some of the studies that have been mentioned. The full method describing the model's development follows in chapter 3.

CHAPTER 3

METHODS

3.1 CHAPTER OBJECTIVES

This chapter presents a full description of the model used in this study. An overview of the entire model is provided first, in order to elucidate its structure and to facilitate an understanding of the detailed descriptions provided further on in the chapter. Justifications for the assumptions made in the structure of the model are listed here. After the overview, two of the elements of the model that are common to both acoustic and electrical hearing are described; namely, the input signal and the rate vs intensity model. Following this is the model of acoustic cochlear stimulation. This part of the model is simpler and should be more familiar to most than the model of electrical stimulation. The model of the CANS (which is also common to both hearing modalities) is detailed after this. This model consists of the rate estimator and the decoders used in this study. Finally, the model of the CI is described, followed by the model of electrical hearing.

3.2 OVERVIEW OF THE MODEL FOR FREQUENCY AND INTENSITY DISCRIMINATION

As was discussed in section 1, the central idea behind this dissertation was to develop a perceptual model that is able to predict both intensity discrimination and frequency discrimination data for both NH listeners and CI listeners. The main criteria in developing this model was that it should have an interchangeable input, but possess the same basic structure for the two different types of listeners.

The current level of understanding of the auditory system's functionality is incomplete. If science offered a complete theory of the auditory system, no new models of the auditory system would be required, and far superior CIs would conceivably be available. As a result of the incomplete description of the auditory system that is available, certain assumptions need to be made in order to realise an interchangeable model of perception. The first and most fundamental of these assumptions is that the CANS must be "blind" to the peripheral input process. This is to allow for the interchange of stimulation models between an acoustical stimulation model and an electrical stimulation model. In order to make the two stimulation types readily comparable, the central processing mechanism needs to be the same for both. It is therefore necessary that both the acoustical and electrical stimulation models yield the same type of output.

The "blind" central processor assumption is not entirely biologically accurate, as there are active feedback mechanisms present between the outer hair cells and the basilar membrane (Nobili, Mammano and Ashmore, 1998; Ruggero and Rich, 1991), but the assumption is necessary for the "black box" approach that was used with regard to the CANS in the development of the model. This assumption is consistent with most other perceptual models that do not take feedback to the auditory periphery into account (e.g. Hanekom and Krüger, 2001; Heinz *et al.*, 2001b; Erell, 1988; Goldstein and Srulovicz, 1977; Siebert, 1970), although some more recent models have taken certain feedback mechanisms into account (Jepsen *et al.*, 2008; Dau *et al.*, 1996a). Thus, the model was developed with the general structure as depicted in figure 3.1. A more detailed representation of the model is shown in figure 3.2.

With reference to figure 3.1, a pure tone serves as input to the model of either acoustic or electrical stimulation, depending on the mode of operation that is required. The stimulation model generates a space-time AP output pattern for a population of neurons. The space-time

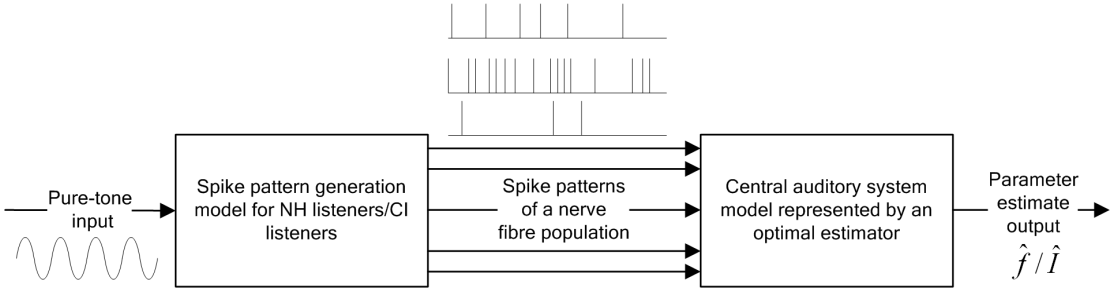

Figure 3.1.
The general representation of the model that was developed.

AP patterns have the same format for both hearing modes: they provide the firing times of every neural spike on each nerve channel. Owing to this feature, the input to the CANS model is interchangeable. The model of the CANS receives the AP patterns and predicts the discrimination threshold by means of an optimal estimation mechanism.

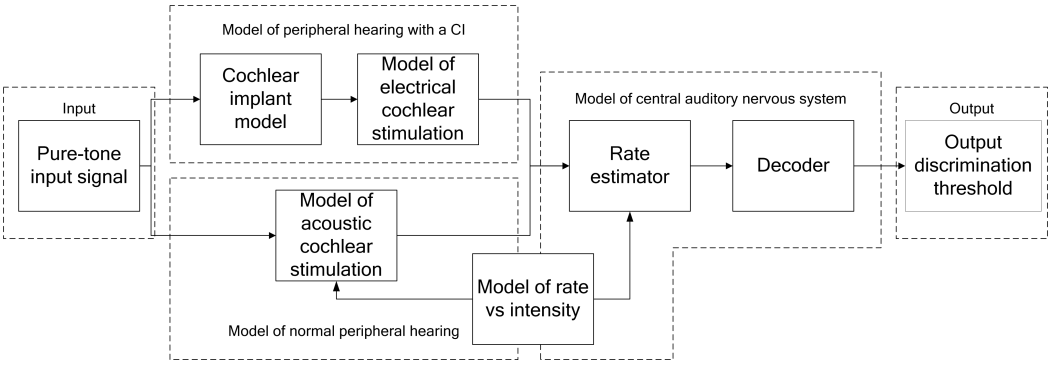

Figure 3.2.
A diagrammatic representation of the model. The components of the model that are discussed in section 3.3 are indicated by the solid blocks. The dashed lines outline the conceptual groups of the modules, as shown in figure 3.1.

Chapter 2 describes some of the auditory perceptual models that appear in literature. These models can be broadly grouped into a few classes, which include:

- rate-place models, which utilise the average rate of the nerve firings as well as the

tonotopic location of the nerve fibres (e.g. Siebert, 1970),

- temporal models, which make use of the timing information provided by the phase-locked neural firings to pure-tone input signals (e.g. Hanekom and Krüger, 2001) and
- all-information models which utilise all the information available, including the average rate, place information and temporal information (e.g. Heinz *et al.*, 2001b).

The three groups of models mentioned (and specifically the examples cited) typically make use of optimal detectors to make predictions from neural output, as described in chapter 2. The model described in this study could be broadly identified with the category of rate-place models, except that the model uses instantaneous rate information along with place information, instead of simply an average rate. This is a similar concept to the spatiotemporal models described in chapter 2. The use of instantaneous rate information is thought to increase the biological plausibility relative to average rate models as it is unlikely that the auditory system only uses average rate information over a period longer than 5 to 10 ms while processing auditory input (Viemeister and Wakefield, 1991). The mechanics of the processing are described later in section 3.3.4. For the purposes of differentiation, the type of model used in this study will be referred to as an instantaneous rate-place model, as opposed to an average rate-place model.

3.2.1 Justification for the model structure

Much of the structure and design of the model of this study was necessitated by the requirements set out in chapter 1. In order to create a model that is interchangeable between two hearing modalities, the central processing mechanism is required to be flexible. The models of peripheral processing need to produce similar types of output and should be roughly equivalent in simplicity in order that information included in one of the models is not included in the other – such a situation might affect the results, and comparisons between the two modalities would potentially be made more difficult. A further advantage of a model of this type is that it can be expanded in future studies owing to its modular nature: the peripheral processing sections can be upgraded to include more complex neural behaviour, such as neural adaptation, or alternatively, the central processor can be altered entirely. This provides a framework in which expansion can facilitate the testing of other perceptual parameters, in-

cluding masking or gap detection. Thus, many hypotheses of central or peripheral processing may be tested.

One of the more successful models of frequency and intensity discrimination, by Heinz *et al.* (2001a; 2001b), utilised SDT. This computational model described the instantaneous rate of auditory neural firings by performing various stages of signal processing on the input signal. This description of the instantaneous rate on each neuron allowed stochastic descriptions of the discharge times to be formulated, to which SDT could be applied and from which discrimination thresholds could be predicted. Unfortunately, no such stochastic descriptions of neural firing are available for the electrically stimulated cochlea in the sound field, which made the model unsuitable for use as a basis for this study. The central processing model proposed by Hanekom (1999), based on optimal estimation techniques, provided a versatile basis on which to construct the model as the inputs to that model could be any population of nerve fibre outputs, including those created by an electrical stimulus.

Although much of the literature suggests that a phase-locking or other temporal mechanism is used by the CANS in the discrimination of frequency at lower frequencies (refer to chapter 2), instantaneous rate-place models have been the focus of only limited study, but have been seen to account for some frequency discrimination trends (see Hanekom, 1999). The fact that the model developed here makes use of instantaneous rate provides a more biologically plausible mechanism than average rate-place models, as it accounts for the effect that each spike has at each point in time, because the instantaneous rate that is estimated by the central processor is dependent on the timing of each individual spike. In this way, temporal information is taken into account. Additionally, this model was designed to account for both frequency and intensity discrimination thresholds, and in the latter, excitation pattern models (place models) have been found to be rather successful. Details of the implementation are discussed in section 3.3.

3.3 DETAILS OF THE IMPLEMENTATION OF EACH MODEL SECTION

In this section, each component of the implemented model is described in detail. Certain minor aspects of the implemented model were not used in the final results, and this is noted where applicable.

The input to the model is a pure-tone signal. The model operates in two different modes, one for CI listeners and one for NH listeners. The difference between the two modes lies in the model of peripheral hearing. Both modes use the same rate estimator and decoder. The output of the decoder is the discrimination threshold of the mode currently in use.

3.3.1 Input signal

The input to the model, used in all simulations (unless otherwise indicated) was a 200 ms pure-tone signal. A pure tone was used for ease of comparison with data on frequency discrimination in the literature (such as that of Sek and Moore, 1995; Moore and Glasberg, 1989; Moore, 1973) for NH listeners. A 500 ms pure tone was used for CI listeners so that comparison with Pretorius and Hanekom (2008), who tested CI listeners' frequency discrimination abilities in the sound field, was possible. The input intensity of the signal is represented in terms of dB SPL, which refers to the decibels of sound pressure level. This measurement refers to the level in dB *re* 20 μPa. Therefore, in order to transform an intensity in Pascals to a measurement in dB SPL, equation 3.1 is used:

$$p_{\text{dB SPL}} = 20\log_{10}\left(\frac{p_{\text{RMS,Pa}}}{20\times 10^{-6}}\right) . \tag{3.1}$$

3.3.2 Model of rate vs intensity

The model of rate vs intensity is largely based on two key functional mechanisms: a model of the RI function[1] (Yates, 1990) and a model of the tuning curve[2] (Colburn, 1973), both of which are simple models. The RI curve has been studied extensively in literature and many models of RI functions exist (Young and Calhoun, 2005; Köppl and Yates, 1999; Yates, 1990; Yates, Winter and Robertson, 1990; Sachs, Winslow and Sokolowski, 1989; Young and Barta, 1986; Sachs and Abbas, 1974; Kiang, 1965). A model of rate vs intensity was necessary in this model in order for the input intensity to be translated into a firing rate on each nerve channel. Owing to the fact that Colburn's tuning curves are included (Colburn, 1973), the firing rate is frequency dependent. The translation of intensity to rate is necessary to generate the AP patterns for the acoustic model, and to provide a reference for the rate

[1] See appendix A for more detail
[2] See appendix A for more detail

estimator. The concept of the rate vs intensity model is illustrated in figure 3.3.

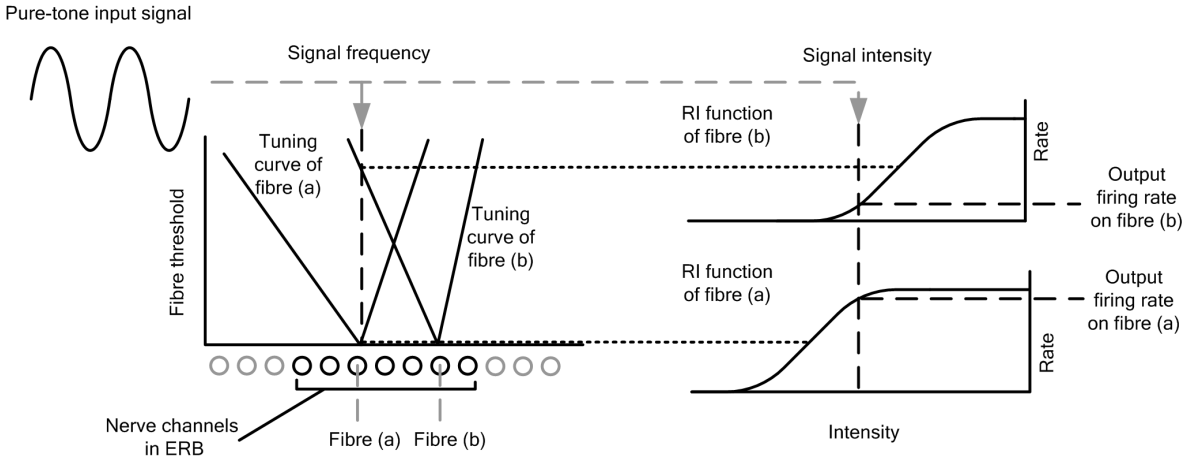

Figure 3.3.
The concept of the rate vs intensity model.

In figure 3.3, the frequency of the input signal determines the threshold of each fibre by means of a tuning curve. This threshold is applied to the RI function. The intensity of the signal is then used to determine the output firing rate of the given fibre from the RI function. More detail is provided in the sections below.

3.3.2.1 Quantitative description of rate-intensity functions

A full description of the mathematics of RI functions is presented in this section. This is to provide a complete picture of the model that was used in this study. All three types of RI functions are described here and were implemented for illustrative purposes, but only flat-saturating RI functions were used to obtain the results in chapters 4 and 5.

According to Köppl and Yates (1999), RI functions can be classified into three different groups: flat-saturating, sloping-saturating and straight. They are named for their behaviour at charasteristic frequency. These three curve types are best described visually, and are shown in figure 3.4.

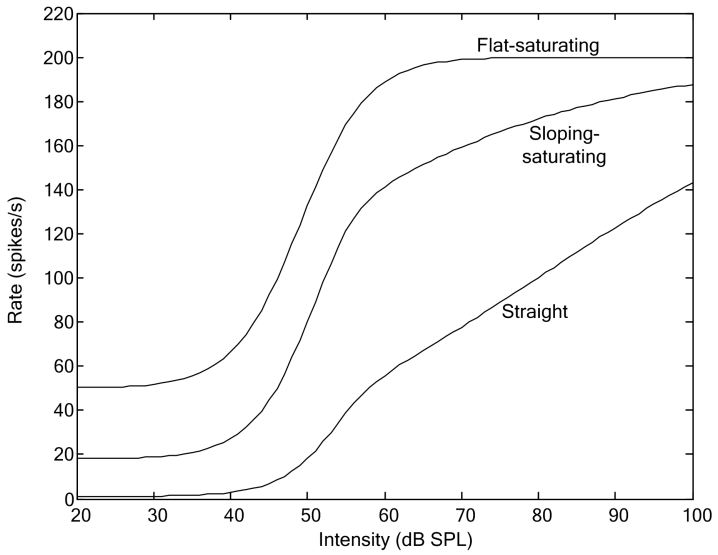

Figure 3.4.
The three different RI curve types: flat-saturating, sloping-saturating and straight, with spontaneous rates of 50, 18 and 1 spikes per second respectively. All three have the same maximum rate of 200 spikes per second and the same threshold of 49.1 dBSPL.

Each of the curves has been described quantitatively by Yates (1990); equation 3.2 describes a flat-saturating RI function:

$$R = A_0 + \frac{A_1 \cdot p^2}{{A_2}^2 + p^2} \ . \tag{3.2}$$

R is the nerve firing rate (spikes/s), A_0 is the spontaneous rate (spikes/s), A_1 is the maximum firing rate – the spontaneous rate (spikes/s), A_2 is the threshold of the fibre (Pa) and p is the input sound pressure level (Pa). For the purpose of this study, the threshold was defined as the point where the firing rate is at a value of 50% of the difference between the maximum firing rate and the spontaneous rate. This allowed a level of consistency with the model of electrical stimulation used (see Bruce *et al.*, 1999c; Bruce *et al.*, 1999a), but deviates from the more commonly used threshold definition of a firing rate 10% to 20% above spontaneous rate.

The sloping-saturating and straight curves are characterised differently from the flat-saturating curves. It is clear from figure 3.4 that these curves saturate very slowly; this is due to the non-

linearity in the basilar membrane displacement, which affects the shape of the RI functions (Yates *et al.*, 1990). This is illustrated in a sketch in figure 3.5. Flat-saturating RI functions do not operate in the nonlinear region of basilar membrane displacement.

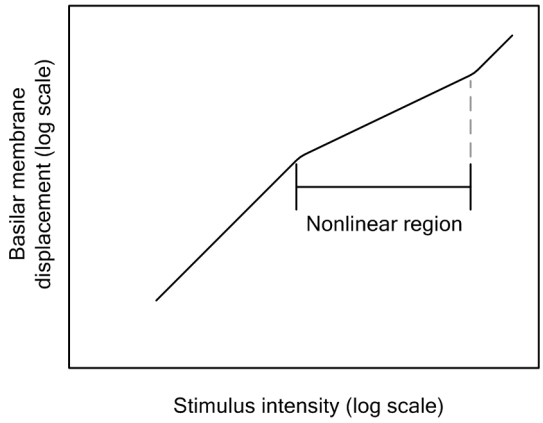

Figure 3.5.
Illustration of the nonlinear behaviour of basilar membrane motion when plotted against input stimulus intensity. Both axes possess a logarithmic scale. Adapted from Yates (1990).

In the case of sloping-saturating and straight fibre curves, the equations are slightly different. The rate (R) is defined by equation 3.3:

$$R = A_0 + \frac{A_1 \cdot d^2}{{A_2}^2 + d^2} \ , \tag{3.3}$$

where R is the nerve firing rate (spikes/s), A_0 is the spontaneous rate (spikes/s), A_1 is the maximum firing rate minus the spontaneous rate (spikes/s), A_2 is the value of d that produces half of $A_1 + A_0$ (unitless) and d is the displacement of the basilar membrane at the given point (unitless), described by equation 3.4. The equation describing the basilar membrane displacement d is

$$d = \left[\frac{A_3^{(1/A_4-1)} \cdot p^{(1/A_4)}}{A_3^{1/A_4-1} + p^{(1/A_4-1)}} \right]^{A_4} \ , \tag{3.4}$$

where d is the displacement of the basilar membrane at the given point (unitless), A_3 is the sound pressure level, where the basilar membrane changes to nonlinear behaviour (Pa), A_4 is the slope of the nonlinear region of basilar membrane displacement (dB/dB) and p is the input sound pressure level (Pa).

3.3.2.2 Changing parameters of the rate-intensity functions

The parameters of the RI functions are not constant over all frequencies. The curves are named for their behaviour at their CFs, but the curves change in type as the stimulating frequency moves further away from the fibre's CF. High spontaneous rate (above 18 spikes/s), low threshold fibres show flat-saturating behaviour for frequencies up to CF, after which they take on sloping-saturating and eventually straight properties. Sloping-saturating fibres exhibit flat-saturating behaviour almost up to CF, sloping-saturating behaviour around CF and straight behaviour thereafter. Straight fibres only show flat-saturating behaviour well before CF, which eventually becomes sloping-saturating behaviour just below CF and then straight curves at and above CF (Yates, 1990).

Parameters A_0 and A_4 – spontaneous rate and slope. The two parameters A_0 (the spontaneous rate of the fibre) and A_4 (the power-law exponent of the nonlinear region of basilar membrane displacement or the the slope of the nonlinear region as a function of input stimuli) are both constant parameters across frequency (Yates, 1990; Yates *et al.*, 1990; Köppl and Yates, 1999). Although A_0 can vary between different fibres, the value of A_4 remains constant for most fibres, and according to Yates *et al.*, this value is 0.2.

Parameter A_1 – firing rate range. It is clear from literature such as the work of Sachs and Abbas (1974) and Evans (1978) that at frequencies higher than CF, the maximum firing rate begins to decrease, and thus the parameter A_1 decreases too, because spontaneous rate remains constant. There does not appear to be any data available quantifying the change in maximum firing rate in literature. It was therefore necessary to use the available data to model the apparent changes. This modelling was performed by making rough approximations to the data. This was considered to be acceptable owing to the largely variable nature that RI functions tend to have from nerve to nerve. Data was gathered from Sachs and Abbas (1974), who measured the RI functions of many different fibres and provided a fair amount of data from which modelling could be performed. Specifically, the percentage decrease in frequency (after CF) was measured and plotted against the percentage decrease in rate. A curve fit was made to the data with a linear function for the sake of simplicity[3]. An optimisation procedure was performed in Matlab[4] to provide a least-squares fit to the provided data.

[3] See appendix B for the data.
[4] The curve fit was performed using the function "lsqcurvefit" in Matlab R2007b.

The equation to which the data was fitted is provided in equation 3.5:

$$\Delta r = m \cdot \Delta f \ , \tag{3.5}$$

where Δr is the percentage decrease in the maximum rate, Δf is the percentage difference in frequency and m is the free parameter to which a least-squares fit was performed. The data that was used from Sachs and Abbas (1974) is shown as a scatter graph in figure 3.6, along with the fitted linear function, where the free parameter was found to be equal to 1.1.

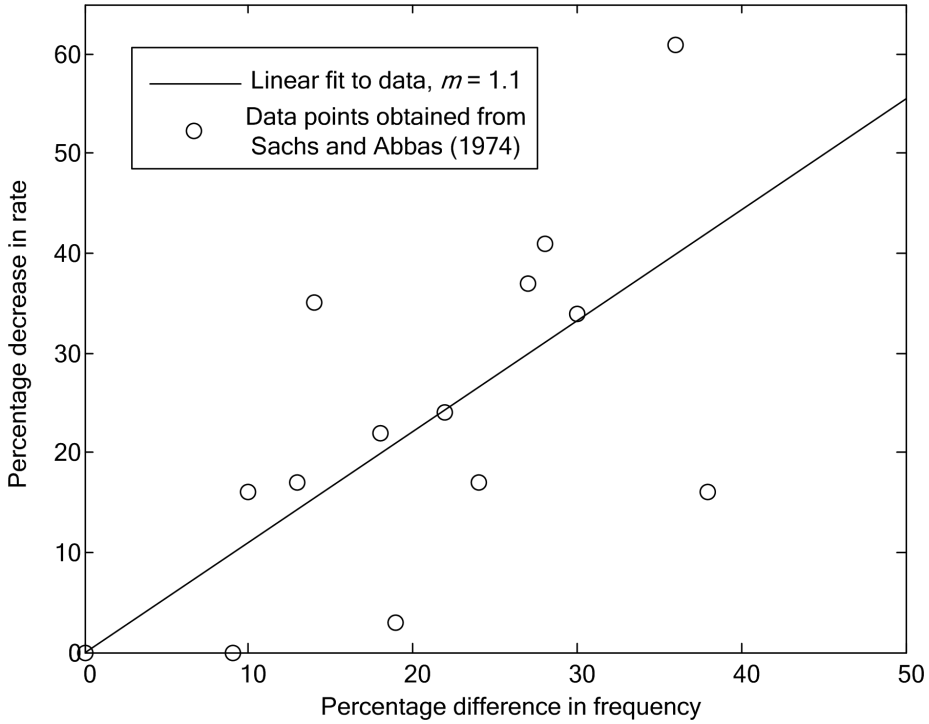

Figure 3.6.
Data from Sachs and Abbas (1974) on a scatter diagram, with the function $\Delta r = m \cdot \Delta f$ fitted to it by least-squares optimisation.

Parameter A_2 – threshold. Much research has been performed in the study of tuning curves (e.g. Kiang, 1965; Kluk and Moore, 2004; LePage, 1987), and Colburn (1973) developed a simple model of tuning based on the experimental results of Kiang (1965). This model provides the threshold of the fibre at a given frequency as a function of its CF and an α parameter, which determines the quality factor or Q of the tuning curve. The tuning curve

is essentially represented as a filter (equations 3.6 and 3.7), where the value of the transfer function of the filter is given by H_m:

$$H_m(f) = \begin{cases} \left(\frac{f}{f_{CF}}\right)^{\alpha(f_{CF})}, & \text{for } f \leq f_{CF} \\ \left(\frac{f_{CF}}{f}\right)^{2\alpha(f_{CF})}, & \text{for } f > f_{CF}\ , \end{cases} \tag{3.6}$$

where α is given by

$$\alpha(f_{CF}) = \begin{cases} \alpha_0, & \text{for } f \leq f_{CF} \\ \alpha_0 \cdot \frac{f_{CF}}{f_T}, & \text{for } f > f_{CF}\ , \end{cases} \tag{3.7}$$

and f is the stimulating frequency (Hz), f_{CF} is the CF of the fibre (Hz) and f_T is the frequency at which the Q factor of the filter increases (Hz). In order to determine the threshold value at a specific frequency, the value of A_2 is then divided by H_m in order to obtain the new threshold value. Examples of the tuning curve shapes can be observed in figure 3.7.

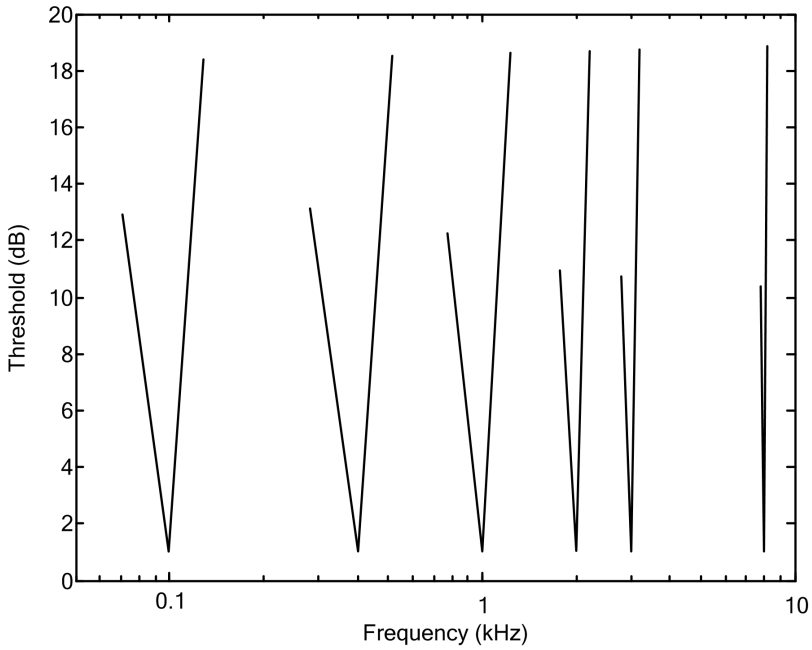

Figure 3.7.
Example tuning curves generated using equations 3.6 and 3.7. The curves were generated with α_0 = 4 and f_T = 800 Hz, for CFs of 100 Hz, 400 Hz, 1000 Hz, 2000 Hz, 3000 Hz and 8000 Hz.

The tuning curves in figure 3.7 show how the width of the filter stays constant below f_T (800 Hz in this case), and sharpens beyond that point. This is consistent with data showing an increase in tuning at higher frequencies (Liberman, 1978).

Parameter A_3 – the nonlinearity breakpoint. There is a small amount of data relating the breakpoint parameter (A_3) to frequency; Köppl and Yates (1999) measured this parameter in barn owls and found there to be an approximately linear relationship between the threshold at CF of a fibre and the breakpoint parameter at CF. Other than this, there does not appear to be any data relating A_3 to frequency. In order to fill this gap, data was derived from RI functions in literature. Again, this was deemed to be an acceptable approximation as a result of the variability in the RI functions of different nerve fibres. The breakpoint at the CF was deemed to be constant at all frequencies above characteristic which is consistent with the model of Sachs and Abbas (1974), who chose a value of 73 dB. From the data obtained in literature (Evans, 1978; Sachs and Abbas, 1974) the average value found was within a similar region (about 76 dB)[5]. This assumption does not hold for all types of RI function, but would hold for flat-saturating curves particularly (see section 3.3.2.2).

3.3.2.3 Resulting rate-intensity functions

The model of rate vs intensity that was developed provides a platform from which modelled RI functions can be obtained. For the purposes of the overall model, the RI functions were restricted to flat-saturating curves that only acted in the linear region of basilar membrane displacement (refer to figure 3.5). This means that the breakpoint parameter was not included in the applicable sections of the model; only equations 3.2, 3.5, 3.6 and 3.7 were used. This assumption was used for the sake of simplicity and so that some of the weaker assumptions (such as the assumption made in paragraph regarding the nonlinearity breakpoint) did not affect the model. Figure 3.8 shows the differences between the resulting RI functions, i.e. with and without the basilar membrane nonlinearity included.

[5] See appendix B for details.

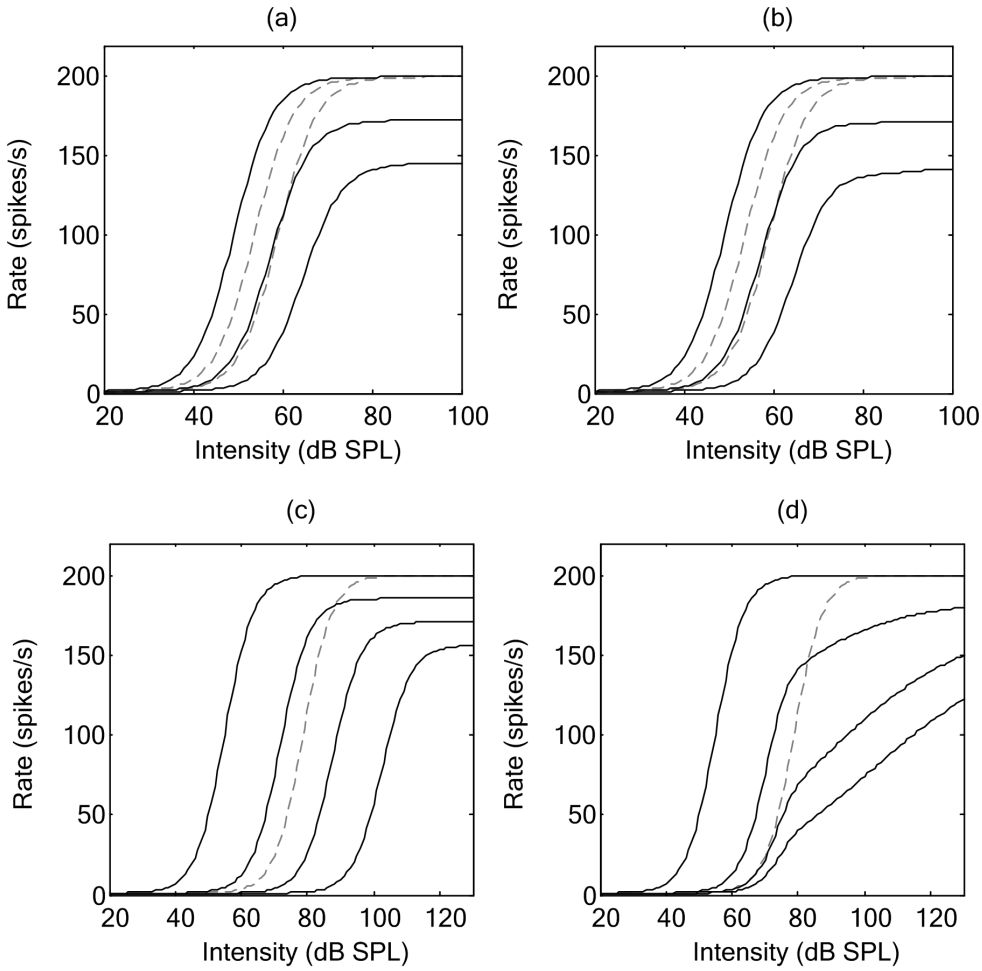

Figure 3.8.
RI functions generated with and without the basilar membrane nonlinearity included. (a) A standard nerve fibre's RI functions for five different stimulus frequencies without basilar membrane nonlinearity taken into account. The nerve fibre parameters are: spontaneous rate = 1 spike/s, maximum rate = 200 spikes/s, threshold = 49.1 dBSPL, CF = 800 Hz and α_0 = 4. The input frequencies are 600 Hz, 700 Hz (both shown with grey dashed lines), 800 Hz, 900 Hz and 1000 Hz (shown with the black solid lines). (b) The same as in (a), except the basilar membrane nonlinearity is included. (c) Another nerve fibre's RI functions, without basilar membrane nonlinearity taken into account. The nerve fibre parameters are: spontaneous rate = 1 spike/s, maximum rate = 200 spikes/s, threshold = 55 dBSPL, CF = 3000 Hz and α_0 = 4. The input frequencies are 2500 Hz (shown with the grey dashed line), 3000 Hz, 3200 Hz, 3400 Hz and 3600 Hz (shown with the black solid lines). (d) The same as in (c), except the basilar membrane nonlinearity is included.

It can be observed that when a lower threshold and frequency is used, there is not much difference between the RI functions that are generated (a and b in figure 3.8). However, at higher frequencies and thresholds, the nonlinearity does play a large role in the shape of the RI function at frequencies greater than the CF. Nevertheless, the nonlinearity was not used in the overall model of this study. The graphs on the left of figure 3.8 provide a good representation of how the RI model functions in the context of the larger model. The graphs on the right are for illustrative purposes should the rest of the RI model be expanded on in future work. The model does not take a complete representation of RI function into account owing to the diversity of the dynamics of nerve fibres.

3.3.3 Model of acoustic cochlear stimulation

The model of acoustic stimulation essentially consists of two parts: firstly, a critical bandwidth corresponding to the input frequency is determined (detailed in section 3.3.3.1), wherein the firing rate on each fibre contained in that band is determined using the RI model (described in section 3.3.2). The rate for each fibre is then used as an input to a modified Poisson process generator (from section 3.3.3.2), which creates the discharge patterns on each fibre. A block diagram illustrating the concept of the acoustic stimulation model is provided in figure 3.9.

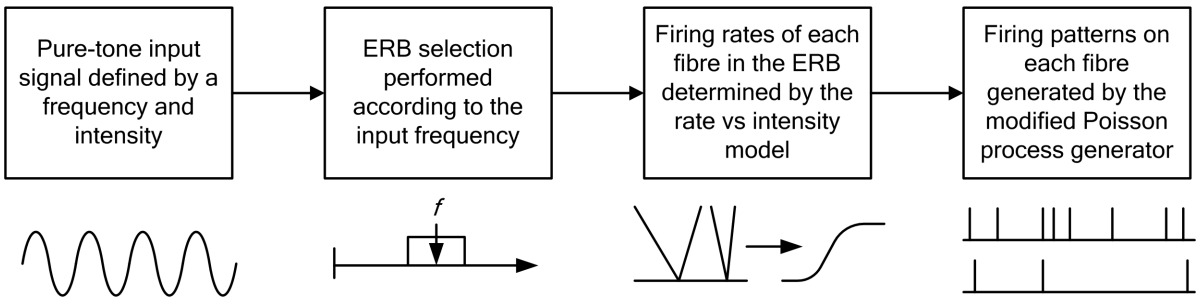

Figure 3.9.
Block diagram of the acoustic model. From left to right, the blocks correspond to descriptions found in sections 3.3.1, 3.3.3.1, 3.3.2 and 3.3.3.2.

3.3.3.1 Selection of the equivalent rectangular band

Instead of using values of critical bandwidth, a measure that approximates the critical bandwidth to an equivalent rectangular filter was used. This is known as the ERB (see appendix A). The ERB was assumed to consist of a set number of afferent channels (N_c), spaced a constant distance (Δc) from one another over a constant length along the cochlea (l_C). This follows the approach taken by Hanekom (1999), which implicitly assumes that there is some method of course filtering in the auditory periphery that isolates the critical band prior to the central processing of the input signal, where the relevant information (frequency or intensity) is extracted from the signal. Therefore, for each input stimulus, N_c channels are used by the estimator to determine the discrimination threshold for that particular stimulus.

The course filtering described above is performed by centring the ERB over the region of the cochlea that tonotopically corresponds to the input frequency. This was done by inserting the input frequency value into the equation developed by Greenwood (1990). Greenwood's equation is provided below in equation 3.8:

$$f = 165.4 \cdot (10^{0.06x} - 0.88)\ , \tag{3.8}$$

where f is the frequency in Hz associated with the place on the cochlea provided by x, the distance in mm from the cochlea apex. Since the number of channels used and the channel spacing are constant, the ERB was set to be the area centred on the point corresponding to the input frequency. Once the outer limits of the ERB are found, each fibre is assigned a CF based on its location within the band, according to equation 3.8. This is illustrated in figure 3.10

Once the CF on each of the channels is known, the model of rate vs. intensity (described in section 3.3.2) uses this information along with the input signal intensity and frequency to determine the rate at which each particular fibre in the band should fire. This is used by the AP model described in the next section (section 3.3.3.2) to create space-time AP patterns for the particular input stimulus.

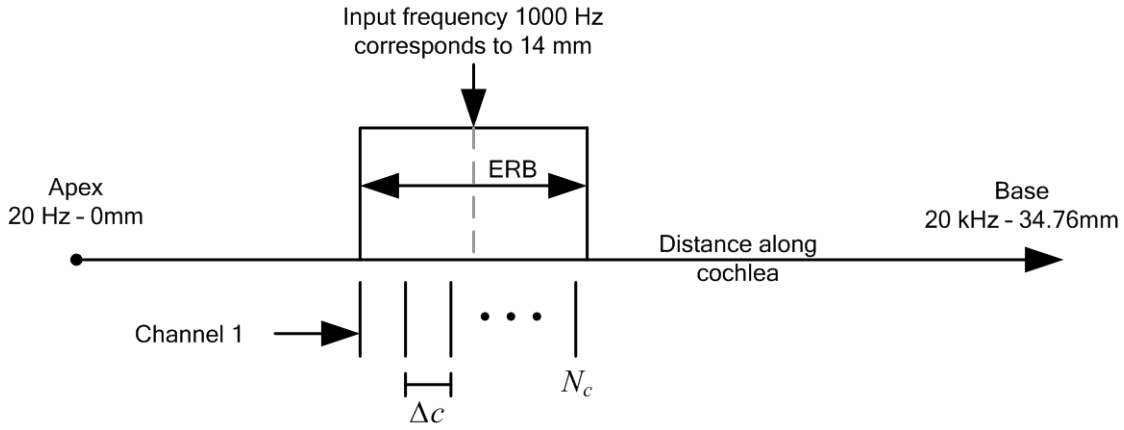

Figure 3.10.
Illustration of the ERB selection along the cochlea.

3.3.3.2 Action potential model

The discharge trains are based on the assumption that in the auditory nerve they can be approximately described by a point process. The use of this approximation is common in literature (Colburn, 1973; Hanekom, 1999; Siebert, 1970; Snyder and Miller, 1991; Young and Barta, 1986). The spike train firing patterns were modelled using a Poisson process, such as that used by Young and Barta, except that an absolute refractory period was included in the process for the sake of consistency with the model of electrical cochlear stimulation (Bruce *et al.*, 1999a). Although the model of Young and Barta (1986) makes use of a Poisson distributed dead-time (which results in roughly the same firing distribution as the relative refractory function used by Bruce *et al.*, 1999a), it was necessary to modify the equations to include the absolute refractory period as well, in order for consistency between the two models to be maintained.

As in the model of Young and Barta (1986), the neural firing model uses pulses generated from a homogenous Poisson process. Veering from the design of this model slightly, the output Poisson process is fed to a fixed dead-time processor, where all spikes within that fixed dead-time are removed. From here, the design follows that of Young and Barta again and the remaining spikes are fed to a variable dead-time processor, where an additional dead-time is appended to the fixed dead-time, and spikes falling within this period are also removed. According to Snyder and Miller (1991), the final output process is a non-paralysable, dead-

time modified Poisson process. The model is shown in figure 3.11.

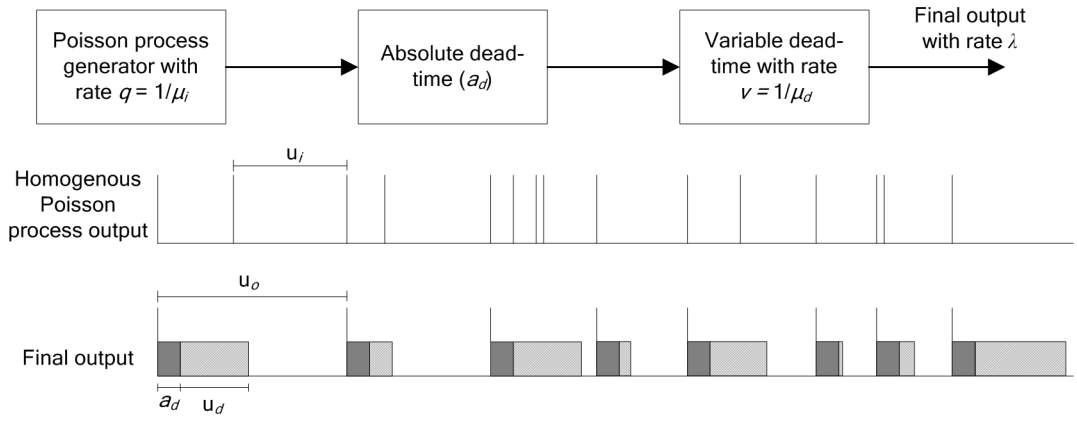

Figure 3.11.
Illustration of the structure and function of the NH listener neural firing model. At the top, the structure of the model can be observed, while in the middle a typical output of the Poisson process generator is shown. The bottom illustration shows the final output for the given input sequence, with the fixed dead-times (dark grey blocks) and variable dead-times (light shaded grey blocks) shown. This image is a modified version of figure A1 from Young and Barta (1986).

For ease of reference, a similar notation is used here as was used in Young and Barta (1986). It is assumed that a homogenous Poisson process with rate q, having a probability density function (of interevent intervals, and in the case of this model, ISIs) given by equation 3.9, describes the behaviour of the homogenous Poisson process generator. The random variable u_i represents the intervals of this process.

$$\text{pdf}_{u_i}(t) = qe^{-qt} \tag{3.9}$$

Thus, the mean ISI is $\mu_i = 1/q$ with a variance of $1/q^2$. The absolute refractory period is modelled as a fixed period of time, directly after each firing, where no firing can occur. The relative refractory period is modelled as a variable period of dead-time after the absolute refractory period, where no firing can occur. This dead-time is assumed to be a random variable (u_d), with a probability density of $\text{pdf}_{u_d}(t)$, with rate v, which takes the same form as equation 3.9. Thus the mean dead-time is given by $\mu_d = 1/v$ with a variance $1/v^2$.

For the modified version of the model, the fixed dead-time (a_d), the variable dead-time (u_d) and the ISI (u_d) are assumed to be independent of one another. Young and Barta (1986)

describe how the mean interval between spikes at the output of their model is the sum of the means of the constituent processes (Cox, 1962); the same principle applies here because of independence and thus the mean time between overall outputs is given by equation 3.10:

$$\mu_o = E[\mathrm{u}_o] = 1/q + \mu_d + a_d \ . \tag{3.10}$$

The $E[\mathrm{u}_o]$ term is the expected value of the random variable u_o, the output intervals. The variance is given by equation 3.11:

$$\sigma_o^2 = 1/q^2 + \sigma_d^2 \ , \tag{3.11}$$

where σ_d^2 is the variance of the dead-time. Note that the fixed dead-time does not contribute to the variance as its variance is zero.

For an equilibrium renewal process, to find the expected number of counts ($E[N_T]$) during an interval, the period of the interval (T) should be multiplied by the average spike rate (Cox, 1962; Young and Barta, 1986), yielding the mean number of spikes in a certain period ($\mu_N(T)$) as shown by equation 3.12:

$$\begin{aligned} \mu_N(T) &= E[N_T] \\ &= T\frac{1}{\mu_o} \\ &= \frac{T}{1/q + \mu_d + a_d} \\ &= \frac{T}{1/q + 1/v + a_d} \ . \end{aligned} \tag{3.12}$$

The approximate rate of the process, $\check{\lambda}$, can be determined by dividing the number of spikes in a given period by the length of the period, and therefore the rate can be approximated by equation 3.13:

$$\begin{aligned} \check{\lambda} &= \frac{\mu_N}{T} \\ &= \frac{qv}{q + v + qva_d} \ . \end{aligned} \tag{3.13}$$

Thus, we have a complete representation of the rate as a function of the absolute dead-time and the rates of the ISI process and variable dead-time process. Following Young and Barta (1986), the maximum rate possible with the aforementioned dead-times can be determined

by equation 3.14:

$$\begin{aligned} \lambda_{max} &= \lim_{q\to\infty} \frac{\mu_N}{T} \\ &= \lim_{q\to\infty} \frac{qv}{q+v+qva_d} \\ &= \frac{v}{1+va_d} \,. \end{aligned} \tag{3.14}$$

By using equation 3.14, and setting the value of $\lambda_{max} = A_1 + A_0$ (the maximum firing rate), the required value of v is determined, given a_d and the maximum firing rate of the fibre. This value of v is in turn used along with a_d and the actual rate of firing of the fibre to determine q, the transformed firing rate from equation 3.13. The value of q determined in this way is used as the rate parameter of the Poisson process generator. A random number is generated using Matlab's default random number generator. This random number is drawn from a uniform distribution. In order to transform the uniform random number to a Poissonian random number, equation 3.15 is used:

$$\mathrm{u}_i = -\frac{1}{q\tau} \cdot \ln(rand) \,, \tag{3.15}$$

where τ is the inverse of the sampling frequency at which the Poisson signal is processed, and *rand* is the uniformly distributed random number. This equation is a discrete time variation on an equation used by Young and Barta (1986).

The output of the Poisson process generator is passed through the absolute dead-time processor and the variable dead-time processor. The output rate (λ) will be that which is determined by the RI function of the specific fibre, except that no discharges will occur during the dead-time periods, as illustrated by figure 3.11.

In order to illustrate the changes that the dead-times apply to the random process, figure 3.12 shows three different histograms, which display different scenarios depending on which dead-time is applied. The black histogram, only considering absolute dead-time, is not consistent with measured data at short intervals. The blue histogram, which considers only the distributed dead-time, provides an adequate approximation to data measured by Kiang (1965); hence, it was used by Young and Barta (1986). The shape of the histogram when both dead-times are applied (shown in red) is consistent with the shape of the histograms measured by Kiang (1965), where absolute and relative refractory effects play a role in the statistics of nerve firing, and it conveniently remains consistent with the model of electrical

stimulation shown later in section 3.3.7.

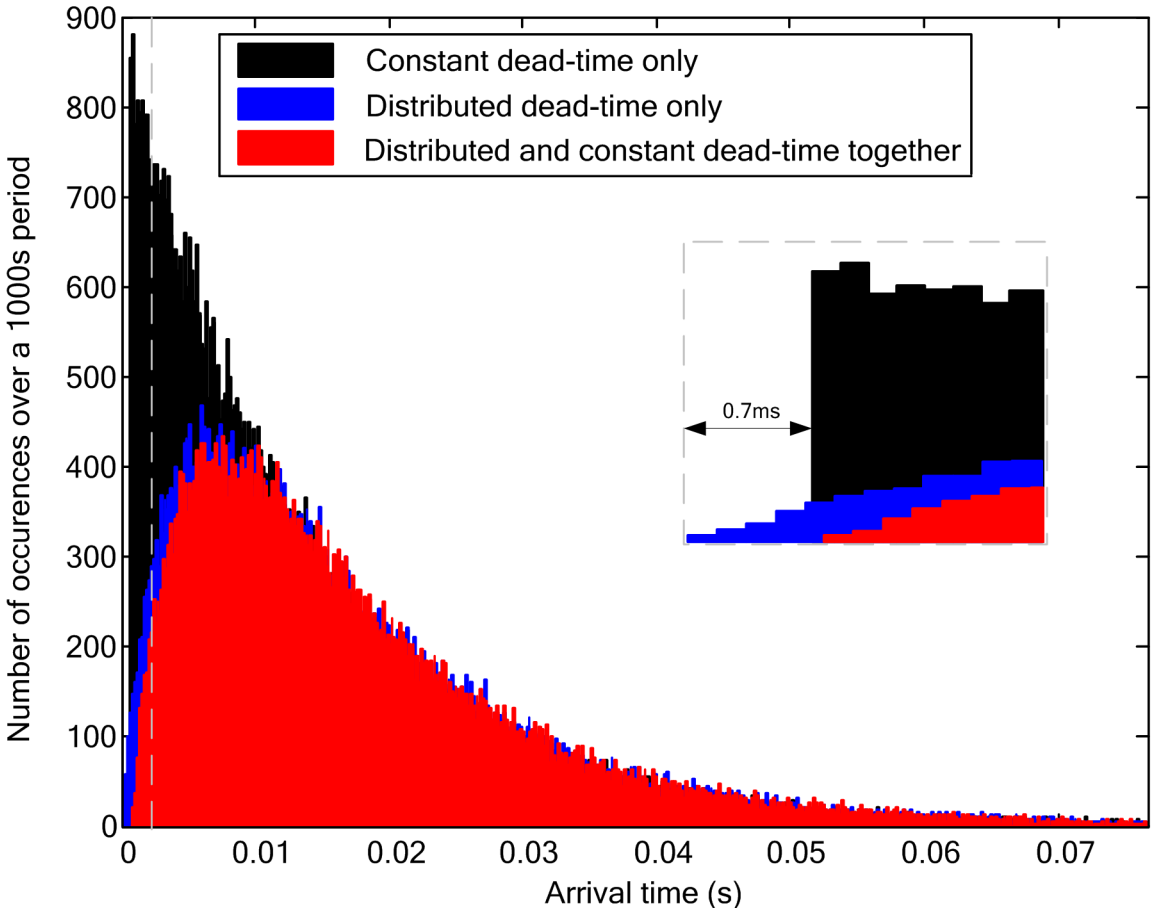

Figure 3.12.
Histograms from 1000 seconds of the Poisson process. The black histogram shows the original process with only the absolute refractory period applied, generated with q = 77/s and a constant dead-time (a_d) of 0.7 ms (the same that was used in Bruce *et al.*, 1999a). The blue histogram shows the same process but with only the distributed dead-time taken into account (q = 77/s, v = 250/s). The red histogram shows the process when both the absolute refractory period and the relative refractory period are taken into account, i.e. the constant dead-time and the distributed dead-time (q = 77/s, v = 250/s, a_d = 0.7ms). The inset shows a detail of the figure where the the effect of the constant dead-time can be clearly seen. The dotted grey line on the left-hand side of the figure indicates where the inset was taken from.

3.3.4 Rate estimator

One of the fundamental assumptions of this model is that the CANS operates "blindly", i.e. that it is unaware of its surroundings, other than the information provided to it via neural information channels from the sensory organs, which it interprets in some way. The central processor is assumed to be ideal and should therefore be optimal in some sense. The optimal estimator utilised by Hanekom (1999; 2000) was developed by Rudemo (1972) and is optimal in the least-squares sense. Rudemo's estimator can determine the approximate rate parameter of the input Poisson process by assuming it to be governed by a Markov process[6], i.e. a certain Poisson rate is associated with a particular state. The estimator determines the current state of the Markov process based on whether an event has occurred at each time point, using an internal representation of the states that could occur. Thus, the Poisson rate at each time instant is estimated. Rudemo (1972) applied his estimator in this way (defining the rate of a Poisson process), but in the context of renewal theory. The mathematical description provided in Rudemo's study is an in-depth one, so only the most important aspects of the estimator are provided here.

3.3.4.1 Description of the rate matrix

In order to make its predictions, the estimator needs to be provided with a description of the possible states of the Markov process for each channel (nerve fibres). Providing the estimator with the states (the rates that it can possibly estimate) implicitly assumes that the CANS has some internal representation that it uses to estimate the rates of a given channel. This is performed by means of a rate matrix or λ-matrix – the λ is representative of the Poisson process rate parameter, after all dead-time has been accounted for (the final output in figure 3.11). The number of states used is set to be constantly equal to the number of channels in the ERB. Therefore, the λ-matrix is square, with $N_c \times N_c$ elements. This was done because the estimator performs a matrix exponential function for each time sample on each channel, which is computationally taxing; thus, the estimator never has to compute with anything larger than a square matrix with dimensions equal to the number of channels. It should be noted that with fewer states, the frequency resolution becomes limited, and so a trade-off needed to be made between the computational time and the number of states used.

[6] See appendix C for more information.

Each row of the matrix represents a channel, and each column represents a possible state (equation 3.16). The value of a state is assigned by the RI function model ($\lambda = R$) according to a distribution of the discrimination parameter. This is different for frequency discrimination tasks and intensity discrimination tasks.

$$\Lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1N_c} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2N_c} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{N_c1} & \lambda_{N_c2} & \dots & \lambda_{N_cN_c} \end{bmatrix} \tag{3.16}$$

Frequency discrimination states. Frequency states are chosen according to the CF of the fibre. The states are determined by an even distribution of frequencies around the CF that lie within the ERB of that specific fibre. This means that some states might represent rates associated with frequencies outside of the ERB of the stimulating frequency. The assumption here is that the CANS does not yet know exactly which frequency has caused the activation of a particular fibre and so it considers all the frequencies that could possibly cause the activation of that fibre.

An example is as follows: if the stimulating frequency is 1000 Hz, and 100 channels spaced 9 μm apart (corresponding to an ERB of 0.9 mm) are assumed, the range of CFs included in the ERB for these parameters is about 931 Hz to 1072 Hz. This range is divided evenly by 100 (the channel number) to yield a resolution of approximately 1.3 Hz per channel. For the fibre representing 931 Hz, the states are determined by all of the frequencies that lie within the ERB that would have been selected had 931 Hz been the stimulating frequency (see section 3.3.3.1), i.e. 866 Hz to 1007 Hz. In the case of the 25$^{\text{th}}$ channel, with a CF of 964 Hz, this range changes to 896 Hz to 1038 Hz. This same procedure applies to all the other fibres in the active band.

Once the frequencies that correspond to each state for each channel have been found, the rate that represents each state can be found by the RI function model. The intensity is kept constant (at the same level as the input stimulus), and the CF of each channel is used with the frequencies representing each state for that channel to determine the value of the rate for each state with the RI function model. Figure 3.13 shows an example λ-matrix.

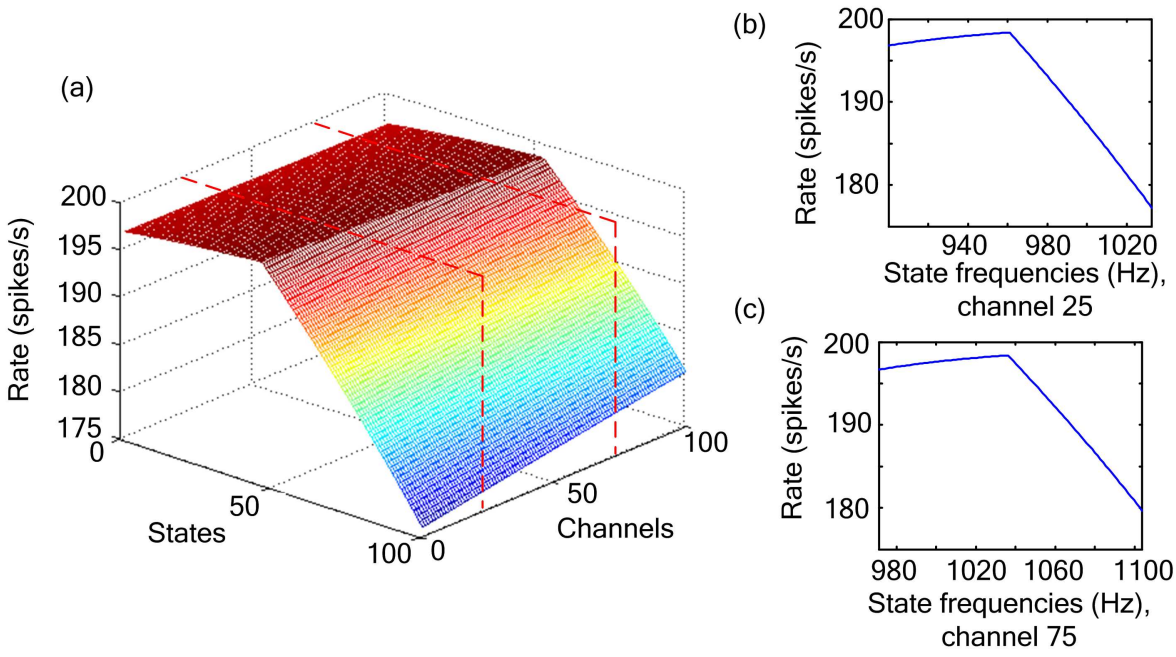

Figure 3.13.
Example of the λ-matrix for frequency discrimination, generated for 100 channels and 100 states. The entire λ-matrix is shown in (a). The CFs of the channels range from 931 Hz to 1072 Hz. The higher channel number corresponds to higher frequencies on the cochlea. The peak provides an approximate indication of where the CFs lie relative to the state number. (b) and (c) show cross sections of the λ-matrix (indicated by the red, dotted lines in (a)) for channels 25 and 75 respectively. It is clear how the state frequency representation differs for each channel.

Intensity discrimination states. The λ-matrices generated for intensity discrimination are created in a similar way to those described above. The process is a degree simpler owing to the fact that the channels are not tuned to intensity. The states are determined by an even distribution of intensities over a constant range of intensities. The same implicit assumption used above still applies. This means that for every channel, the states will have the same intensity resolution, and corresponding intensities on all channels will be represented by the same state. The states are again generated by the RI model, but the frequency at which stimulation occurs is held constant; only the CF of the channel changes for different fibres and the intensity is varied over the specified range of intensities for each state, i.e. in the case of intensity discrimination, the intensity values correspond to individual states instead of frequencies, but again, each individual state is a rate. An example λ-matrix for intensity

discrimination is shown in figure 3.14.

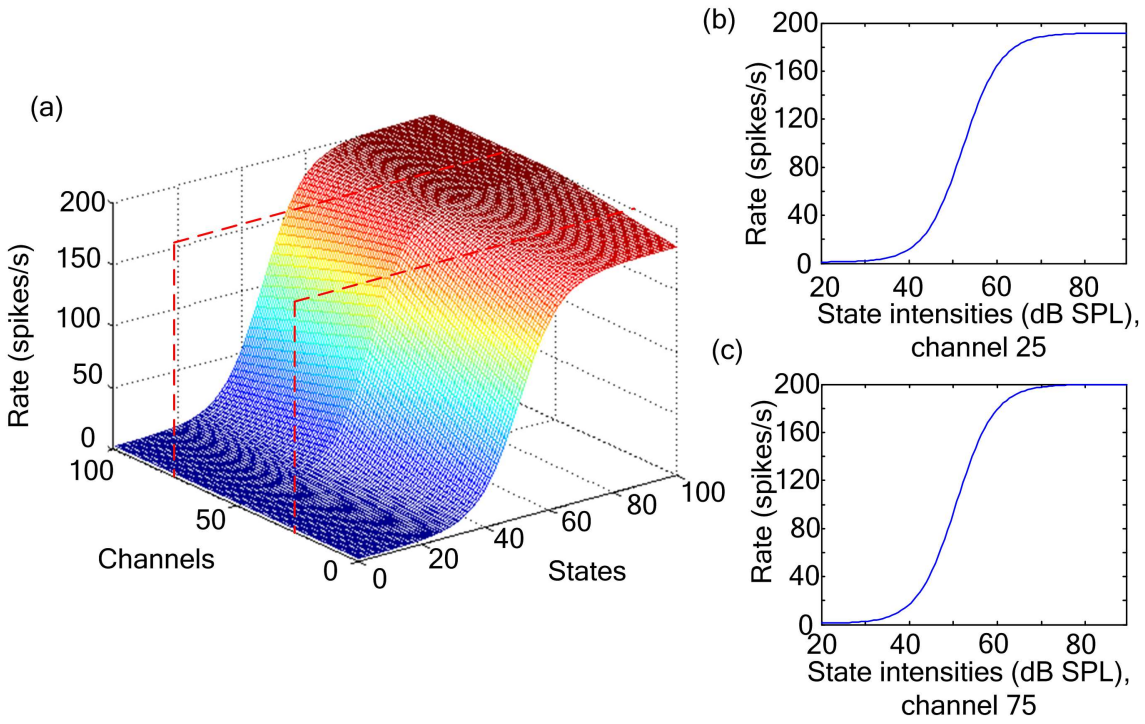

Figure 3.14.
Example of the λ-matrix for intensity discrimination, generated for 100 channels and 100 states. The entire λ-matrix is shown in (a). The shape of the RI function is visible. The ridge running through the centre of the plot marks the fibre closest to the CF. (b) and (c) show cross sections of the λ-matrix (indicated by the red, dotted lines in (a)) for channels 25 and 75 respectively.

3.3.4.2 Estimation of the state

The mathematical description provided by Rudemo (1972) can best be understood if it is translated into an algorithmic form. This is provided here by figure 3.15. After the algorithm has been illustrated and described, the mathematical components are defined.

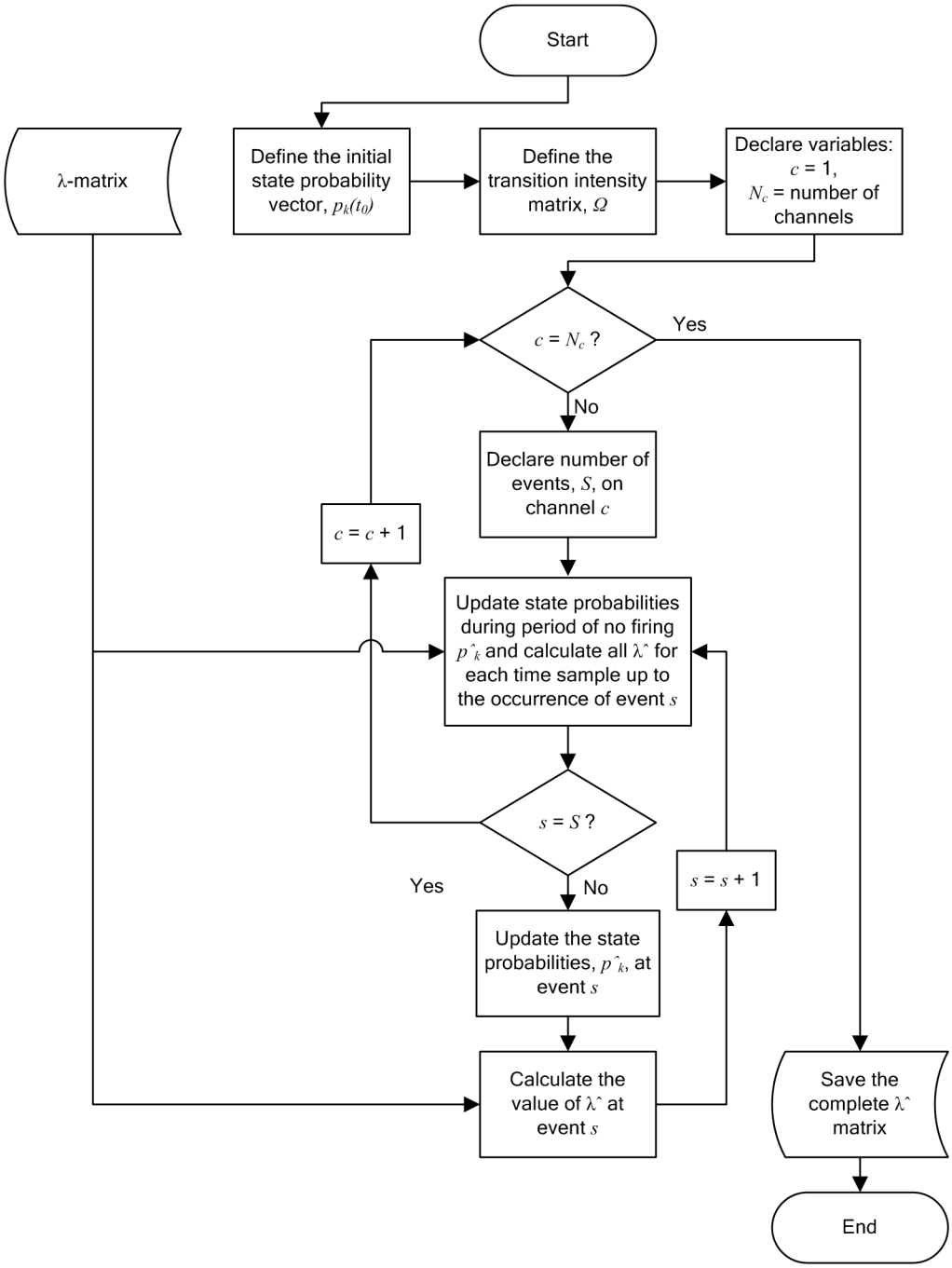

Figure 3.15.
The algorithm used by the rate estimator to determine the instantaneous rate on each channel using the firing pattern, state transitions and transition intensity values.

Figure 3.15 provides a basic representation of the algorithm used to calculate the instantaneous rate on a given fibre at a certain time instant. The paragraphs below summarise the mathematical equations used in the algorithm. In these calculations, a static transition probability matrix is assumed (Rudemo, 1972).

Initial state probability vector. The state probability vector is used by the estimator in determining the rate based on the probability that the Markov process will be in a particular state at a given time, and it represents the state probabilities of a given channel. It was desired that no bias be applied to any particular state, and therefore an equal probability was assigned to each state probability; therefore, where K is the total number of states per channel,

$$p_k(t_0) = [1/K \; 1/K \; \ldots \; 1/K]. \tag{3.17}$$

Transition intensity matrix. The transition intensity matrix (Ω) is very important in the estimator as it defines the intensity[7] by which the Markov process changes when an event occurs. This is the magnitude of the change that the rate estimate makes when transitioning from a specific state to the next. Each state of each channel has a corresponding transition intensity. The transition intensity matrix is determined by the λ-matrix (described in section 3.3.4.1) and predefined transition intensity values ω. These values are defined by the same units as the rate, i.e. $1/s$. The matrix Ω has the following elements:

$$\begin{aligned} \Omega_{ii} &= -(\omega_i + \lambda_{ii}),\ i \in K, \\ \Omega_{ij} &= \omega_{ij},\ i, j \in K,\ i \neq j. \end{aligned}$$

In this context, let K refer to the entire state space. The predefined λ-matrix described in section 3.3.4.1 is used here. A condition set out by Rudemo (1972) is that

$$\sum_{k \neq i} \omega_{ik} = \omega_i \ .$$

Additionally, for simplicity in this model,

$$\omega_{ij} = \omega_i.$$

[7] This is not to be confused with the intensity value used to define the sound level of signal.

By these equations, the transition intensity matrix was determined and used by the rate estimator.

Recursive estimation. In order for the instantaneous rate to be estimated, the estimator undergoes a process of recursive estimation, whereby at each time sample, the state probability vector is updated according to the estimates found in the previous time sample. This is then used to calculate the current rate estimate. The process described below is performed on each individual channel to obtain an instantaneous rate at each time sample on every channel. The equations listed are applied in the state space of a single channel.

In order for the state probability vector estimate to be obtained, Rudemo (1972) defines a linear system of differential equations, whereby a new state probability vector, p^*, is obtained, from which the state probability vector can be calculated anew for each time sample. The details of the mathematics are not provided here, only the practical implementation.

In the initial period before any event has occurred, there is no previous information to draw from, hence the need for the declaration of the initial probability vector (equation 3.17). The set of equations solved recursively in this period is given in equations 3.18, 3.19 and 3.20[8].

$$\hat{p}_k^*[nT] = p_k[0]e^{\Omega(nT)} \tag{3.18}$$

$$\hat{p}_k[nT] = \hat{p}_k^*[nT] / \sum_i \hat{p}_i^*[nT] \tag{3.19}$$

$$\hat{\lambda}[nT] = \sum \lambda_k \hat{p}_k[nT] \tag{3.20}$$

The subscript k indicates that a vector over all the states at that given time for that channel was used. The variables n and T stand for sample number and sampling period respectively. The variable λ_k is obtained from the rows (all the states of the given channel) of the matrix Λ generated in section 3.3.4.1, depending on the percept being tested.

Once a neural firing occurs, the state probability vector is updated differently from the intial set of equations (equations 3.18, 3.19 and 3.20). This results in a jump in the estimated rate when it is evaluated. When such an event occurs, equations 3.21 and 3.22 are used to calculate the state probability vector: this only occurs once per event. The sample at which

[8] All the equations for the estimator are provided in a discretised form as this is how they were implemented in Matlab, although Rudemo (1972) provides them in a continuous form.

an even occurs is denoted m, but is indexed on the same continuum as n.

$$\hat{p}_k[mT] = \hat{p}_k[(m-1)T]\frac{\lambda_k}{\hat{\lambda}[(m-1)T]} \tag{3.21}$$

$$\hat{\lambda}[mT] = \sum \lambda_k \hat{p}_k[mT] \tag{3.22}$$

After an event, the last sample at which the event occurred is considered to be the new state probability matrix. Therefore, in order for the estimation to continue, equation 3.18 must change.

$$\hat{p}_k^*[nT] = p_k[mT]e^{\Omega(n-m)T} \tag{3.23}$$

Equation 3.23 is used before equations 3.19 and 3.20 at every time point recursively. This process continues until another event occurs, at which point 3.21 and 3.22 are used again, and so on.

Estimator output. Once this process has been completed, a space-time representation of the rate on each nerve fibre is available. This is represented by the $\hat{\lambda}$-matrix – the estimated value of λ for all channels and time points. This output is then processed by the decoders described in section 3.3.5. Sample outputs are shown in figure 3.16.

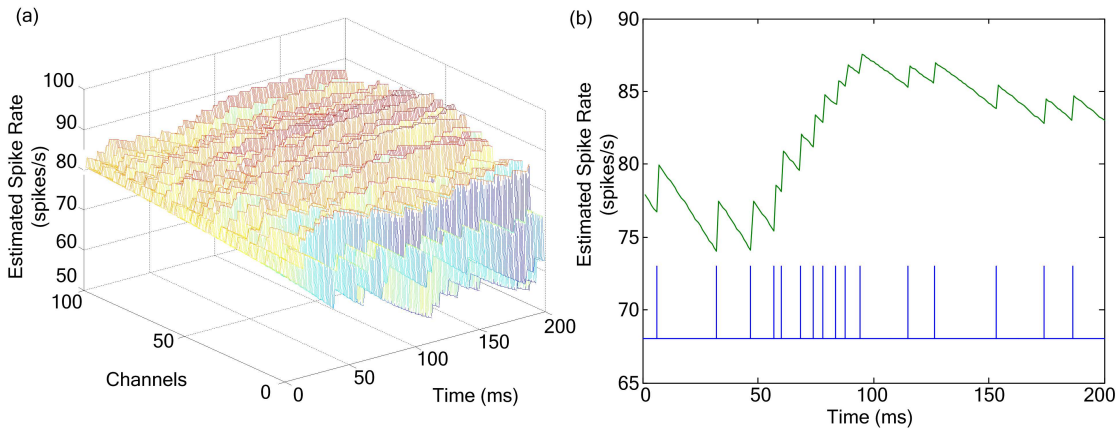

Figure 3.16.
Example output of the rate estimator ($\hat{\lambda}$-matrix). (a) shows the overall output over a population of channels. (b) shows a single channel's output, compared with the nerve firing pattern on that channel. It is clear that jumps in the rate estimate appear when a nerve firing occurs. The average rate on this particular fibre was 80 spikes/s.

3.3.5 Decoders

The purpose of the decoder is to interpret the information generated by the rate estimator (the space-time AP patterns), and translate it into a discrimination threshold. Two different decoders were implemented in this study in order that the two could be compared. A Viterbi decoder (Viterbi, 1967) was implemented in addition to a centre-of-gravity (COG) decoder. When the Viterbi decoder is applied to Markov models, it finds the most likely state sequence that could have occurred, given the information. The COG decoder finds the value at which the centroid of the tuning parameter in question lies (Snippe, 1996).

3.3.5.1 Centre-of-gravity decoder

The COG decoder is a simple decoder that provides a straightforward way of extracting a tuning parameter from a population of neurons (Snippe, 1996). A value is obtained from the COG, representing what the decoder interprets as the point with the greatest weight with regard to the neural responses; thus, the information across all channels is integrated with the COG decoder. As the nerves are tuned to the parameter in question, this correlates to some degree to the value of the parameter stimulating the neural population. In this model, the extracted parameter is frequency. Intensity cannot be extracted by this method as the nerves are not tuned to specific intensities. If a variety of RI function descriptions were to be used, it is conceivable that such a parameter extraction method could be employed, but this is beyond the scope of this work. The COG of frequency in the neural population is defined by equation 3.24:

$$\hat{f}(t) = \frac{\sum_n f_{CF,n}\lambda_n(t)}{\sum_n \lambda_n(t)}, \tag{3.24}$$

where $\hat{f}(t)$ is the estimate of the frequency at time t, $f_{CF,n}$ is the CF of channel n and $\lambda_n(t)$ is the rate on channel n at time t.

When this decoder is applied to the $\hat{\lambda}$-matrix generated by the rate estimator in section 3.3.4.2 by computing the COG for each time sample, the estimated frequency over time can be analysed. Due to noise on the population output, the extracted frequency will vary with time. According to Siebert (1970), the standard deviation of this parameter can be used to directly obtain the JND in frequency. An example output of the COG decoder is shown in figure 3.17.

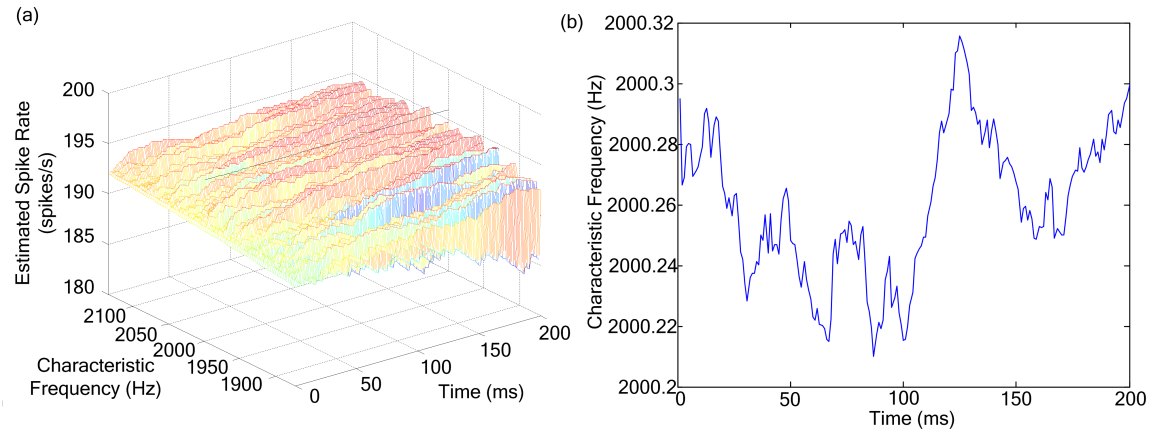

Figure 3.17.
Example of the function of the COG decoder. (a) shows the estimator output of one population of nerve fibres with a stimulating frequency of 2 kHz; the trace of the COG is shown with a black line just above the estimator output. Relative to the range of frequencies the neural population represents, the line appears to be straight. The COG output is displayed in (b) as a function of time and can be seen here to be somewhat varied, but over a small range of frequencies. The standard deviation of this data is used as the estimate of JND in frequency. The standard deviation of this particular decoding is 0.024, which corresponds to a JND in frequency of 0.024 Hz.

3.3.5.2 Viterbi algorithm decoder

The Viterbi algorithm is normally applied in decoding convolutional codes in digital communication systems (Forney, 1973; Proakis and Salehi, 2002; Viterbi, 1967), but it can also be applied to other fields, such as speech and character recognition, in estimating the most likely state sequence of hidden Markov models (HMMs) (Lou, 1995; Rabiner and Juang, 1993). The HMM is so named because the states of the model are “hidden”, but each state is associated with an observation probability or continuous observation probability density (Lou, 1995) so that it can be estimated. For the purposes of this study, the Markov process that the rate estimator estimated was considered to be the HMM, and the most likely state was thus decoded for every time sample, producing the most likely state sequence. The implementation of the algorithm was sourced from Rabiner and Juang (1993), where an alternative implementation is used so that additions are used instead of multiplications, thereby reducing computational complexity. Details of the original implementation can be found in Rabiner and Juang (1993), and the alternative implementation is described here.

The basic purpose of the algorithm is to find the state sequence, $\mathbf{k} = (k_i\ k_2 ... k_N)$ that has the highest likelihood score, $\delta_N(i)$, and is therefore the most probable given the observed information from the model ($\hat{\lambda}$). The observation sequence can be defined in vector format as $\hat{\Lambda} = (\hat{\lambda}_1\ \hat{\lambda}_2 ... \hat{\lambda}_N)$, where each vector represents the information from all channels at one time sample. This likelihood score is defined by equation 3.25.

$$\delta_n(i) = \max_{k_1,k_2,...,k_{n-1}} P(k_1 k_2 ... k_{n-1},\ k_n = i,\ \hat{\lambda}_1 \hat{\lambda}_2 ... \hat{\lambda}_n | \lambda) \tag{3.25}$$

The equation states that the likelihood score at time sample n is given by the highest probability along a path that could possibly account for the observations leading up to point n, ending in state i, given the model λ. The model λ is the same as the λ-matrix generated in section 3.3.4.1.

Transition probability matrix. As mentioned in section 3.3.4.2, the transition probability matrix[9] Π was kept static. Similarly to the initial state probability matrices used by the rate estimator, the transition probabilities were assumed to be equal for all states for every channel, i.e. no state was any more likely to transition to one state than to another. Thus, the transition probabilities were defined by equation 3.26, where K is the total number of states per channel.

$$\Pi = \begin{bmatrix} 1/K & 1/K & \dots & 1/K \\ 1/K & 1/K & \dots & 1/K \\ \vdots & \vdots & \ddots & \vdots \\ 1/K & 1/K & \dots & 1/K \end{bmatrix} \tag{3.26}$$

Observation probability matrix. For the purposes of this study, it was assumed that the observation probability matrix Γ was defined according to a probability density function. More specifically, the probability of a given state at a particular time was defined by the normal distribution of the difference between the estimated rate at a particular time and the modelled rate of each channel at each state. Therefore, for a given time sample (n), state (i) and channel (c), the observation probability would be defined by equation 3.27, with σ_o as the variance in the observation probabiliy.

$$\gamma_c^*(\hat{\lambda}_n) = \frac{1}{\sqrt{2\pi\sigma_o}} e^{\frac{-(\lambda_{c,i} - \hat{\lambda}_{c,n})^2}{2\sigma_o}} \tag{3.27}$$

[9] See appendix C.

The overall observation probability for a specific state at a specific time is given by the cumulative probability over all channels for each time point. This is how the information from the space-time representation of firing patterns is integrated with this form of decoding. For state i at time sample n, the observation probability is given by equation 3.28.

$$\gamma(\hat{\lambda}_n) = \prod_c \gamma_c^*(\hat{\lambda}_n) \tag{3.28}$$

Algorithm implementation. The alternative implementation described by Rabiner and Juang (1993) requires that the natural logarithm of the model parameters be obtained. A parameter that has not yet been discussed is the initial probability that is applied to the observation probability. This is taken directly from the state probability vector. The initial algorithm parameters are defined as follows:

$$\tilde{p}_{i,0} = \ln(p_i[0]), \text{ for } 1 \leq i \leq K\,, \tag{3.29}$$

$$\tilde{\gamma}_i(\hat{\lambda}_n) = \ln(\gamma_i), \text{ for } 1 \leq i \leq K,\ 1 \leq n \leq N \text{ and} \tag{3.30}$$

$$\tilde{\pi}_{ij} = \ln(\pi_{ij}), \text{ for } 1 \leq i,\ j \leq K. \tag{3.31}$$

The initial values of the likelihood score and the tracking array ϕ need to be initialised. The tracking array is initialised to zero for all states at the first time sample. The initial likelihood score is computed by equation 3.32.

$$\tilde{\delta}_1(i) = \ln(\delta_1(i)) = \tilde{p}_{i,0} + \tilde{\gamma}_i(\hat{\lambda}_1), \text{ for } 1 \leq i \leq K \tag{3.32}$$

Once these initialisations have been made, the recursive algorithm can be started. For each time sample, two recursive equations are solved. Equation 3.33 calculates the likelihood score of each state at each time point while equation 3.34 determines which state is the most likely by using the likelihood score.

$$\tilde{\delta}_n(j) = \ln(\delta_n(j)) = \max_{1 \leq i \leq K} \left[\tilde{\delta}_{n-1}(i) + \tilde{\pi}_{ij}\right] + \tilde{\gamma}_j(\hat{\lambda}_n), \text{ for } 2 \leq n \leq N,\ 1 \leq j \leq K \tag{3.33}$$

$$\phi_n(j) = \arg\max_{1 \leq i \leq K} \left[\tilde{\delta}_{n-1}(i) + \tilde{\pi}_{ij}\right], \text{ for } 2 \leq n \leq N,\ 1 \leq j \leq K \tag{3.34}$$

After the recursion has been completed for each time point, the algorithm needs to begin the backtracking process (in order to find the most likely state sequence) by finding the final

state. This is found by equation 3.35.

$$k_N^* = \underset{1 \leq i \leq K}{\arg\max} \left[\tilde{\delta}_N(i) \right] \tag{3.35}$$

The final value of the likelihood score at this point is given by equation 3.36.

$$\tilde{P}^* = \max_{1 \leq i \leq K} \left[\tilde{\delta}_N(i) \right] \tag{3.36}$$

From equation 3.35 and the final ϕ matrix, the backtracking procedure can be performed, and the most likely state sequence over all samples can be found by equation 3.37.

$$k_n^* = \phi_{n+1}(k_{n+1}^*) \tag{3.37}$$

The matrix q^* gives the most likely state sequence over all time samples, and given that the frequencies or intensities represented by each state are known, the state sequence can be translated back into a sequence of frequencies or intensities that the Viterbi algorithm tracked by means of a simple matrix multiplication. The standard deviation taken from this matrix directly yields the JND in frequency or intensity, following Siebert (1970). An example output for intensity and frequency discrimination is shown in figure 3.18.

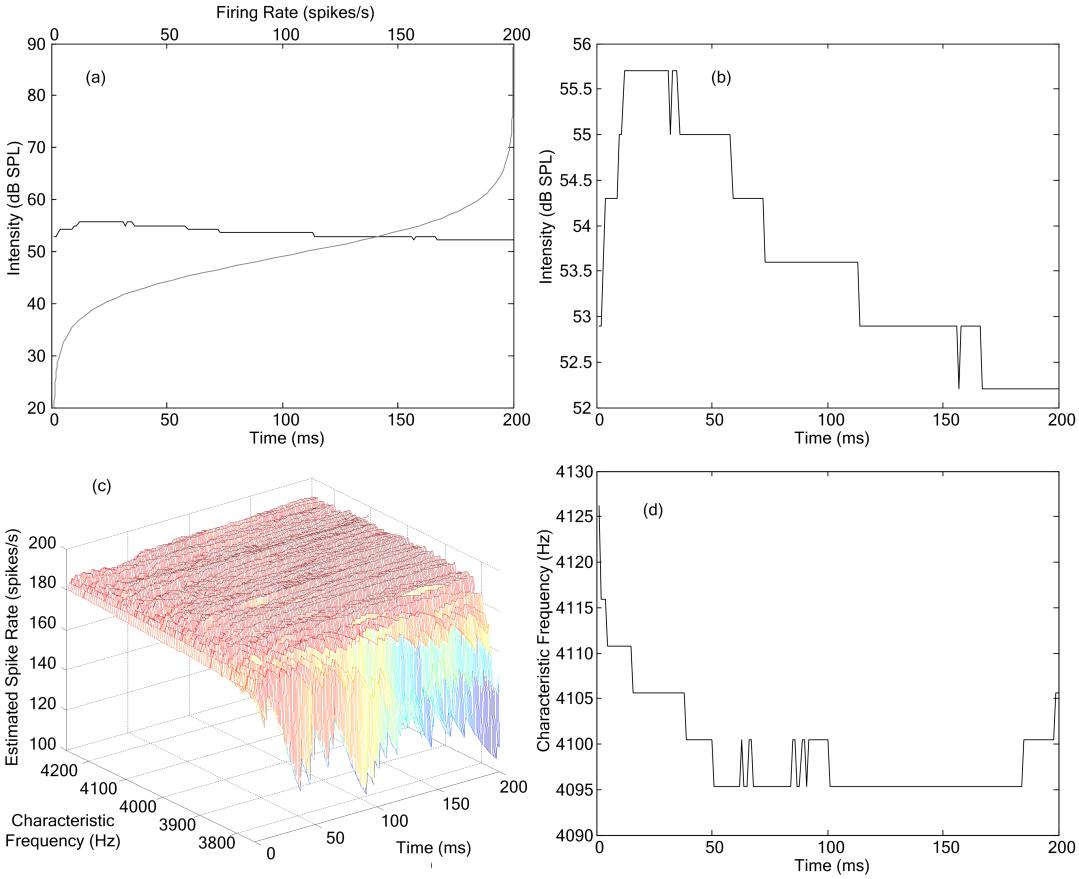

Figure 3.18.
Example of the function of the Viterbi algorithm decoder. (a) The Viterbi decoder output (bottom x-axis, black plot) compared to the RI function of a nerve fibre (top x-axis, grey plot) for intensity discrimination data. The measurement was taken for a fibre at 50 dB SPL at 1000 Hz. (b) The Viterbi decoder output from (a) is displayed at a higher resolution. The standard deviation of this data is used as the estimate of JND in intensity, but can be applied in the same way to obtain frequency discrimination data. The standard deviation of this particular decoding is 1.14, which corresponds to a JND in intensity of 1.14 dB SPL. (c) The Viterbi decoder output of one population of nerve fibres with a stimulating frequency of 4 kHz for frequency discrimination data; the trace of the Viterbi decoding is shown with a black line just above the estimator output. (d) The Viterbi decoder output from (b) as a function of time can be seen here to be varied over a larger range of frequencies when compared to the COG decoder example in figure 3.17. The standard deviation of this data is used as the estimate of JND in frequency. The standard deviation of this particular decoding is 5.45, which corresponds to a JND in frequency of 5.45 Hz.

3.3.6 Cochlear implant model

In order to model the peripheral processing of a cochlear implanted subject, it is important to include the signal processing performed on the input signal that transforms it into pulsatile stimuli. For this study, this processing was performed on the stimulus by means of a Nucleus 24 CI[10] with a SPrint speech processor, utilising the ACE speech processing strategy. The processing used in this particular implant is completely described by the documentation from Cochlear Ltd (2002), but parts of the processing are also described by Clark (2003), Nogueira *et al.* (2005) and Vandali *et al.* (2000). Some of the Matlab routines provided in the Nucleus Matlab Toolbox, version 4.03[11] were used by the model. This particular processing strategy was decided upon because of its widespread use in the CI community and the readily available literature on it. The structure and details of the model are given here.

3.3.6.1 Overview of the cochlear implant model

The CI processing model is described here in a modular way and figure 3.19 shows a breakdown of it on the next page. The processing did not need to be performed in real time as it is done in actual CI processors, as the input signal was stored in computer memory.

Time windowing. The first stage of the signal processing involves sampling the windows of information from the input signal. In the Nucleus 24 SPrint processor with ACE, audio information is sampled at a rate of 16 kHz (Nogueira *et al.*, 2005; Vandali *et al.*, 2000). This sampling rate is used so that sufficient information is sampled in the signal (the Nyquist criterion is fulfilled for frequencies up to 8 kHz) for speech recognition. Owing to the fact that the filtering process uses fast Fourier transforms (FFTs), set periods of time are needed in order to obtain an adequate representation of frequency information. The 16 kHz signal is broken down into 8 ms time windows, which translates into 128 samples per time window (Vandali *et al.*, 2000). However, the time windows are not completely independent, and overlap each other. The frequency at which a new window is analysed is known as the analysis rate. In the Nucleus CI, this is typically equal to the stimulation rate (when the stimulation rate is below approximately 760 pps according to Vandali *et al.*, 2000). A lower analysis rate was used in this study to reduce computational complexity. This does not have a substantial

[10] Manufactured by Cochlear Ltd.
[11] The Nucleus Matlab Toolbox was provided by Cochlear Europe Ltd (www.cochlear.com).

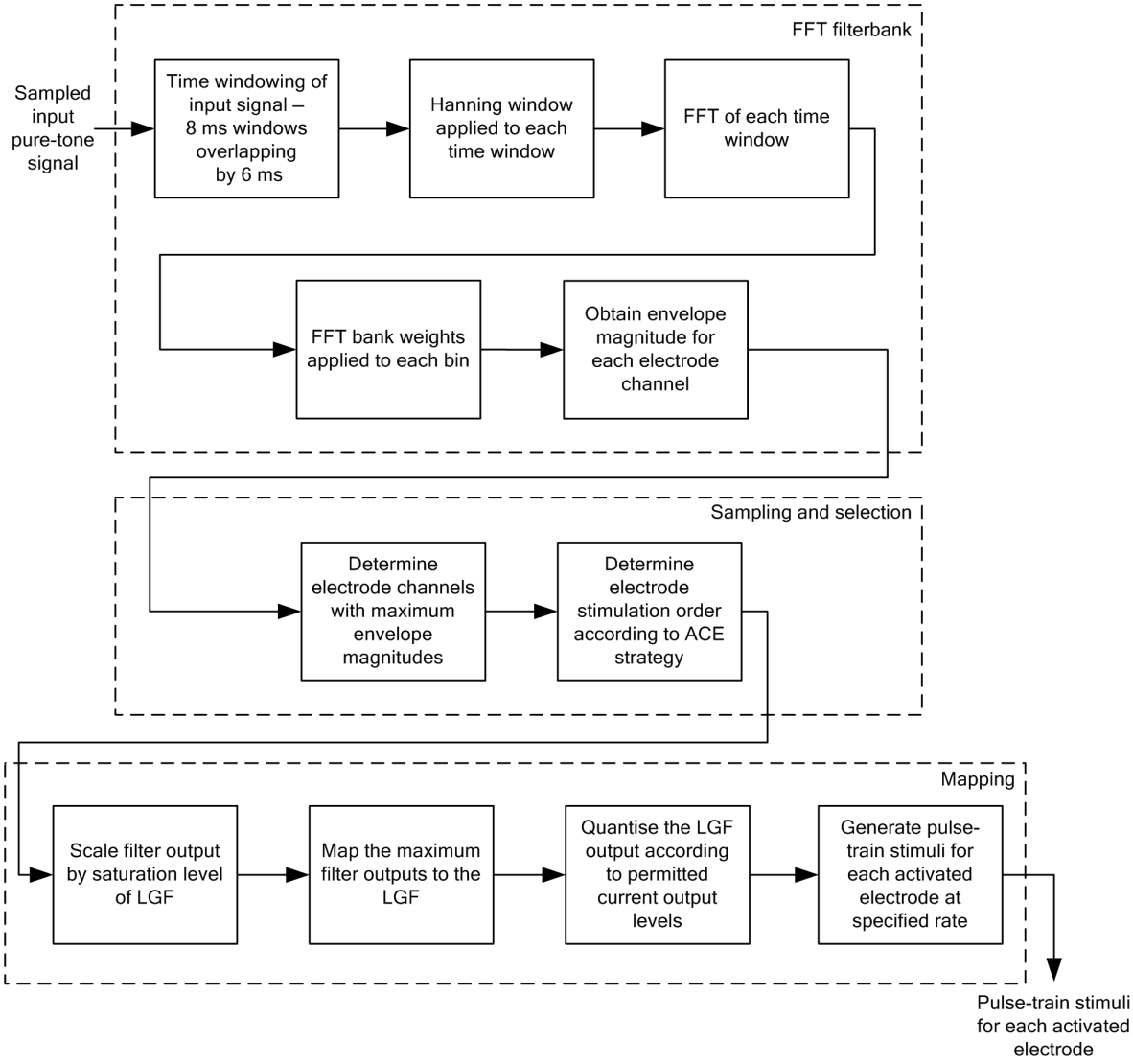

Figure 3.19.
Overview of the CI processor model. The solid boxes indicate the individual processing steps described in this section. The broken line boxes indicate the main processing components as described in Nogueira *et al.* (2005) and Nogueira, Kátai, Harczos, Klefenz, Buechner and Edler (2007).

effect on the output current stimuli, as only pure tones were used in this study. Pure tones are periodic by nature and result in repetitive output pulses. The analysis windows overlap by 6 ms (96 samples). This means that every 2 ms, a new 8 ms time window is sampled by the processor. The overlap in time windows facilitates a better time resolution in the sampling

process while maintaining the frequency resolution. An illustration of this process is shown in figure 3.20.

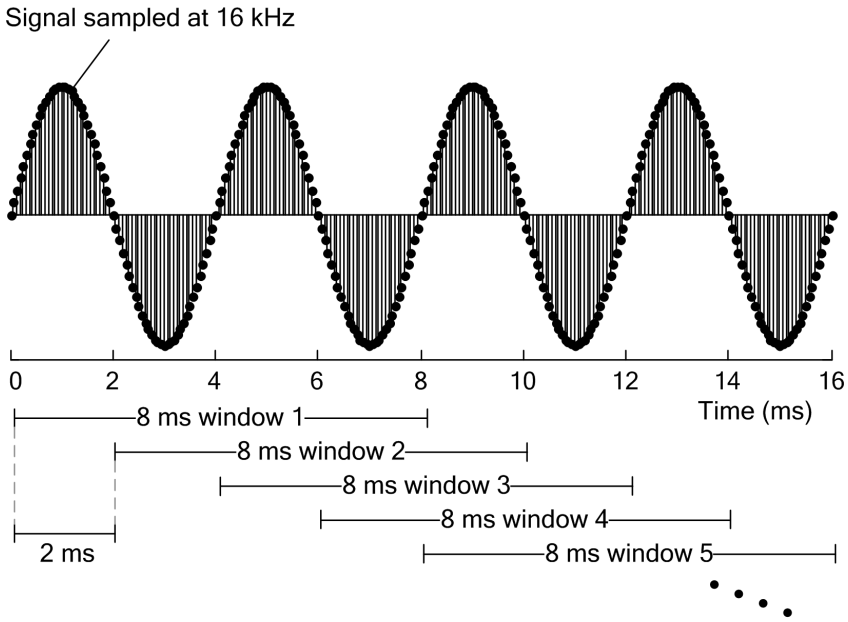

Figure 3.20.
Illustration of the time windowing carried out by the processor on a 250 Hz signal. Individual windows are illustrated and the time difference between the first and second windows is indicated.

Hanning window. In order to remove any high-frequency components created on the windowed signal as a result of the rectangular window, a Hanning window is applied to each time window (Nogueira *et al.*, 2005). The Hanning window is described by the function given by equation 3.38:

$$w(n) = 0.5\left(1 - \cos\left(\frac{2n\pi}{L}\right)\right) \quad \text{for} \quad n = 0, 1, ..., L-1\ , \tag{3.38}$$

where L is the length of the window in number of samples, i.e. 128 samples, and each value of n corresponds to a sample in the window. In order to apply the Hanning window, a multiplication must be performed on each sample of the current time window by each sample

of the Hanning window. This multiplication is shown in equation 3.39:

$$w_o(n) = w(n) \cdot w_i(n), \quad \text{for } n = 0, 1, ..., L-1 \tag{3.39}$$

where w_i is the input window and w_o is the input window with the Hanning window applied. Figure 3.21 shows the effect of the Hanning window on a sample input signal. The next step is the FFT.

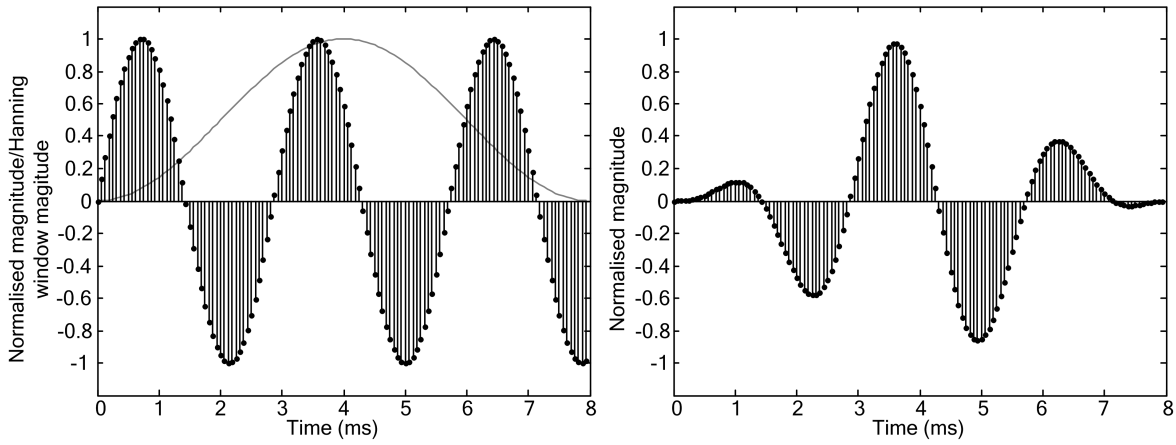

Figure 3.21.
Illustration of the effect of the Hanning window. The figure on the left shows the 8 ms window of a 350 Hz input signal sampled at 16 kHz, indicated by the black stem plot. The grey line plot shows the Hanning window magnitude over the window length. The figure on the right shows the output once the Hanning window has been applied to the signal shown in the figure on the left.

The FFT filter. A 128-point FFT is performed on the Hanning windowed signal. This means that each bin of the FFT will correspond to centre frequencies spaced 125 Hz apart (Vandali *et al.*, 2000). The bins are indexed from 0 to 127. Bin 0 will correspond to 0 Hz; the 64th bin will correspond to 8000 Hz. The 65th to 127th bins are discarded (Nogueira *et al.*, 2005). According to Vandali *et al.* (2000), the Hanning window bestows a 250 Hz, 6dB bandwidth on each FFT bin. Each bin is characterised by a real part and an imaginary part. Therefore, the magnitude squared of the κ^{th} bin is given by equation 3.40, in which the real

part of the κ^{th} bin is given by $x(\kappa)$ and the imaginary part of the κ^{th} bin is $y(\kappa)$.

$$r^2(\kappa) = x^2(\kappa) + y^2(\kappa), \quad \text{for} \quad \kappa = 0, 1, ..., L/2 \tag{3.40}$$

An illustration of the FFT acting on a 350 Hz, Hanning windowed signal is shown in figure 3.22.

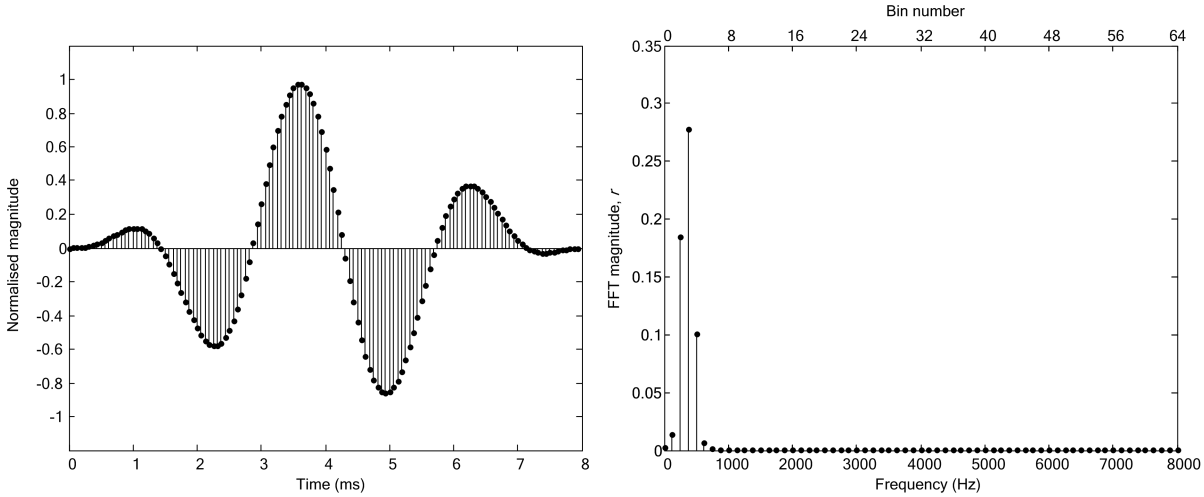

Figure 3.22.
Illustration of the effect of the FFT. The figure on the left shows the 8 ms window of a 350 Hz input signal sampled at 16 kHz, after a Hanning window has been applied. The figure on the right shows the output of the FFT. The top axis of the right-hand figure indicates the bin number of the FFT and the bottom axis indicates the corresponding frequency.

Bin weighting and envelope magnitude. After the FFT has been performed, the processor takes the first step towards assigning a magnitude to each electrode. This is done by determining the envelope magnitudes for each filter band (each of which corresponds to an individual electrode), by applying the weighted sum of the FFT bin magnitudes squared. The bin weights are determined by the number of channels in use by the subject. It was assumed for the purposes of this study that the model subject utilised 22 channels throughout over the full range of frequencies available, i.e. up to 8 kHz. Therefore, the total number of filters Z was 22. Given the assumptions above, the filter table that appears in Nogueira *et al.* (2005) was used, and it is reproduced here for convenience in table 3.1. The column showing the

number of FFT bins per band indicates how many FFT bins are included in the weighted summation of FFT bin powers. The column labelled “Centre frequency” indicates the centre frequencies of each filter.

Table 3.1.
CI FFT filter table. The leftmost column indicates the electrode with which the filter is associated. The middle column indicates the number of FFT bins that are included in the filter. The column on the right indicates the centremost points between which the magnitude response of adjacent filters are equal. This table is adapted from Nogueira *et al.* (2005).

Filter number	Number of FFT bins per band	Centre frequency
1	1	250
2	1	375
3	1	500
4	1	625
5	1	750
6	1	875
7	1	1000
8	1	1125
9	1	1250
10	2	1437
11	2	1687
12	2	1937
13	2	2187
14	3	2500
15	3	2875
16	4	3312
17	4	3812
18	5	4375
19	5	5000
20	6	5687
21	7	6500
22	8	7437

The magnitude of the z^{th} filter-band envelope ($v(z)$) is calculated by equation 3.41 (adapted

from Nogueira *et al.*, 2005).

$$v_o(z) = \sqrt{\sum_{\kappa} h(z,\kappa) r^2(\kappa)}, \quad \text{for } z = 1,2,...,Z \tag{3.41}$$

The weighting factors h are determined according to the cross-over frequencies of the filters, the sampling frequency and the FFT length. The cross-over frequencies are derived from table of centre frequencies (table 3.1). The weights were determined by a function provided with the Nucleus Matlab Toolbox.

Sampling and selection. The envelope magnitudes are available for each filter and therefore for each electrode, but electrical stimuli are not applied to every electrode in the ACE strategy. For this study, the maximum number of activated electrodes, Z_{max}, was set to six. No more than this was required as only pure tones were used as stimuli and the resulting FFTs of such stimuli have very narrow bandwidths, which results in few electrodes being activated. The six filters with the highest envelope energies had their corresponding electrodes selected as the stimulating electrodes. According to the stimulation order of the ACE strategy, the electrode closest to the base is stimulated first. The second electrode to be stimulated is the activated electrode that is next closest to the base, and so on. The stimulation order is thus high to low frequencies and most basal to most apical electrodes (Nogueira *et al.*, 2005).

For safety, the maximum rate of stimulation of the implants is 14.4 kHz (over all electrodes) and the maximum electrode stimulation rate is 2.4 kHz in the Nucleus implant (Clark, 2003; Vandali *et al.*, 2000). For consistency with the model of electrical stimulation (detailed later, in section 3.3.7), the electrode stimulation rate, R_e, was chosen to be 600 Hz. This low rate also facilitated faster simulations being carried out, as the sampling rate could be much lower than with a stimulation rate as high as 2.4 kHz. With an electrode stimulation rate as given above, the maximum implant stimulation rate, R_I, is given by equation 3.42:

$$R_I = Z_{max} \cdot R_e \ . \tag{3.42}$$

Despite the fact that up to six electrodes can be activated at a given time, this rarely occurs when a pure tone (in absolute quiet) is picked up by the implant. This happens because even some of the Z_{max} maximum envelopes may fall below the base level of the loudness growth function (LGF). Envelopes that fell below base level were assumed to not be able to stimulate

any nerve fibres and were therefore not considered in the generation of pulse trains.

Mapping envelope outputs to the loudness growth function. In Nucleus 24 implants, the output envelope magnitudes are first applied to a LGF. This is a logarithmically shaped function that maps the envelope magnitude to the an output magnitude. This LGF is used because CI users have a much smaller dynamic range than NH listeners. The LGF allows for the compression of a larger range of loudness into the smaller dynamic range. It was assumed that the same LGF was used by all electrodes. In order to match the envelope magnitudes output from the FFT filter banks with the range of the LGF, the envelopes were first scaled up by multiplying the maximum envelope magnitudes by the saturation level, M, of the LGF (equation 3.43) i.e.

$$v_{LGF}(z) = M \cdot v_0(z), \quad \text{for } z = 1, \ldots, Z_{max}\,, \tag{3.43}$$

where v_{LGF} represents the envelope magnitude that is used in the calculation of the output magnitude, ρ.

The LGF is defined by three main parameters: the base value (B), the saturation value (M) and the ε value. The base value indicates the envelope magnitude at which any lower value will produce the same output magnitude as at base level: it corresponds to a user's threshold. The saturation value indicates the envelope magnitude at which any higher value will produce the same output magnitude as at saturation level: it corresponds to a user's maximum comfort level[12]. The parameter ε dictates the steepness of the logarithmic function between these two values. The equation that defines the LGF is provided by equation 3.44 (Nogueira *et al.*, 2005).

$$\rho = \begin{cases} \dfrac{\log_{10}\left(1+\varepsilon\left[\frac{v_{LGF}-B}{M-B}\right]\right)}{\log_{10}(1+\varepsilon)}, & \text{for } B \leq v_{LGF} \leq M \\ 0, & \text{for } v_{LGF} \leq B \\ 1, & \text{for } v_{LGF} \geq M \end{cases} \tag{3.44}$$

A suitable method for determining the value of ε is provided in Cochlear Ltd (2002). A sample LGF is shown in figure 3.23.

To transform the LGF output, ρ, to a current level, the quantisation performed by the Nucleus implant has to be taken into account. First, the maximum (I_{max}) and minimum (I_{min}) current output allowed by the implant, and the number of quantisation levels, Ψ, must be defined.

[12] The threshold at which louder sounds begin to cause the CI user discomfort.

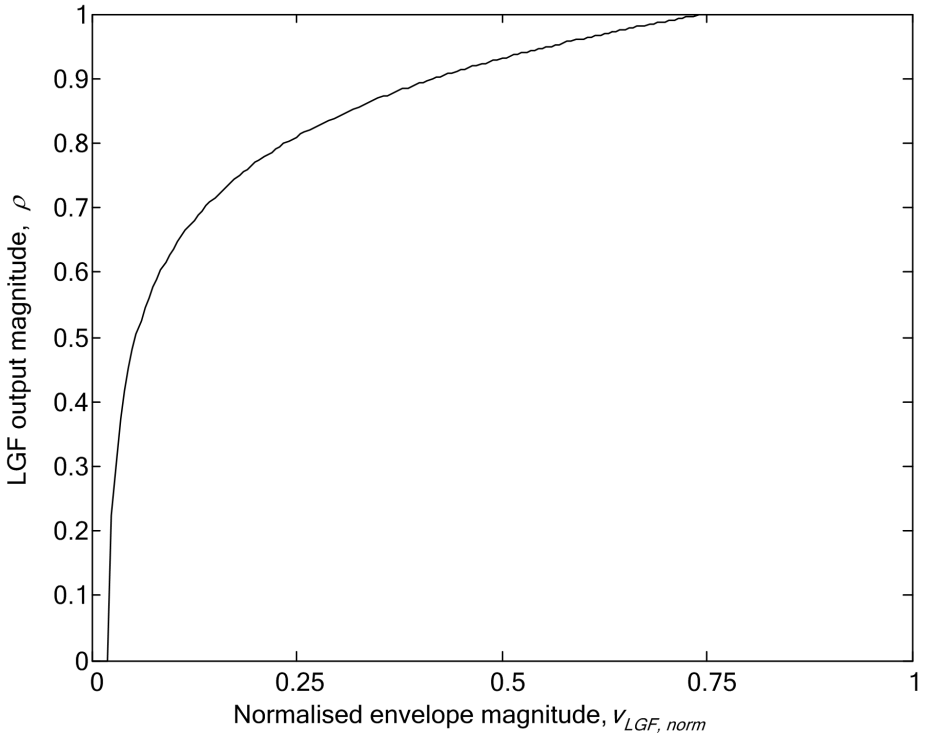

Figure 3.23.
Sample LGF, similar to the LGF used in this study. The envelope magnitude is normalised relative to the base and saturation levels, adapted from Nogueira *et al.* (2007).

These values were selected as typical values of Nucleus implants (Clark, 2003):

$$\begin{aligned} I_{max} &= 1750\mu A, \\ I_{min} &= 10\mu A \text{ and} \\ \Psi &= 256. \end{aligned}$$

The actual output current level is defined by equation 3.45:

$$I_{\psi} = I_{min} \cdot \frac{I_{max}}{I_{min}}^{(\psi/\Psi)}, \quad \text{for} \quad \psi = 0, 1, ..., \Psi - 1 \,, \tag{3.45}$$

where the output current I is in μA. With the parameters provided above, equation 3.45

reduces to equation 3.46 (Clark, 2003):

$$I = 10 \cdot 175^{(\psi/255)} \ . \tag{3.46}$$

The output of the LGF is assigned to one of the 256 levels; this is done by dividing the maximum value of ρ by 256. Each of these levels will correspond to one of the quantisation levels. The output of the LGF is rounded down to the nearest quantisation level, ψ, which is then used in equation 3.46 to determine the output current magnitude. However, this study assumes the pulse width is always constant. This is a necessary assumption imposed by the model of electrical stimulation (see section 3.3.7). A plot of the quantisation levels is given as a function of LGF output in figure 3.24.

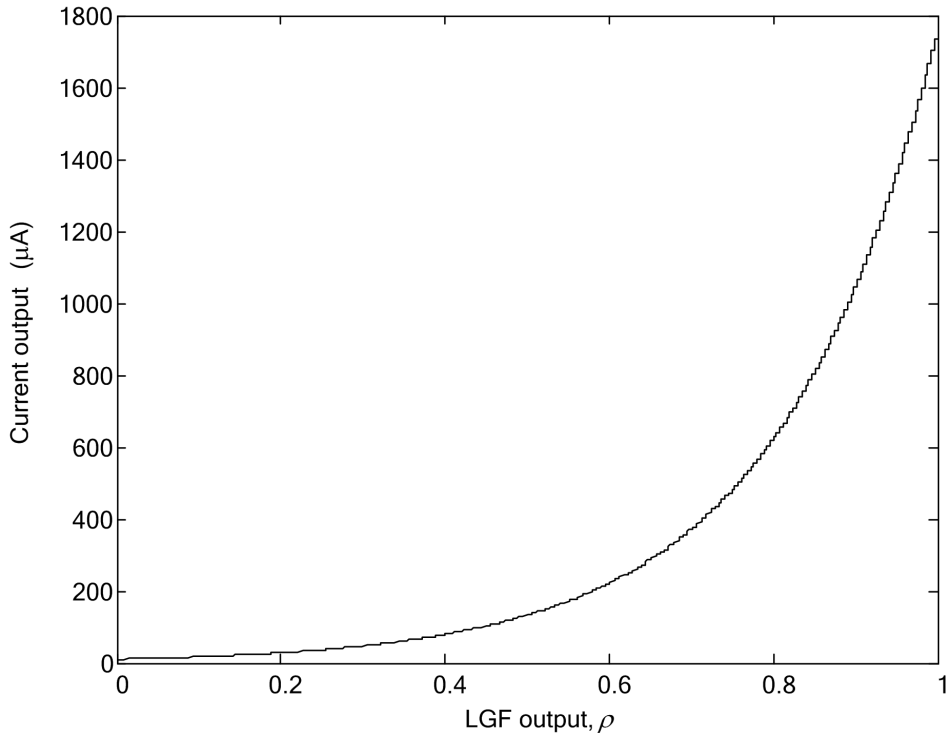

Figure 3.24.
The quantised current levels plotted against the output of the LGF. This plot is generated using equation 3.46 with 256 levels.

This is applied to each active electrode in every stimulation period. Owing to the fact that the update rate of the time windows is lower than the stimulation rate on each electrode, the same pulse level will occur on two consecutive pulses every few pulses. This repetition occurs in

the actual implementation of the processor as well (Vandali *et al.*, 2000). The model only processes pure tones, so this has negligible effect as the pulses are very repetitive as a result of the periodic nature of the signal.

3.3.6.2 Electrode array model and excitation spread

For the purposes of this study, the electrode array was assumed to be a typical Nucleus electrode array, consistent with the use of the Nucleus CI model. These arrays consist of 22 electrodes, spaced 0.75 mm apart centre to centre (Clark, 2003). The array was assumed to be fully inserted into the cochlea, with the most apical electrode situated 10 mm from the apex. This corresponds to a depth of approximately 25 mm from the base. This assumption was consistent with better performing CI subjects, as discussed in chapter 2.

Because of the fact that the model is designed to simulate the perception of CI listeners in a sound field environment, it is important to consider the phenomenon of excitation spread that occurs in multichannel CIs (see chapter 2). Only one current decay rate was considered in this study. The value of the current decay rate chosen was 4 dB/mm. This value is the same as that used by Bruce *et al.* (1999b) and O'Leary *et al.* (1995) for bipolar stimulation; however, Bingabr *et al.* (2008) describe this value as being more typical of monopolar stimulation. Hence, the value is considered typical enough to be used as an approximation for the current spread in a cochlea. The decay is considered to be linear (on a dB scale) over the length of the cochlea, following the assumption made by Bruce *et al.* (1999b). Equation 3.47 provides the relation between current and the distance from the electrode generating the current:

$$I_{stim} = I \cdot \Delta d \cdot \delta \ , \tag{3.47}$$

where I_{stim} is the current at a distance Δd from the stimulating electrode producing current I, given a decay rate of δ. Only the lateral distance between the electrodes and the nerve plane[13] is considered. Figure 3.25 illustrates the interaction between the electrode and the nerve plane.

On any given channel, the stimulating current as a function of time can be calculated by summing the constituent output currents of all active electrodes after modification of those

[13]The nerve plane refers to a hypothetical plane along which all the nerve channels in the model are considered to lie. For the purposes of this study, the plane is considered to lie parallel to the electrode array. This is illustrated by figure 3.25

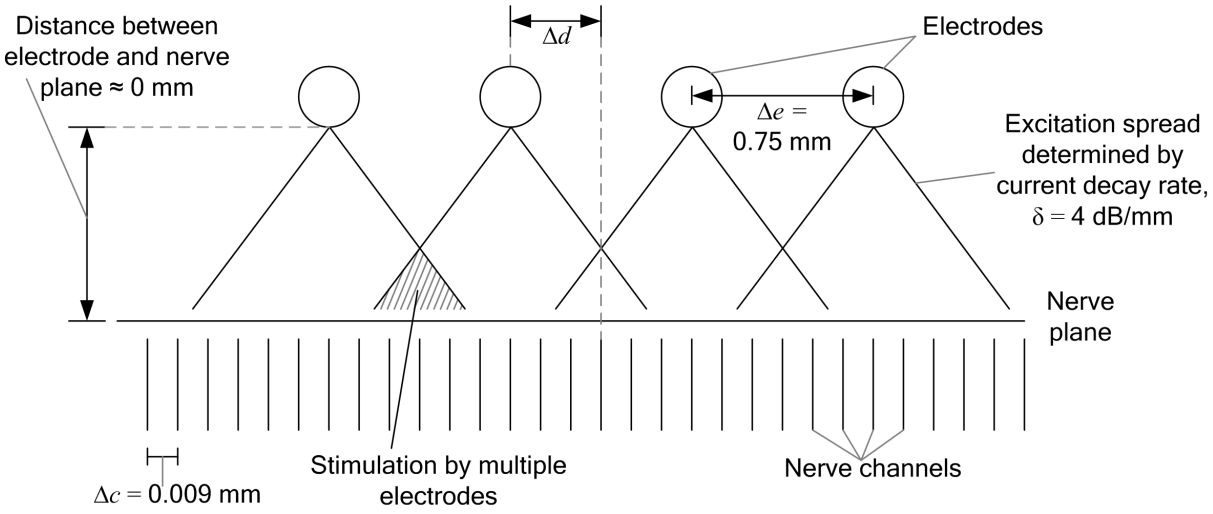

Figure 3.25.
Illustration of the excitation spread from the electrode array with some important dimensions used in the model.

currents by equation 3.47. Therefore, the equation governing the current over time for each channel in the chosen band (see sections 3.3.3.1 and 3.3.7.1) is given by equation 3.48:

$$I_{stim}(c) = \delta \sum_{z=1}^{Z} \Delta d_c(z) \cdot I(z) \ , \tag{3.48}$$

where c is the channel index. The equation shows that the sum over all electrodes (indexed by z up to a total of Z) is performed. This does not affect the assumption that only a maximum of Z_{max} electrodes can be active in a given stimulation cycle, because the inactive electrodes contribute zero current at all time samples. Figure 3.26 shows examples of the effect of excitation spread. The figure illustrates how the actual stimulation current that appears on each channel differs from the stimulus current originating from three different electrodes.

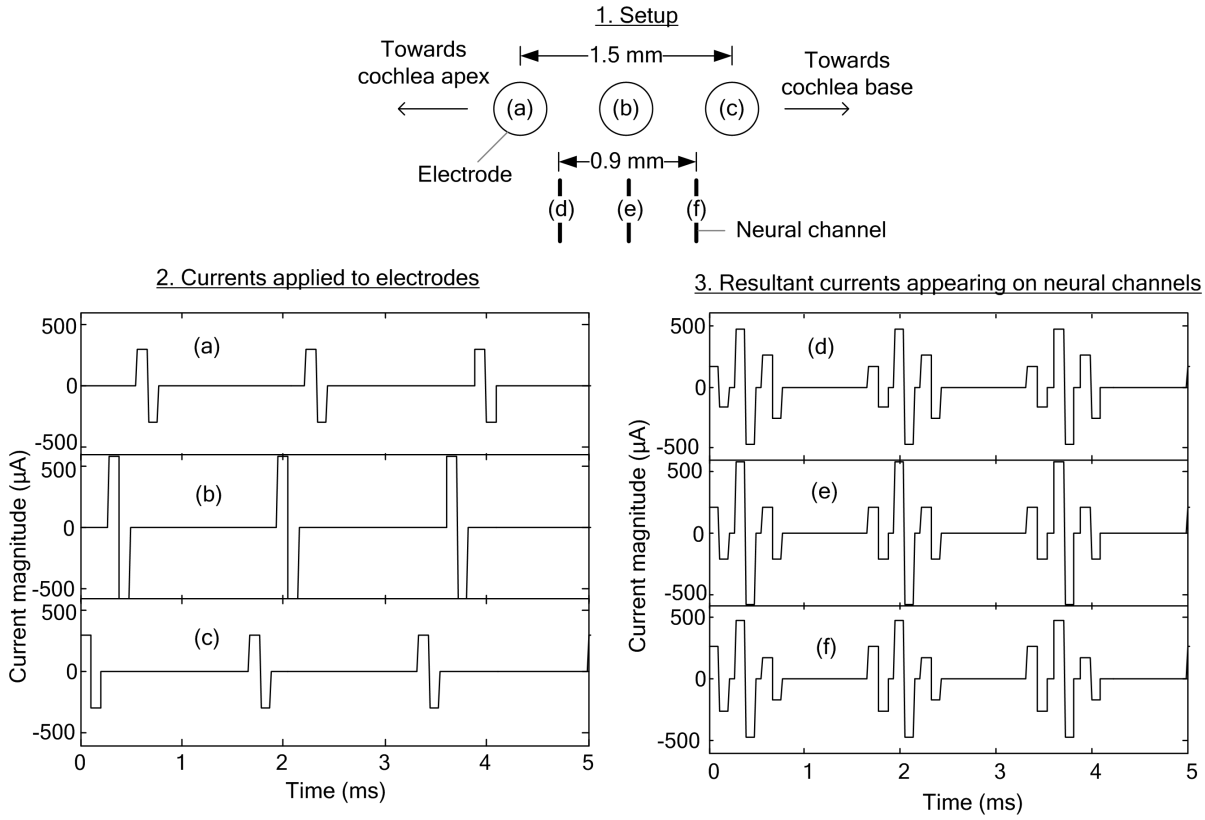

Figure 3.26.
Examples of the effect of excitation spread. The first part of the figure (1.) shows the orientation of the electrodes (denoted by (a), (b) and (c)) relative to the neural channels (denoted by (d), (e) and (f)). The second part of the figure (2.) shows the stimulating current that originates from each electrode (a), (b) and (c). The third part of the figure (3.) shows the current that each neural channel, (d), (e) and (f), experiences. The magnitudes change because of the current decay (4 dB/mm in this example) and the position of each neural channel relative to the each electrode. Channel (d) is the most apical channel about 0.45 mm from the electrode providing the highest stimulation current (b), channel (e) is the centremost channel, almost directly under electrode (b) and channel (f) is the most basal channel, also about 0.45 mm from electrode (b). In all cases, the pulse width is 100 μs/phase. Other parameters used here are shown in figure 3.25. The relative timing of the pulses is clearly visible from 2. and 3. – the stimulation rate on each electrode is 600 pps.

3.3.7 Model of electrical cochlear stimulation

Bruce *et al.* (1999a) describe a model of electrical stimulation in detail. Their stochastic computational model is approximated by an analytical model in Bruce *et al.* (2000). The analytical, renewal process model is far more computationally efficient than the stochastic computational version, but for the purposes of this study the renewal process approximation was not sufficient and the full computational model described in Bruce (1997) was used. This approach was taken because the renewal process approximation from Bruce *et al.* (2000) assumes that the pulses of electrical stimuli are of equal magnitude and are spaced uniformly apart such as the stimuli shown in figure 3.27 (a). The model implemented in this study requires that pulses of varying magnitude and non-uniform spacing be used, such as the stimuli shown in figure 3.27 (b).

Even though the full version of the model described by Bruce *et al.* (1999a) is more computationally intensive than the equivalent approximation, it is useful for this model and is indeed not the most resource-consuming component of the entire model (the rate estimator is: see section 3.3.4). The structure of the model is shown in figure 3.28 on page 85. The model shown here is based on the original in Bruce *et al.* (1999a) – Bruce *et al.* used potentials in describing the model of electrical stimulation, instead of currents. The description here is similar to that of Xu and Collins (2004; 2005; 2007).

The model works by comparing the stimulus current (I_{stim}) with the sum of three other currents ($I_{thr} + I_{noise} + I_{refr}$). This idea was based on a concept by Hill (1936), who referred to a threshold potential and a local potential. The equivalents here are the stimulus current and the summed currents respectively. If the stimulus current is equal to, or exceeds the summation of the threshold, noise and refractory currents (if $I_{thr} + I_{noise} + I_{refr} \leq I_{stim}$), neural firing occurs. This comparison is performed at every time sample in the cathodic phase of stimulation in order to determine whether excitation has occurred. Neural firing will not occur at each and every time sample as the refractory function plays a role in determining when the next spike could possibly occur. The stochastic nature of the noise current creates a variation in the precise firing times.

The component I_{stim} represents the input pulse train that appears on the nerve fibre, which could be a constant pulse train (figure 3.27 (a)) or a varied, non-uniform set of pulses (figure 3.27 (b)). In fact, the actual model does not consider both phases of stimulation, but rather only the cathodic stimulation shown in figure 3.27 (c). A key assumption to the func-

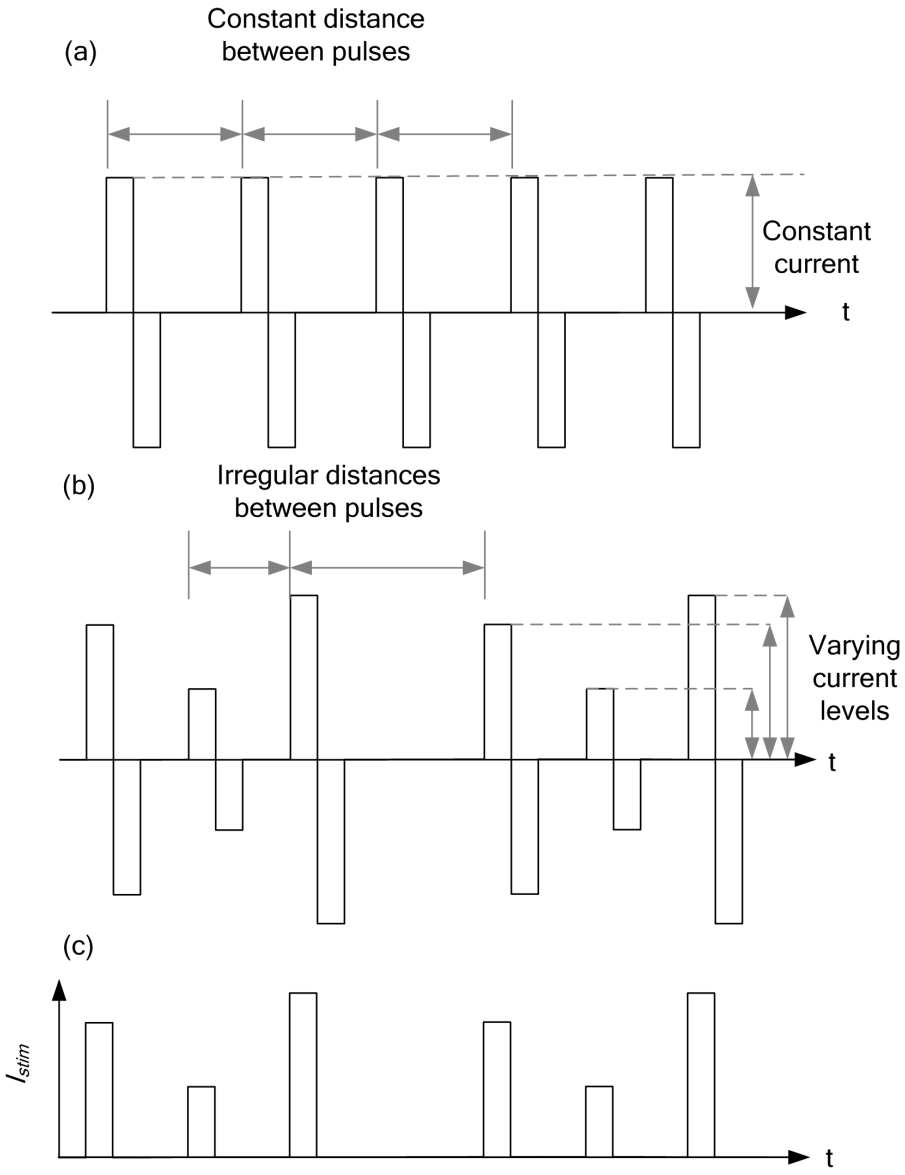

Figure 3.27.
Illustration of pulse trains on nerve fibres. (a) A pulse train that can be used by the renewal approximation version of the model of Bruce *et al.* (1999a), with constant current levels and constant distances between pulses. (b) A pulse train that can be used by the computational version of the same model, with varying pulse magnitude and time between pulses. (c) The actual representation of I_{stim}, the stimulus current, to the nerve fibres owing to the assumption that firing will only occur in the cathodic phase of the biphasic pulse (Bruce, 1997).

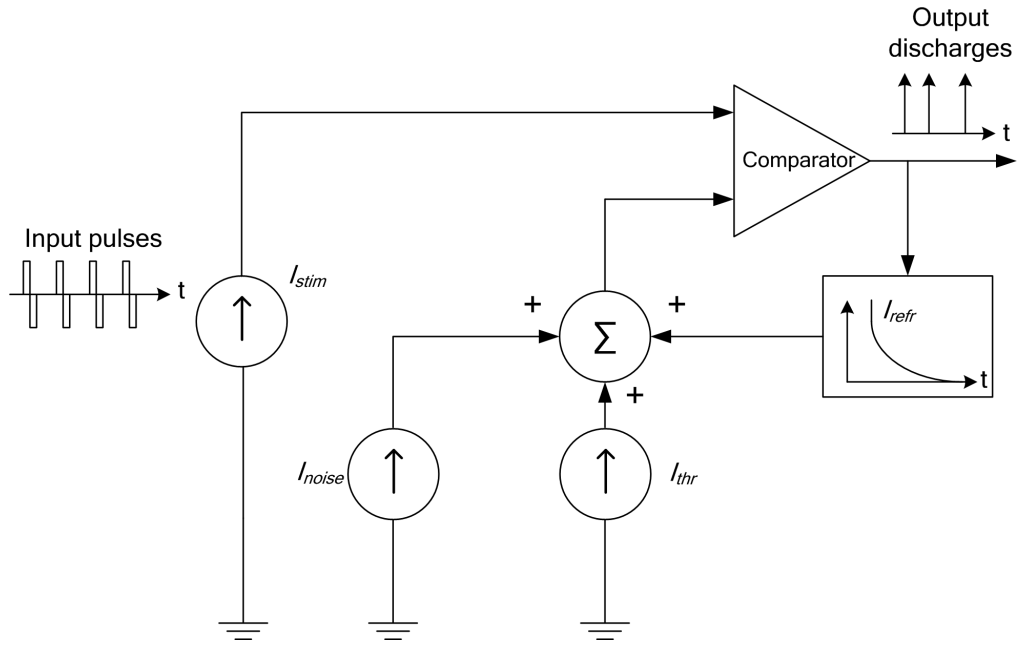

Figure 3.28.
The circuit model of the model of electrical stimulation adapted from Bruce *et al.* (1999a).

tioning in the model is that neural firing can only occur in the cathodic phase of the pulse. This is based on the fact that thresholds during the cathodic phase of biphasic stimulation are lower than during the anodic phase, and thus discharges are more likely to occur in the cathodic phase (Bruce *et al.*, 1999c). Furthermore, it is assumed that no spontaneous firing occurs (hence assuming that most inner hair cells are destroyed), and thus no firing occurs between pulses (Bruce *et al.*, 1999c).

I_{thr} is a constant value representing the threshold of the nerve fibre. As in the NH model and consistent with the work of Bruce *et al.* (1999a; 1999b; 1999c), the threshold is assumed to be the value at which the firing rate was 50% of the maximum firing rate minus the spontaneous rate. The noise (I_{noise}) is assumed to be normally distributed, with a standard deviation (σ_{noise}) related to the threshold and the RS represented by equation 3.49.

$$\sigma_{noise} = RS \cdot I_{thr} \tag{3.49}$$

The RS and therefore the noise variance are also assumed to be constant values for all neural

channels over time. The value of I_{noise}[14] added to the threshold is constant for the duration of the pulse, but is varied from pulse to pulse, following the assumptions made by Bruce *et al.* (1999a). The noise is assumed to be uncorrelated between pulses. This random noise creates variation in the firing probability of a nerve and ultimately results in a RI function similar in shape to those in figure 3.8.

Refractory effects are discussed briefly in section 2.2.2. In the models of electrical hearing, this is implemented by means of a monotonic refractory function that is added to the nerve threshold, thus increasing it, for a finite duration. This refractory function immediately comes into effect when a neural spike occurs, owing to the electrical stimulus being greater than the sum of the three additional currents (I_{thr}, I_{noise} and I_{refr}). The refractory function is provided in equation 3.50:

$$\begin{aligned} & = \infty, \text{ for } 0 \leq t \leq 0.7 \text{ ms} \\ \frac{I_{thr}+I_{refr}}{I_{thr}} & = 1+0.97e^{\frac{-(t-0.7)}{1.32}}, \text{ for } 0.7 \text{ ms} < t \leq 20 \text{ ms}, \\ & = 1, \text{ for } t > 20 \text{ ms} \end{aligned} \tag{3.50}$$

where t is the time in ms. The first line in the equation represents the absolute refractory period, where no firing can occur as a result of the refractory function imposing an infinitely high value on the threshold. The comparison between the stimulus pulses (I_{stim}) and the summed threshold, noise and refractory currents ($I_{thr}+I_{noise}+I_{refr}$) will always choose the summation as the higher value during the absolute refractory period ($I_{thr}+I_{noise}+I_{refr} > I_{stim}$), thus preventing any discharges in that time. The same absolute refractory time is used in the model of NH described in section 3.3.3.2. The second line of the equation represents the relative refractory period, where the probability of firing gradually becomes greater, until it normalises completely after 20 ms (the third line of the equation), where firing can occur without refractory effects affecting the outcome. A plot of the refractory function is provided in figure 3.29.

[14]This value is determined by a normally distributed random number generator in Matlab, "randn.m", the output of which is multiplied by σ_{noise}.

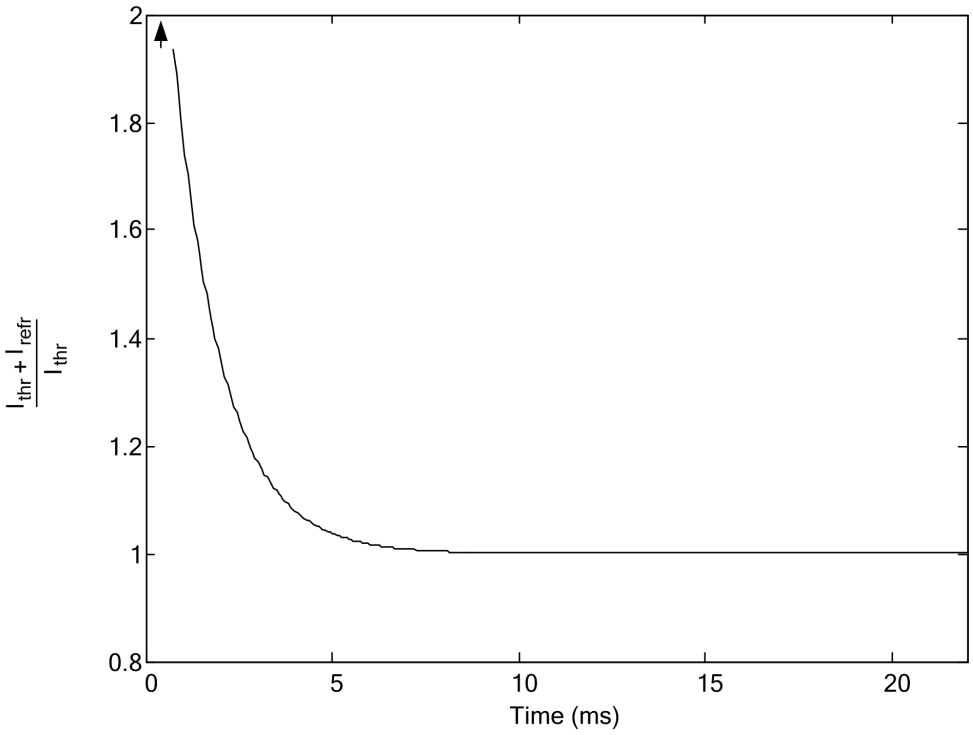

Figure 3.29.
A plot of the refractory function. The time axis shows the duration of time passed after an AP has occurred. The arrow indicates the infinity value of the function below 0.7 ms.

Bruce (1997) did not define a specific sampling frequency for his model, but rather defined the number of samples per phase. The minimum number of samples per phase required is related to the slope of the refractory function and Bruce (1997) found that increasing this number beyond 10 had a negligible effect. This standard was maintained for the purposes of this model.

Figure 3.30 shows the different currents and levels acting in the model. Note that in figure 3.30 the APs are shown to have a magnitude: this is for illustrative purposes only, as they do not hold any magnitude in the model. Three different circumstances can be observed in this figure: APs occurring due to the summed currents being lower than the stimulus current, no AP occurring as a result of the summed currents being higher than the stimulus current for the duration of an entire stimulus cycle, and an AP occurring late during a pulse as a result of the reduction in the refractory function's value over time.

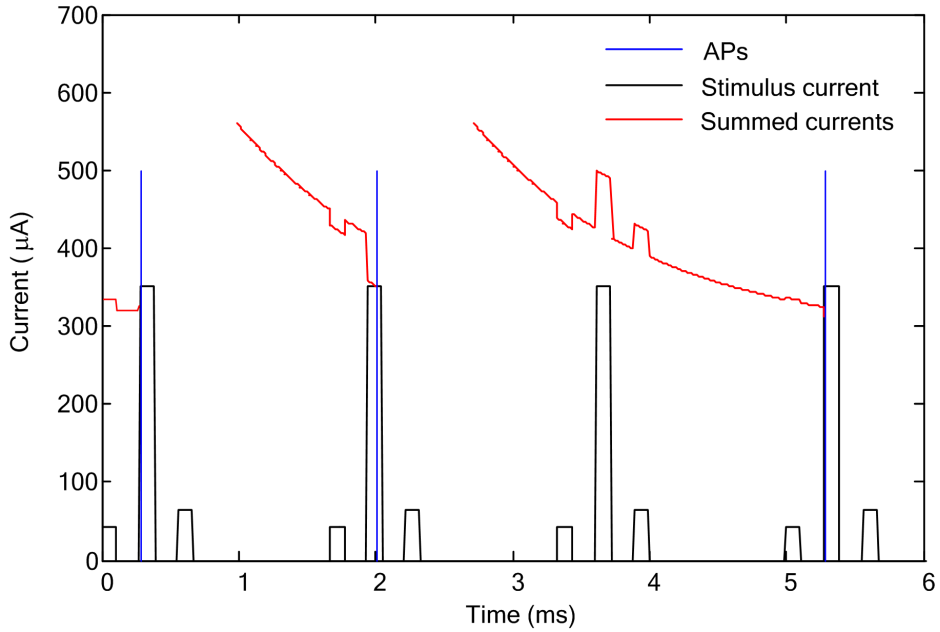

Figure 3.30.
Example of the functioning of the model of electrical stimulation. The APs are indicated by the blue spikes. The stimulus current (I_{stim}) is indicated by the black line and the sum of the currents ($I_{thr} + I_{noise} + I_{refr}$) that is compared to the stimulus current is indicated by the red lines. In the first and the last pulses that elicit an AP, it is clear that the sum of the currents would be lower than the stimulus current. In the second pulse that elicits an AP, it is clear that the transient reduction in the refractory function causes an AP to occur at a point later in the stimulus pulse than the previous two mentioned. In the third set of pulses, it can be observed that the noise current (visible by the notches in the recognisable refractory function) does not bring the summed currents lower than the stimulus current, and thus no AP can occur. Where the summed currents are not visible, the value of the refractory function is infinity. Before t = 0.2 ms, the refractory function is in steady state. The threshold of this particular fibre is 49.1 dB *re* 1μA and the RS is 0.151.

The CI model described in section 3.3.6 generates the stimulus currents for each neural channel in the critical band. The model mentioned here uses these stimulus currents to determine the spike train pattern for each neural channel. Once the AP patterns for each channel have been obtained, the rate estimator is used to estimate the instantaneous rate on each fibre at each time point, and then a decoder can be used to estimate the given percept, either frequency or intensity discrimination. Sections 3.3.4 and 3.3.5 describe how these processes are performed. The processing is performed in exactly the same way for both

normal and CI hearing, except that in CI hearing, the rate matrix is generated in a slightly different manner.

3.3.7.1 The rate matrix for electrical hearing

In section 3.3.4.1 a full description of how the rate matrix is generated for use by the rate estimator is provided. The central estimation mechanism for electrical hearing is the same as that used in NH, which implies that a rate matrix must be provided for this modality as well.

Because of the inexact mapping of electrode frequencies to the cochlea's frequency-place map (Baskent and Shannon, 2005), it is necessary to change the method of course filtering that is employed by the CANS that was described in section 3.3.3.1. This is the only difference in the central processes of the acoustic and electrical hearing models. The difference is necessary as a result of the mapping mismatch mentioned above, and also because it is not known how the critical band is selected by the auditory system; however, the processes used for selecting the ERB for each hearing modality are very similar and any inaccuracies introduced by the dissimilarities in the methods should not result in a substantial margin of error.

To choose the ERB for NH, the tonotopic location corresponding to the frequency at which the stimulation occurs is considered to be the centre of the band. For electrical stimulation, an approximation of the activated regions is made and the from this estimate the centre of the ERB is chosen.

The approximation is obtained by using the renewal process approximation of Bruce *et al.*'s model (Bruce *et al.*, 2000), i.e. the analytical version of the computational model used in this study (see also Bruce *et al.*, 2000). As mentioned before, this approximation is too restrictive to use to model electrical stimulation of the nerves induced by a CI in the sound field, but the approximation is adequate for obtaining a rough representation of the stimulated regions of the cochlea. In principle, this could have been improved by initially simulating the rate on all the electrodes and obtaining the critical band from the larger estimated rate profile, but this would have been too computationally taxing, and the difference would be insignificant.

The main equation describing the renewal approximation model is given here by equation 3.51. This does not provide a complete mathematical description of the model, but

the role that the analytical version played in the model of electrical hearing does not justify the mentioning of the extensive mathematics derived in Bruce *et al.* (2000), where the full analytical model is described. Bruce *et al.* (1999a) provide further details on the analytical model.

$$P_g(\upsilon_i|j) = \frac{1}{2}\left(1+\text{erf}\left[\frac{I_{stim}(\upsilon) - I_{thr}(\upsilon) - I_{refr}(\upsilon_i|j)}{\sqrt{2}\,\sigma_{noise}}\right]\right) \tag{3.51}$$

In equation 3.51, $P_g(\upsilon_i|j)$ is the probability that the stimulus level (I_{stim}) is greater than or equal to the refractory and noise-modified threshold level at bin i of pulse υ, given that the last neural spike occurred υ pulses previously in bin j. The bins refer to the samples within the pulses, and as mentioned above, the sampling frequency is determined by this standard. The error function (erf[x]) is given by equation 3.52:

$$\text{erf}(x) \triangleq \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt\ . \tag{3.52}$$

The approximation is obtained by using the effect of the stimulus from each individual electrode on each channel to obtain an estimate of the effect that the particular electrode has on the activation profile of the cochlea. The model of Bruce *et al.* (2000) provides a probability of firing as an output and this provides a means to determine the effect that each electrode might have if it were acting in isolation. This simulation is not computationally intensive. The maxima over all the activation profiles are determined and the COG (see equation 3.24) of the resulting maximum profile is used to determine the centre of the ERB. A sample of an activation profile is shown in figure 3.31.

Once the centre of the ERB is found, the positions of the neural channels (1 up to N_c) are determined. Figure 3.10 in section 3.3.3.1 shows this for NH and the same principle applies in this case except that the centre is determined by the means described above.

The rate matrix is then determined by exactly the same methods as described in section 3.3.4.1 for NH, the fundamental difference between the two hearing modalities being that the ERB could possibly be centred around a different point for the same stimulus frequency. Both normal and electrical hearing subject models use a rate matrix that is determined using the RI function model for NH listeners.

The parameters that generate the rate matrix are kept consistent with the parameters used

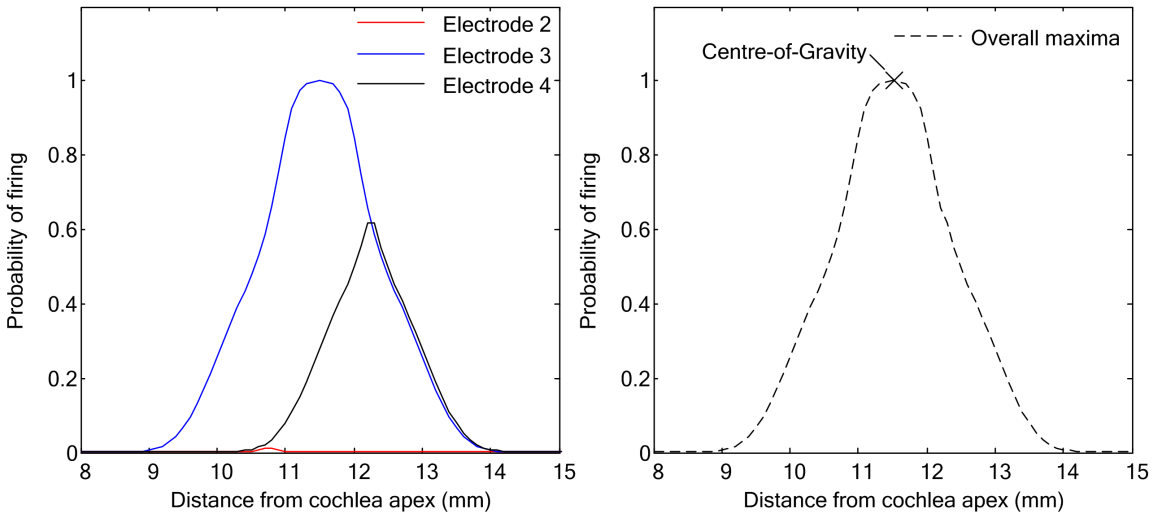

Figure 3.31.
Example of the activation profile used for finding the centre of the ERB used in electrical hearing. The figure on the left shows the firing probabilities of the stimulating electrodes, calculated as though each electrode is the only stimulating electrode. The figure on the right shows the trace around the maxima of the three activation functions, from which the COG (indicated by the large cross) is calculated and used as the centre of the ERB. The input frequency here is 531 Hz, the stimulation rate is 600 pps with a pulse width of 100 μs/phase and the neural threshold is 49.1 dB *re* 1μA with an RS of 0.151. The current decay rate is 4 dB/mm.

as input to the CI model. If the input intensity is 70 dB SPL, the rate vs intensity model will utilise that sound pressure in generating the rate matrix. The same analogue is applied to the threshold levels, i.e. if an electrical threshold of 49.1 dB *re* 1 μA is used, then a threshold of 49.1 dB SPL is used in generating the rate matrix. Although this means that the stimulating current and the internal representation of the signal will not necessarily be matched, there is no way to match current with a corresponding SPL. This is because no data is available from which such information can be obtained. Binaurally deaf subjects can no longer experience normal auditory sensation and are not able to match electrically induced sensation to acoustically induced sensation. It is therefore assumed that there is a one-to-one correspondence between dB *re* 1 μA and dB SPL.

The approach used in this study follows from the fundamental assumption made by the model that the CANS is blind to the type of input that it receives, and assumes that the modelled CI subject learnt to hear using auditory inputs, as is the case in post-lingually deafened

adults. This assumption was based on the fact that much of the work in the literature uses post-lingually deafened adults, who had auditory abilities at some point in their lives before cochlear implantation, as subjects (e.g. Pretorius and Hanekom, 2008; Drennan and Pfingst, 2005).

Once the rate matrix for the electrical hearing modality has been generated (whether it be for frequency or intensity discrimination), it is applied to the rate estimator, as described in section 3.3.4. Hence, the outputs of the rate estimator are the same for electrical and acoustic hearing. Similarly, the decoders are applied to the rate estimator output to finally provide an estimate of the percept threshold that is being tested, either the JND in frequency or the JND in intensity.

3.4 SUMMARY

In this chapter, the methods that were used in the development of the model were explained. The systems used in the implementation of the model were described in detail. This model is used in the following two chapters to predict the outcomes of classic perceptual experiments involving the discrimination of frequencies and intensities in sound field environments. In chapters 4 and 5, these predictions are compared with data found in literature and the similarities and discrepancies are noted.

CHAPTER 4

FREQUENCY DISCRIMINATION

4.1 CHAPTER OBJECTIVES

In this chapter, the model presented in chapter 3 is used to predict the outcomes of frequency discrimination experiments for NH and CI listeners. This chapter aims to provide insight into the frequency discrimination data prediction capabilities of the model. The chapter begins by presenting all of the parameters that were used to generate the results, along with the justifications for choosing those parameters. This is followed by a presentation of frequency discrimination results for NH listeners, where the model-predicted results are shown alongside data from experiments and models in literature. The model-predicted results for CI listeners are then presented in conjunction with the available literature data. Finally, a discussion of the presented results considers the implications of the model and the data that it yields.

4.2 METHODS

The majority of the methods that were used to generate the results are provided in chapter 3, but the specific parameters used are expounded here. Any deviation from the standard parameters are indicated in the text were applicable. The parameter naming convention used here is kept consistent with the parameter names used in chapter 3. The tables that follow

provide the parameters used in each section of the model. Justifications and origins of the parameters are provided after each table.

4.2.1 Normal hearing parameters

The parameters relevant to NH are presented in table 4.1. These parameters are only pertinent to the peripheral part of the NH model; the central processor's parameters are listed later and are standard for both acoustic and electrical hearing.

Table 4.1.
List of standard parameters used for simulations of NH listeners.

Parameter	Description	Value
Input signal		
f_{in}	Input frequency	1000 Hz
I_{in}	Input intensity	70 dB SPL
T	Input signal duration	200 ms
Rate vs intensity model		
A_0	Spontaneous Rate	1 spike/s
A_1	Maximum rate − spontaneous rate	199 spikes/s
$A_{2,\text{dB SPL}}$	Threshold	49.1 dB SPL
α_0	Base α value	4
f_t	Tuning curve change frequency	800 Hz
Cochlear dimensions		
F_{apex}	CF at cochlear apex	20 Hz
F_{base}	CF at cochlear base	20 000 Hz
l_C	Total length of Cochlea	34.761mm
Δc	Channel separation distance	9 μm
l_{ERB}	Equivalent rectangular bandwidth	0.9 mm
N_c	Number of neural channels per ERB	100
Acoustic stimulation model		
a_d	Absolute refractory period	0.7 ms

4.2.2 Cochlear implant hearing parameters

The parameters for CI listeners follow in table 4.2. Some of these parameters overlap with and are the same as those used for NH and have therefore not been repeated.

Table 4.2.
List of standard parameters used for simulations of CI listeners.

Parameter	Description	Value
Input signal		
T	Input signal duration	500 ms
CI model		
F_s	CI acoustic sampling frequency	16 000 Hz
L	Number of samples per time window	128
F_a	Time window analysis rate	500 /s
Z	Number of electrodes in cochlea/FFT filters	22
Z_{max}	Maximum number of stimulating electrodes	6
R_e	Stimulation rate of each electrode	600 pps
w	Stimulating pulse width	100 μs/phase
δ	Current decay rate	4 dB/mm
Δe	Electrode spacing	0.75 mm
x_{e_1}	Distance from apex of most apical electrode	10 mm
Electrical stimulation model		
I_{thr}	Fibre current threshold	49.1 dB *re* 1 μA
RS	RS of fibre	0.151

4.2.3 Central estimator parameters

The parameters for the central estimator include those used for the Viterbi decoder and the rate estimator. As mentioned before, these parameters, provided in table 4.3, are common to both acoustic and electrical hearing.

Table 4.3.
List of standard central estimator parameters used for simulations of NH and CI listeners.

Parameter	Description	Value
Rate estimator		
ω_i	Transition intensities	0.5 /s
K	Number of states per channel	100
Viterbi decoder		
π	Transition probabilities between states	0.01
σ_0	Observation probability variance	10

4.2.4 Parameter justification

All of the free parameters, and a few set parameters in the model, are listed in tables 4.1, 4.2 and 4.3. Justifications for the parameter choices are provided below.

4.2.4.1 Input signal

The properties of the input signal were largely based on the stimulus signals used in other studies. In Heinz *et al.* (2001a), 1000 Hz is used in the frequency discrimination experiments where frequency is not the variable parameter. Sek and Moore (1995) used tones of 200 ms duration at a level of 70 dB SPL. This level was maintained for the CI model, but the signal duration was increased to 500 ms to be consistent with the study of Pretorius and Hanekom (2008).

4.2.4.2 Rate vs intensity model

All the fibres in both hearing modalities were assumed to have static, flat-saturating RI functions for the sake of simplicity. Variability in these properties would have confounding effects on the results. The spontaneous rate and maximum rate were chosen based on the model of Hanekom (1999), where a low spontaneous rate (1 spike/s) and relatively high maximum firing rate (200 spikes/s) were used. Both values are realistic neural parameters (Sachs and Abbas, 1974). These values are representative of a large range of possible firing rates.

The threshold of the fibres was chosen to be as consistent as possible with the model of electrical hearing; therefore, 49.1 dB SPL was used. However, as mentioned in section 3.3.7.1, dB SPL and dB *re* 1μA are assumed to correspond, but this is not necessarily the case. The values relating to the tuning curves are sourced directly from the model of Colburn (1973) in which an α_0 value of 4 and a f_T of 800 Hz, were used.

4.2.4.3 Cochlear dimensions

The cochlear dimensions were based on the average physiological cochlear dimensions of humans. According to Hartmann (1998), the range of frequencies that the average human can perceive is 20 Hz to 20 kHz. When these values are applied to Greenwood's equation (equation 3.8), the length of the cochlea, from apex to base, can be found to be 34.761 mm.

The channel spacing was based on the fact that low spontaneous-rate fibres were used. According to Shofner and Sachs (1986), such fibres account for 15% of the afferent nerve population of about 28 000 (Kim, 1984). This yields a fibre spacing of nearly 9 μm, following that used by Hanekom (1999), where a similar model was used. This spacing was also necessary owing to the increased computational requirements incurred with a higher number of channels, and given the ERB size of 0.9 mm, this would mean 100 channels are simulated. If the length of all ERBs over the length of the cochlea are obtained by using the equation of Glasberg and Moore (1990) (equation A.1 in appendix A) and Greenwood's equation (equation 3.8), the approximate average ERB over the length of a human cochlea is 0.9 mm. This value was also cited in a review by Clark (2003), therefore 0.9 mm was chosen as the constant ERB length.

4.2.4.4 Acoustic stimulation model

The absolute refractory period for the acoustic stimulation model was taken directly from the model of electrical stimulation, i.e. the assumption by Bruce *et al.* (1999a) that the absolute refractory time of the auditory neurons is 0.7 ms. The other parameters that relate to this part of the model can all be calculated using this value and the maximum firing rate (see the equations in section 3.3.3.2).

4.2.4.5 Cochlear implant model

Most of the parameters for the CI model were sourced from typical parameters provided by Nogueira *et al.* (2005) and Vandali *et al.* (2000). In the Nucleus 24 CI with a SPrint processor using the ACE speech processing strategy, the audio sampling rate is always 16 kHz. The time windows consist of 128 samples and are therefore 8 ms long. These 8 ms windows overlap one another by 96 samples (6 ms), which results in a time window being analysed every 2 ms, i.e. at a rate of 500 windows per second.

The modelled CI can support a maximum of 22 FFT filters, all of which were used to supply the model with as much information as possible, over the largest range possible. The number of active electrodes in a stimulus cycle was chosen to be six because this is sufficient to represent any pure-tone signal over the entire frequency range at high stimulation levels.

The pulse width and pulse rate (100 μs/phase and 600 pps respectively) were based on the values used in the model of Bruce *et al.* (1999a), which yielded the nerve fibre threshold and RS that was used. The current decay rate of 4 dB/mm was used by Bruce *et al.* (1999b) to model bipolar current decay based on the work of O'Leary *et al.* (1985). This value is also consistent with values given by Bingabr *et al.* (2008).

A Nucleus implant electrode array consists of electrodes with centre-to-centre distances of 0.75 mm (Clark, Patrick and Bailey, 1979), which is the reason why this value was used for the spaces between electrode centres. Such an array is normally inserted up to approximately 25 mm into the cochlea, which corresponds to a distance of approximately 10 mm between the first electrode and the apex (Dorman *et al.*, 1997a).

4.2.4.6 Electrical stimulation model

The electrical stimulation model parameters were taken from the properties of one of the neurons studied by Javel *et al.* (1987), neuron 3-21. This particular neuron was used extensively in the work of Bruce *et al.* (1999a; 1999b; 1999c) and Bruce (1997) to demonstrate the use of the electrical stimulation model. The threshold of 49.1 dB *re* 1μA and the RS of 0.151 corresponds to the pulse width of 100 μs/phase according to the above-mentioned studies.

4.2.4.7 Rate estimator

There was no apparent source data from which to draw for the rate estimator and as a result, the transition intensities were chosen to be 0.5 so that the rate of change of the estimated rate was not very erratic. A very high transition intensity would have resulted in swift changes in the estimated rate in 1 ms which, given the integration time of the brain, seems unlikely (approximately 5 ms integration time is the norm according to Viemeister and Wakefield, 1991). The number of states was simply set to be equal to the number of channels, so that a fair representation of the states was provided to the estimator, traded off with the simulation time required for a large number of states.

4.2.4.8 Viterbi decoder

The probabilities of each state occurring in the Viterbi algorithm (as well as the initial states of the rate estimator) were set to be equal for all states, thus removing any particular bias to a specific state. This yields a 0.01 probability for 100 states. The observation probability variance was set to be 10. This value is large enough to provide a low probability of the rate differing substantially from the internal representation. This implicitly assumes that the internal model (λ-matrix) provides a reliable representation of the incoming signal.

4.3 RESULTS

Using the model described in chapter 3, and the parameters detailed above, predictions of the frequency discrimination abilities of NH listeners and CI listeners were obtained. The results from the NH listeners are used as a benchmark by which the reliability and efficacy of the model are judged, i.e. the model must work for NH data before it can be used to predict CI data. It is for this reason that the majority of results are generated for NH. The model used in this study is from here on referred to as the instantaneous rate estimator model, or simply the IRE model. Only objective observations of the results are provided in this section. Reasoning for the result origins and insights into the model's performance are provided later, in the discussion section of this chapter (section 4.4).

4.3.1 Normal hearing frequency discrimination

There is much data available on NH frequency discrimination in literature for comparative purposes, as discussed in chapter 2. This study will present similar results to those presented in Heinz *et al.* (2001a), namely frequency discrimination as a function of frequency, intensity and duration of the input signal. These three fundamental parameters are often measured in studies using human subjects, and provide a reliable indicator of the generality of a model.

4.3.1.1 Frequency discrimination as a function of frequency

The results for modelled, NH frequency discrimination as a function of frequency are provided on the next page in figure 4.1. The modelled results using both the COG decoder and the Viterbi algorithm decoder are compared to results from literature, including those of a similar model (Heinz *et al.*, 2001a) as well as measured human data (Moore, 1973; Sek and Moore, 1995).

It can be observed from figure 4.1 that the results of the model using the COG decoder are substantially lower than those of observed human data, and the bowl-shaped trend noticeable in human experimental data is quite apparent in the results of the IRE model, although the bowl is narrower. When the frequency at which the tuning curves begin to narrow is set to 2000 Hz, the shape is observable for the same range of frequencies as those in Heinz *et al.* (2001a) for their all-information model. In both cases, the performance of the IRE model with a COG decoder is marginally closer to that of human experimental performance than that of Heinz *et al.* (2001a), but is similar to their data in magnitude. The similarity in trend is maintained up to approximately 5000 Hz (when f_T is 800 Hz), at which point the trend begins to deviate from the expected continuation of the increase. This deviation is absent in the COG results for f_T = 2000 Hz over the frequency range tested.

The Viterbi decoder results are substantially different from those of the COG decoder. When using the Viterbi algorithm decoder, the bowl-shaped trend is still observable for up to 3000 Hz, and the values of the discrimination thresholds are in the region of those seen in the experimental data of Sek and Moore (1995) and Moore (1973). After the 3000 Hz point, the regular trend degrades and begins to become erratic. This is similar to the degradation in trend seen with the COG decoder, but it occurs at a much lower frequency, and the variation in value appears to be higher, too. When a gradually narrowing ERB (according to

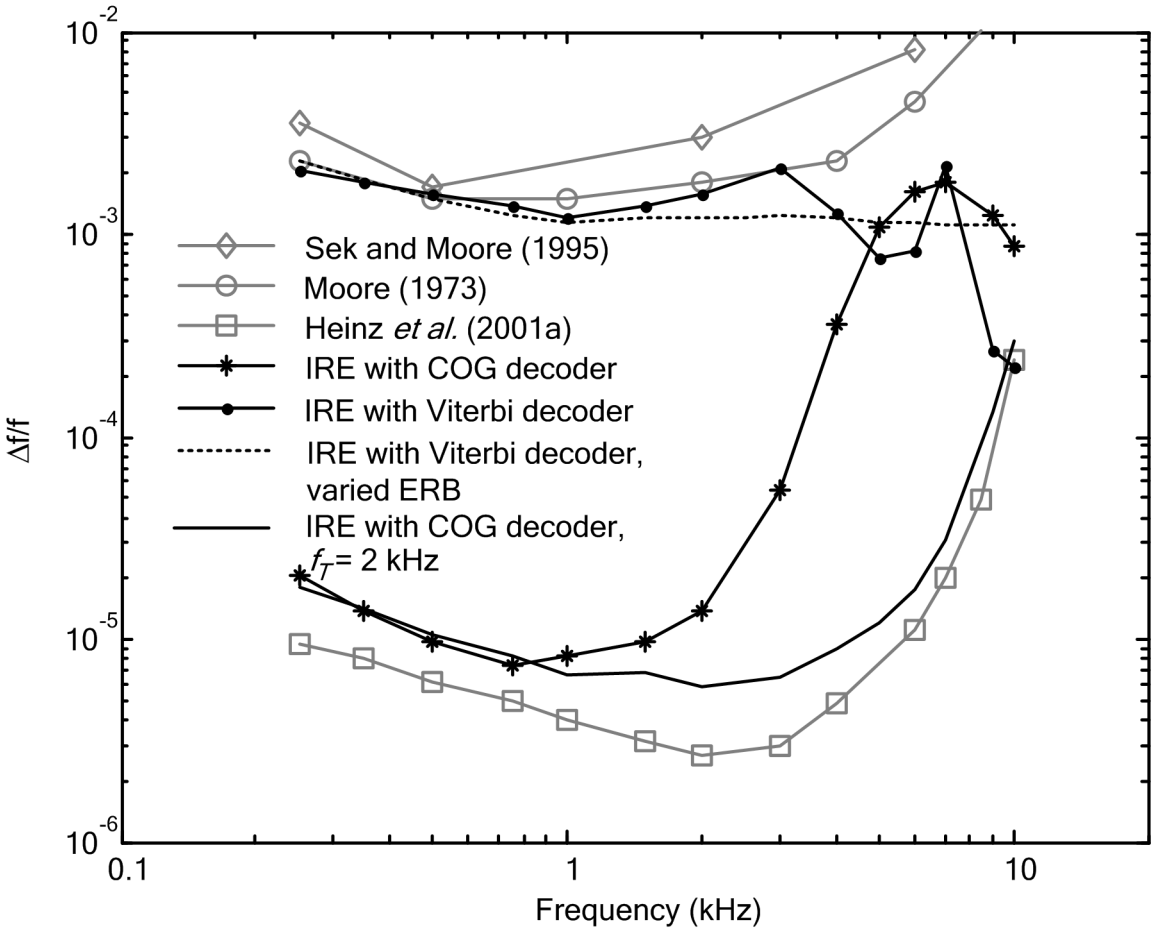

Figure 4.1.
Frequency discrimination data for NH listeners as a function of frequency. Data from literature is shown by the grey plots representing data from Sek and Moore (1995) (diamonds, 70 dB SPL, 200 ms), Moore (1973) (circles, equal loudness across frequencies, 60 dB SPL at 1 kHZ, 200 ms) and Heinz *et al.* (2001a) (squares, 40 dB SPL, 200 ms). The data from Sek and Moore (1995) and Moore (1973) is experimental human data, while that of Heinz *et al.* (2001a) is from their all-information model. Data from the model used in this study is indicated by the black plots (70 dB SPL, 200ms). The two sets of data using the COG decoder differ in the values of f_T used (see section 3.3.2.2); stars show the standard 800 Hz and the unmarked line shows when f_T is 2 kHz. The Viterbi algorithm data using standard parameters is indicated by the solid line punctuated by dots. Finally, the dotted line shows IRE model data using the Viterbi decoder, except that the ERB changes according to equation A.1 and the α value is set to be constant after 2 kHz. The data for this study is the mean of 50 simulations.

equation A.1) across frequency is utilised and tuning curve sharpening is stopped at 2000 Hz ($\alpha(f_{CF} > 2kHz) = \alpha(f_{CF} = 2kHz)$), the resulting frequency difference limens flatten out (as indicated by the dotted line in figure 4.1).

The IRE model for NH adequately predicts the frequency discrimination trends as a function of frequency found in experimental data for lower frequencies (Moore, 1973; Sek and Moore, 1995) when using both the COG decoder and the Viterbi decoder. However, the magnitudes are far lower than those obtained in the literature data when using the COG decoder. The IRE model can be set to yield similar values to optimal detection models of frequency discrimination (Heinz *et al.*, 2001a) if the value of f_T is set to 2000 Hz.

The behaviour of the results appears to depend to a large extent on the tuning curves and the ERB. Figure 4.2 shows the resulting firing rate profiles at different frequencies for the standard parameters. It is clear that the tuning is broadest at 1000 Hz, and gradually narrows as the frequency is increased. Eventually, the profile begins to include very low firing rates (at 10000 Hz). Over all the frequencies, the peak of the activation is centred on the middle fibre in the band, but as the tuning changes with frequency, the degree of symmetry around this point varies. For comparative purposes, an excitation pattern for low spontaneous rate fibres is shown in figure 4.2 (the 1500 Hz data from Shofner and Sachs, 1986). The excitation patterns from the model are similar to the data from literature.

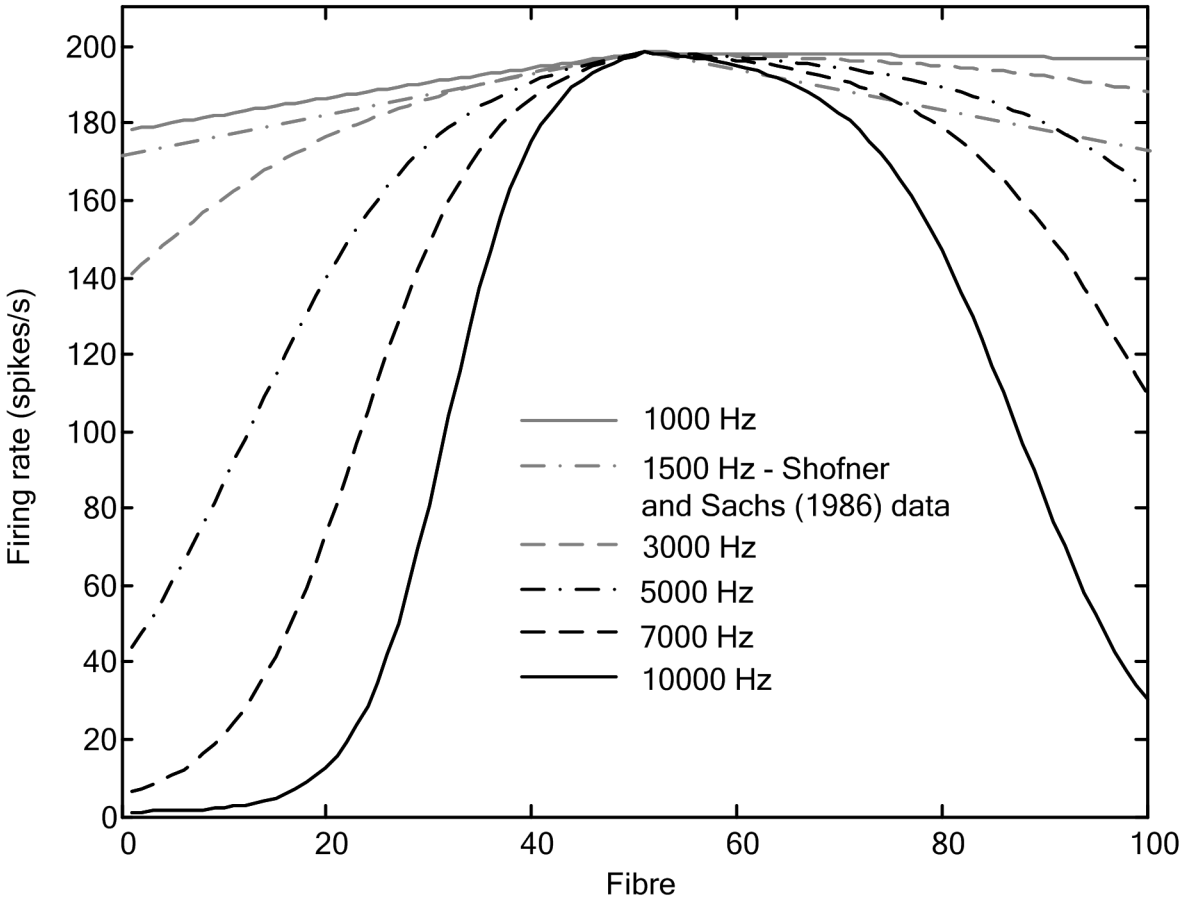

Figure 4.2.
The spatial firing rate profiles for different frequencies shown for each of the fibres within the 0.9 mm ERB of each frequency. The data for 1000 Hz, 3000 Hz, 5000 Hz, 7000 Hz and 10000Hz were used in the generation of the firing patterns for the modelled data in figure 4.1 and they form a part of the λ-matrices used in the rate estimation and decoding of the signal generated at each of the given frequencies. These are analogous to excitation patterns generated for each frequency. The excitation pattern for 1500 Hz is taken from Shofner and Sachs (1986) and is normalised relative to the maximum firing rate of the other excitation patterns shown here: approximately 200 spikes/s. In all cases, the data is for low spontaneous rate fibres and the middle fibre corresponds to the fibre with a CF equal to the stimulating frequency.

4.3.1.2 Frequency discrimination as a function of intensity

Human frequency discrimination ability as a function of intensity has been tested for NH listeners (Wier *et al.*, 1977) and is one of the parameters analysed by Heinz *et al.* (2001a). The modelled results for this experimental parameter are shown in figure 4.3. For these results, the threshold of the nerve fibres was set to be lower than the standard described in section 4.2.1. This was done so that a larger range of intensities could be tested. If the standard threshold were used, the values far below the threshold would not yield predictions, as no firing would be elicited by the stimuli.

Figure 4.3 shows the comparison of the modelled results with the results from literature. The Viterbi decoder results come very close in magnitude to the experimental data from Wier *et al.* (1977), and the trend is similar: an initial decrease in the frequency discrimination ability can be observed, followed by a plateauing of the ability at a certain input intensity level (around 40 dB SPL). The magnitude is slightly higher than the values found by Wier *et al.* (1977) but not substantially so. Relative to the results generated by the model of Heinz *et al.* (2001a), the Viterbi decoder did not show discrimination thresholds as low in magnitude, but exhibited a similar trend.

The COG decoder also exhibited a trend similar to the data of Wier *et al.* (1977), but was also much lower in magnitude than the values found experimentally, i.e. the COG decoder performed better than humans. This is also the case with Heinz *et al.*'s model (Heinz *et al.*, 2001a), although the COG does not perform as well as the all-information model that Heinz *et al.* proposed. In the case of both decoders, the largest change in the model's magnitudes occurs around the neural threshold. The plateau in the results is slightly different to the gradual decrease observed in the experimental data by Wier *et al.* (1977).

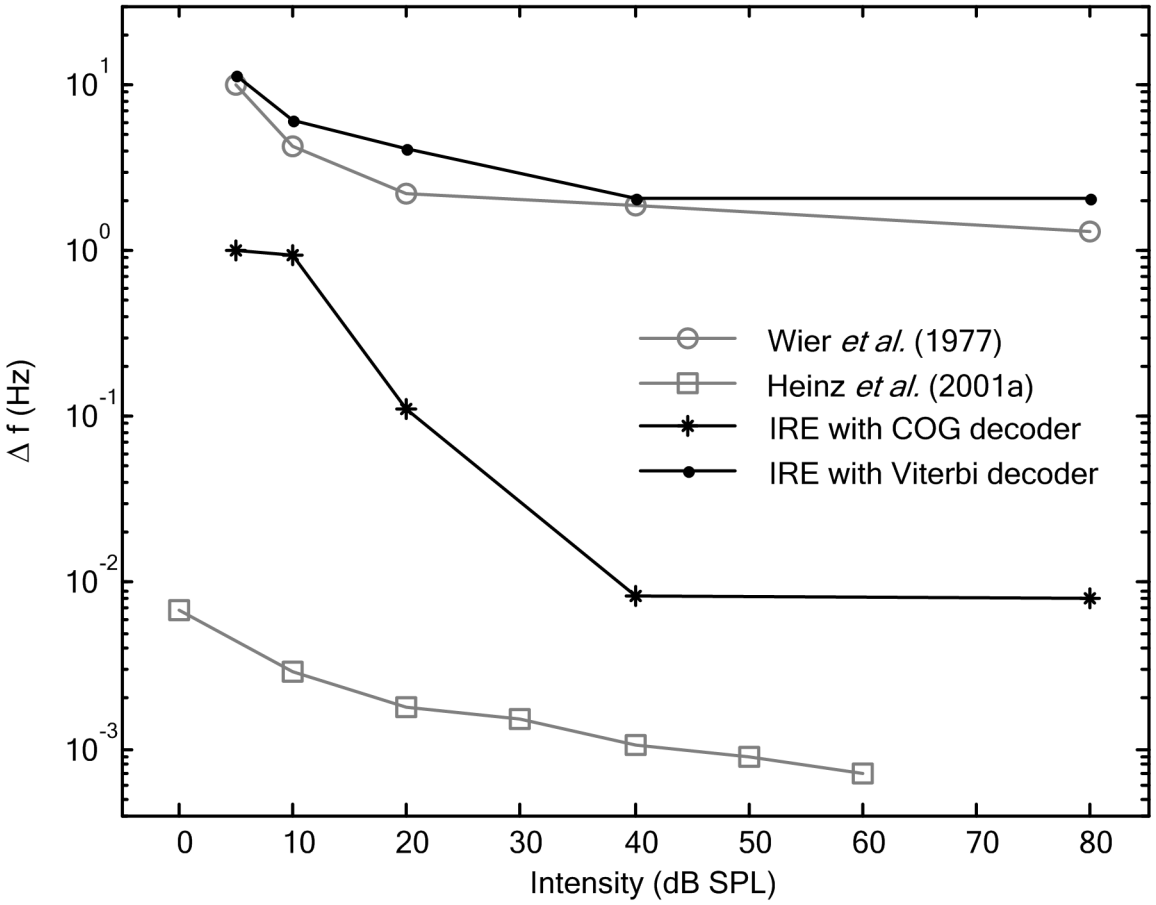

Figure 4.3.
Frequency discrimination data for NH listeners as a function of stimulus intensity. Data from literature is shown by the grey plots illustrating data from Wier *et al.* (1977) (circles) with an input frequency of 1000 Hz for 500 ms, and data from Heinz *et al.* (2001a) (squares) at 970 Hz for 500 ms. Data from the model used in this study is indicated by the black lines at 1000 Hz for 500 ms, showing the Viterbi decoder results (dots) and the COG decoder (stars) results. The data for this study is the mean of 50 simulations, with a lower threshold of 10 dB SPL used instead of the standard 49.1 dB SPL.

4.3.1.3 Frequency discrimination as a function of duration

The frequency discrimination thresholds for NH listeners were also tested as a function of frequency. The results from the simulations are shown in figure 4.4, compared with data from Moore (1973) and Heinz *et al.* (2001a).

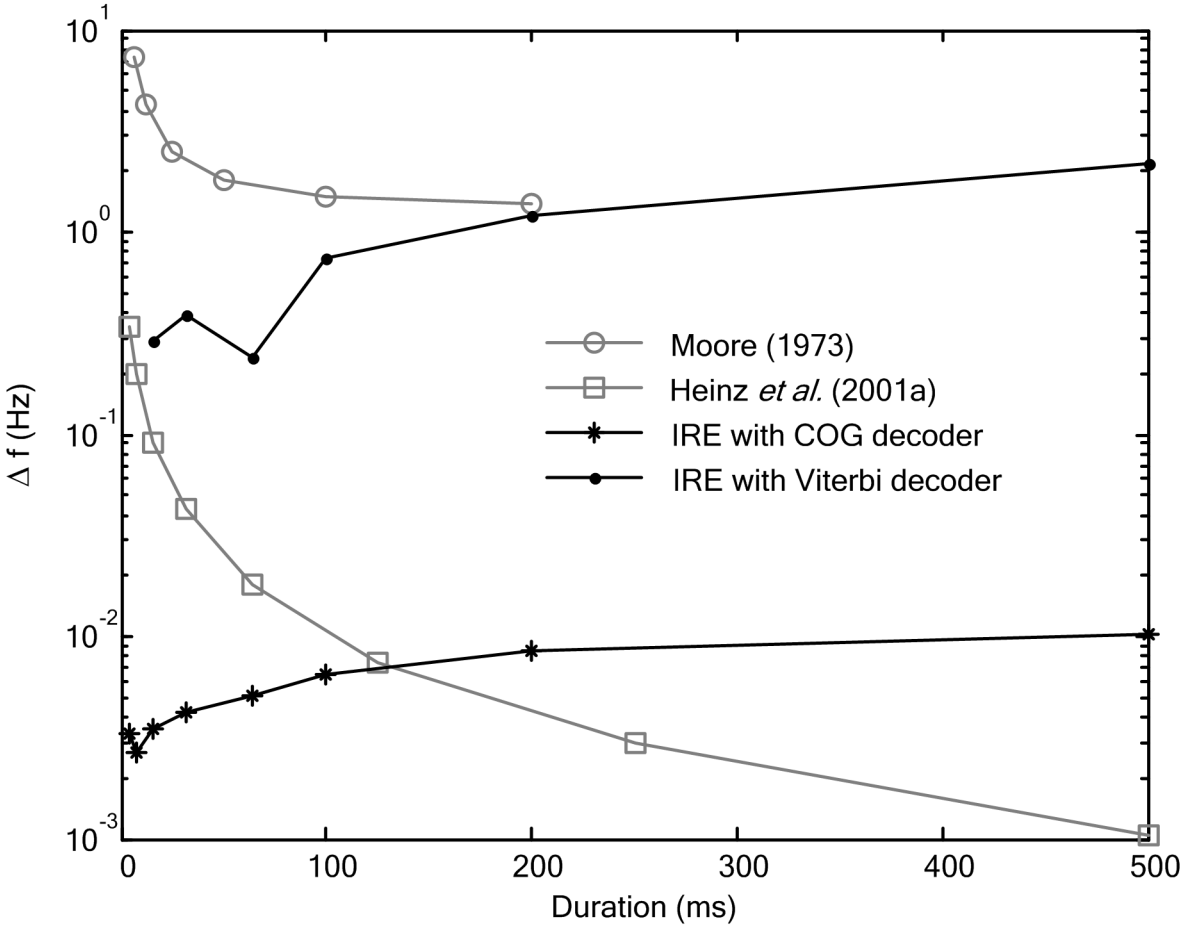

Figure 4.4.
Frequency discrimination data for NH listeners as a function of duration. Data from literature is shown by the grey plots, illustrating data from Moore (1973) (circles) with an input frequency of 1000 Hz at 60 dB SPL, and data from Heinz *et al.* (2001a) (squares) for 970 Hz at 40 dB SPL for experimental human data and all-information model data respectively. Data from the model used in this study is indicated by the black lines for 1000 Hz at 70 dB SPL, showing the Viterbi decoder results (dots) and the COG decoder results (stars). The data for this study is the mean of 50 simulations.

It is clear from figure 4.4 that the trends obtained using the IRE model are not similar to those in literature. Both decoders show a similar trend: an initially low value grows up to a point where it begins to plateau in value. In the case of the Viterbi decoder, the plateau appears to be in the region of the plateau of the NH experimental data, which allows the model to be useful at longer signal durations (200 ms and longer). In the case of the COG decoder, the plateau is higher than the value provided by an all-information model for longer durations,

but in the regions being tested for this study (around 200 ms) the data is comparable to Heinz *et al.*'s modelled data (Heinz *et al.*, 2001a). The discrepancy between the two sets of data is discussed in section 4.4.1.

4.3.2 Cochlear implant frequency discrimination

Given that the model of NH frequency discrimination performed adequately, the next step in testing the hypothesis was to predict the outcomes of CI frequency discrimination experiments. In order to accomplish this, the model described in chapter 3 was used with the parameters listed in section 4.2.2. The frequency discrimination values as a function of frequency are provided in this section, where they are compared with experimental results taken in a sound field environment by Pretorius and Hanekom (2008), as well as results from NH listeners. It is somewhat difficult to draw comparisons between frequency discrimination ability that is studied in the sound field and existing studies of frequency discrimination using electrical stimulation, as the majority of studies have investigated pulse rate (rate discrimination, Barretto and Pfingst, 1992) or single electrode discrimination (Busby and Clark, 1996; Collins, Zwolan and Wakefield, 1997; Dawson, McKay, Busby, Grayden and Clark, 2000; Zwolan, Collins and Wakefield, 1997). The assumption used in this study is that the simulated CI listener uses a constant stimulation rate and that the stimulation takes place in a multi-electrode stimulation processing strategy such as ACE (Loizou, 2006). Therefore, the published data for pulse rate discrimination would not be valid for comparison, and because there is little variation in the active electrodes between adjacent frequencies tested in this study, electrode discrimination is not really a factor. Hence, comparisons have mainly been drawn with data obtained in sound field conditions. These results are shown in figure 4.5 on the next page, with the frequency difference limens normalised to the subject's implant filter width. In the case of the illustrated data, the filter widths were all 125 Hz. This metric is maintained here for consistency with published data. A comparison is provided with simulated and experimental NH data in figure 4.6 on page 109. There does not appear to be data available for CI frequency discrimination as a function of intensity or duration in sound field conditions, and therefore there is no data with which to compare simulated data for these experimental parameters.

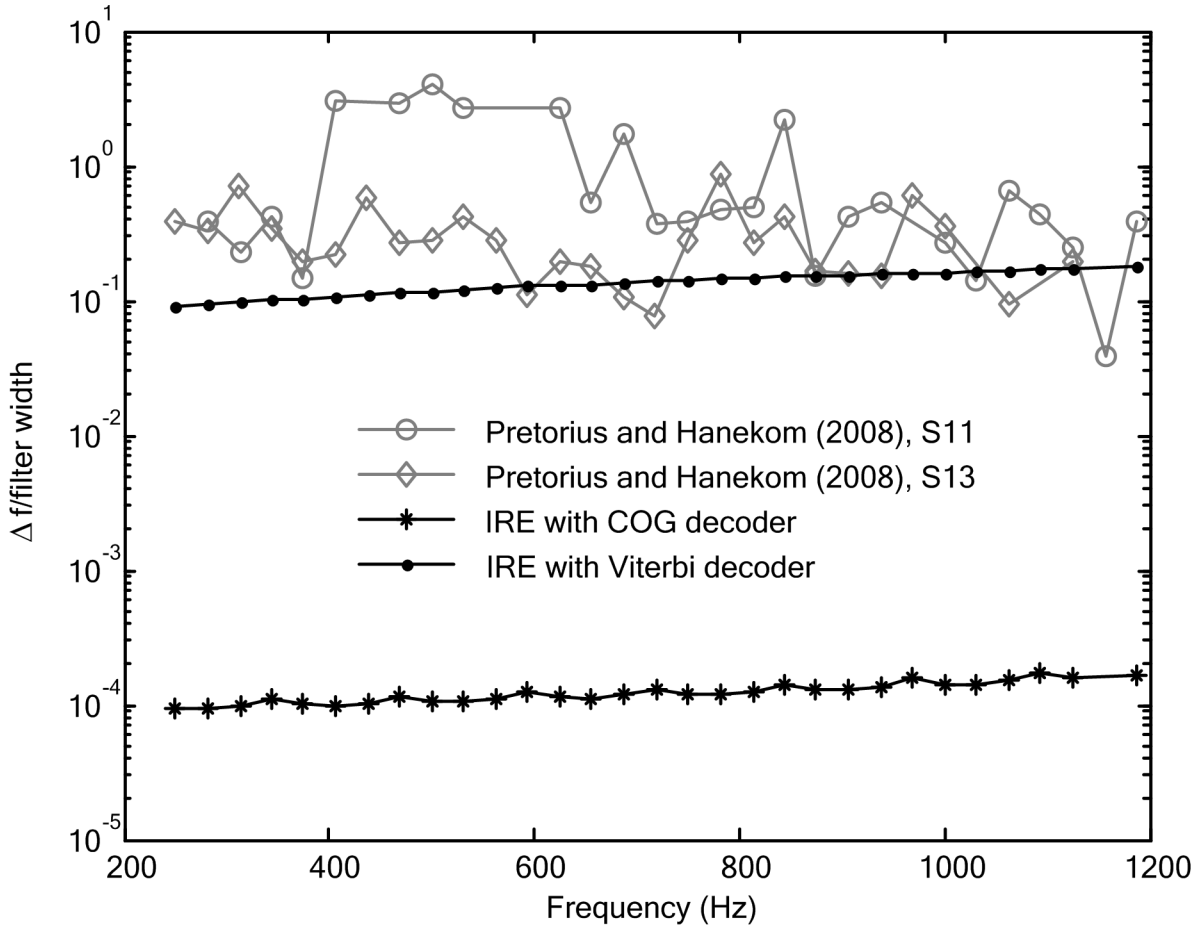

Figure 4.5.
Frequency discrimination data for CI listeners as a function of frequency. Data from human subjects is shown by the grey plots representing data from subjects S11 and S13 from Pretorius and Hanekom (2008) respectively (75% of maximum comfortable loudness, 500 ms). Data from the IRE model is indicated by the black lines (70 dB SPL, 500ms), showing the Viterbi decoder (dots) and COG decoder (stars) results respectively. All data is normalised to the filter width of the subject, which in all cases is 125 Hz.

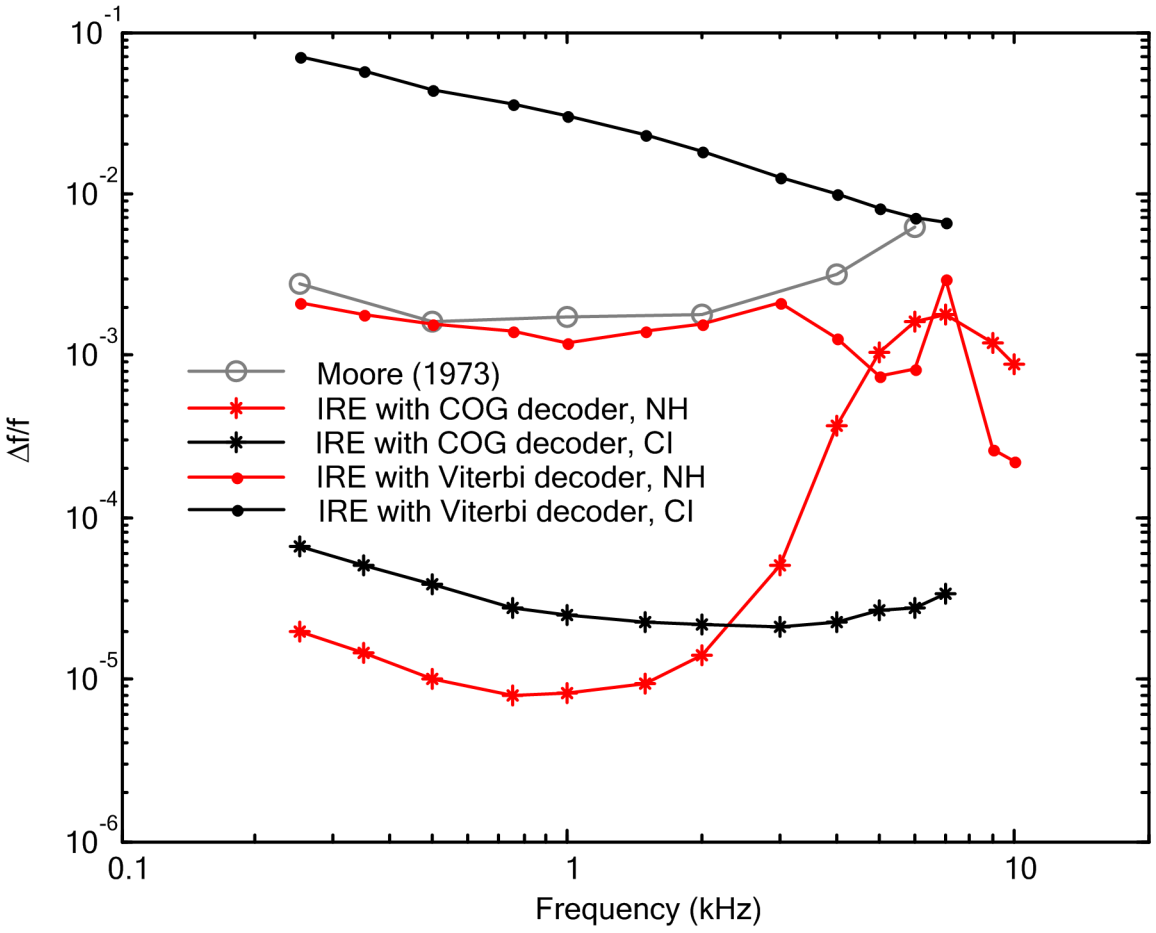

Figure 4.6.
Frequency discrimination data for CI listeners compared to NH listeners as a function of frequency. The grey plot shows experimental NH data from Moore (1973), as shown in figure 4.1. Red plots represent NH listener data, whereas the black plots represent CI hearing data. In both cases, the plots punctuated by stars indicate the COG decoder results and the plots punctuated by dots indicate the Viterbi decoder results. All data points are the mean of 50 simulations, except for the experimental data from Moore (1973).

The results of the simulated CI listener using a COG decoder in figure 4.5 show substantially lower magnitudes than those obtained from the measured experimental results from Pretorius and Hanekom (2008). Subtle oscillations over the frequency range can be observed, as well a very slight positive slope. The Viterbi decoder output shown in figure 4.5 is once again much higher than the COG decoder output and is in the same order of magnitude as the experimental data. The same slight upward trend in the discrimination threshold is observable, but the plot is flat over the entire range of frequencies.

When the COG decoder results from a simulated cochlear implantee are compared with a simulated NH listener (figure 4.6), it can be observed that the CI results are higher than NH results for lower frequencies, but because the CI results are relatively flat, there is a turning point (just above 5000 Hz) at which the experimental NH data from Moore (1973) is higher than simulated CI values. It appears as though this turning point occurs at approximately 6000 Hz in the case of the Viterbi decoder. There is a slight discrepancy in the trends of the two decoders above 3000 Hz. Both the COG and Viterbi decoders show a relatively constant decrease in the Weber fractions over the range of frequencies tested, although a wide bowl-shaped trend is noticeable for the COG decoder.

4.4 DISCUSSION

The results presented in this chapter are used to explore the value of the predictions made by the IRE model for frequency discrimination experiments for NH and CI listeners. The model-predicted outcomes have been compared to results from literature, and observations regarding the model's performance have been presented alongside the figures. In this section, the origins and implications of these results are discussed. Further discussion is provided in chapter 6.

4.4.1 Normal hearing frequency discrimination

The bowl-shaped curve in figure 4.1 is a typical characteristic of frequency discrimination curves (Moore, 1973; Moore, 2003; Sek and Moore, 1995), although the origin of the decrease in performance after the 2 to 3 kHz point is attributed to the roll-off to phase locking of spike trains to the acoustic stimulus (Johnson, 1980; Joris, Carney, Smith and

Yin, 1994a; Joris, Smith and Yin, 1994b). The IRE model does not take this phenomenon into account, yet still reproduces the bowl shape. This indicates that phase locking of stimuli may not be necessary to explain the existence of this shape, or at least it is not the only mechanism that can possibly explain it. In this model, the origin of the shape is attributable to the change in the tuning width of the tuning curves, as observed by Kiang (1965) and modelled by Colburn (1973), and is observable in both of the different decoders. This is supported by the fact that by increasing the value of f_T to 2000 Hz, the bowl shape becomes wider and the change in performance occurs at 2000 Hz.

The prediction of the bowl shape in the frequency discrimination curve is unexpected owing to the fact that the model uses a rate-place mechanism instead of a mechanism that takes synchronous spike timing information into account. Average rate-place models (such as those of Siebert, 1970; Henning, 1966) do not normally manage to obtain the bowl-shaped trend, but rather a continual downward trend, related to the increasing sharpness of the tuning curve, is obtained. In the IRE model, the increase in tuning sharpness (Colburn, 1973; Glasberg and Moore, 1990; Liberman, 1978) with increasing frequency is taken into account, but the performance drops. This is consistent with the study of COG decoders by Snippe (1996), who proved that performance worsens for such decoders as tuning narrows. This narrowing in the tuning is clearly shown in figure 4.2. Therefore, this trend may be specific to rate-place population codes that utilise the instantaneous rate of spike trains as opposed to conventional rate-place codes that utilise the average rate of spike trains.

The COG decoder's unrealistically low predicted magnitudes for frequency discrimination are expected and displays near equivalence to Heinz *et al.*'s values (Heinz *et al.*, 2001a) under certain conditions, given that the COG decoder performs optimally with broadening in tuning (Snippe, 1996). The broadest excitation pattern exists at f_T (refer to figure 4.2), where the best performance in frequency discrimination can be observed. The slight difference between Heinz *et al.*'s all-information model and this model is likely due to the lower amount of information utilised by the IRE model, as the former uses all the information available over the length of the cochlea, as well as timing information in the spike patterns. This result is consistent with the above-human performance of optimal decoders that have been presented before (Gresham and Collins, 1998; Gresham and Collins, 1999a; Heinz *et al.*, 2001a).

A clue to the unusual trend in performance predicted by the COG decoder is provided in figure 4.2, where it can be observed that the unusual pattern that occurs between 5000 Hz and 10000 Hz could be brought about by the very low firing rates included in the critical

band. This outlines a potential weakness of the model in that the results are very sensitive to the tuning that is used. This effect might have been mitigated by centring the critical band over the region with the highest firing rates on all fibres instead of using the maximum firing channel as the centre. It has been shown in animal data (Kiang, 1965; Liberman, 1978) and in human data (Glasberg and Moore, 1990) that the sharpness of neural tuning curves increases with frequency; it is possible, however, that the simplistic model of triangular tuning curves that is used in the IRE model overestimates this degree of tuning curve sharpening, which affects the results. Shera, Guinan Jr. and Oxenham (2002) present a model of human tuning sharpness that increases as a power function above 1000 Hz. The use of such a model may provide a better prediction of experimental results. Additionally, the IRE model utilises only low spontaneous rate fibres in the predictions which results in a restricted area of activation. The inclusion of medium and high spontaneous rate fibres would widen the area of activation over all frequencies (Shofner and Sachs, 1986), which could affect the discrimination results at higher frequencies. The change in the RI functions as frequency changes (see equation 3.5) is based on very little data, and so this could also affect the tuning. More realistic representations of the tuning curves (Glasberg and Moore, 1990) and excitation patterns may also alleviate the problem, but even the simplistic model of the acoustically stimulated periphery still provides a good representation of the trend observed up to 5000 Hz with the standard parameters, and provides data very close to the model of Heinz *et al.* (2001a) for f_T = 2000 Hz.

The Viterbi decoder is clearly even more sensitive to slight inadequacies in the model when predicting frequency discrimination data. This is probably due to the reliance of the Viterbi decoder on the internal representation of the excitation pattern (in the form of the λ-matrix), which is a function of many variables, such as the tuning curves used, the number of fibres in the ERB and the RI functions. This complex dependence might cause the acoustic stimulation assumptions to become readily apparent in the predictions as the simplifications are compounded. The dependence on these parameters, especially the tuning and ERBs, is illustrated by the dotted line in figure 4.1, showing a variation on the Viterbi algorithm's performance.

Despite the Viterbi algorithm's difficulty in predicting the frequency discrimination trend over the entire range of frequencies, the near-experimental performance of the Viterbi decoder implies that it is a better candidate for predicting frequency discrimination data than the COG decoder for lower frequencies. The difference between the two decoding mechanisms is substantial considering that in both cases exactly the same information is provided to

the decoder to make the frequency discrimination predictions, and all predictions are based on the variance in the estimated frequency over the course of time. The key distinction is that the COG decoder finds the centre of the region of maximum stimulation over the population, and using tonotopic mapping, assigns a frequency value to this. The Viterbi algorithm uses prior knowledge (the internal representation by the λ-matrix) of what it expects and makes its frequency estimate based on a probabilistic mechanism by comparison of the prior knowledge with the estimated input rate (see equation 3.27). The Viterbi algorithm is therefore a form of maximum-likelihood decoder (Lou, 1995). Furthermore, the CANS may use this type of decoder in making predictions. This implication is discussed further in chapter 6.

The IRE model is also rather sensitive to the threshold parameter. This is illustrated by the results in figure 4.3, where the threshold was reduced to 10 dB. The distinctive shape of the intensity dependence graph of frequency discrimination seems to be brought about near to the threshold value, which in this case is low. Hanekom and Krüger (2001), Heinz *et al.* (2001a) and Wakefield and Nelson (1985) attribute this trend to synchronisation information available to the processor. The IRE model shows how the trend can be accounted for without phase information. It has been shown by experimental studies on animals that there tend to be more lower threshold fibres than high threshold fibres with CFs below 10 kHz (Kiang, 1965; Sachs and Abbas, 1974), which explains the necessity of reducing the threshold using the IRE model. It is expected that if the model provided a more realistic spread of thresholds and spontaneous rates (e.g. as documented in Winter and Palmer, 1991), the discrimination results as a function of intensity should improve.

Another interesting trend for both decoders is that the frequency discrimination ability saturates. This is probably due to the saturating RI functions in the model. It is therefore conceivable that if the representative RI function is a sloping-saturating or straight function, the trend could possibly be set to be closer to experimental data. The usual spread of RI curve types is not included in the model (Yates *et al.*, 1990; Yates, 1990). A better representation of the intensity dependence of this particular percept might be yielded if this were to be included in the model, if the intensity dependence of frequency discrimination is indeed related by the RI functions. The trend of NH data is well predicted for intensity, which shows that the simple RI description provides adequate information in the spike patterns to explain the data. This is surprising when considering the other models, which utilise timing information to predict the same trends (Hanekom and Krüger, 2001; Heinz *et al.*, 2001a).

The COG decoder's predictions are not as close to the values of Heinz *et al.* (2001a) as in

figure 4.1, but this is a result of the fact that the model does not adequately account for the duration dependence of frequency discrimination (refer to figure 4.4). However, the COG decoder still performs far better than human capabilities and the Viterbi decoder, but displays the same trends over the same intensity range, once again illustrating its efficiency relative the Viterbi decoder.

The duration dependence of frequency discrimination is not well predicted by the model. A plateau effect is seen at longer durations for both decoders, which is similar to what occurs in experimental conditions, but at lower durations there is no correspondence in performance levels. In terms of magnitude, both decoders still compare well to their analogues, further highlighting the difference between the two decoding mechanisms mentioned above. The IRE model's failure to do so is a result of the absence of any temporal-based dynamics in the model, such as neural adaptation (Elliot and Fraser, 1970; Rieke *et al.*, 1997; Viemeister and Wakefield, 1991). The lack of synchronisation information may also contribute to this. Additionally, the absence of onset and offset effects would have a substantial effect on results at lower durations. This is because the onset of the sound results in a higher discharge rate than the sustained signal, which is very different from the description of the Poisson model (Heinz *et al.*, 2001a). It has generally been found that models that take timing information into account produce an accurate representation of the trend seen in experimental conditions (see Hanekom and Krüger, 2001; Heinz *et al.*, 2001a; Goldstein and Srulovicz, 1977). The IRE model is useful for the purposes of predicting frequency and intensity discrimination, but this should be done with longer tone durations (particularly, durations of 100 ms or higher) in order for the magnitudes to be representative of performance levels. This is a typical problem of rate-place models (Siebert, 1970; Heinz *et al.*, 2001a) and if additional timing or neural adaptation features were to be added to the model, it is conceivable that the dependence on duration trends would be improved.

4.4.2 Cochlear implant frequency discrimination

Despite the fact that the model has limited capabilities in describing all of the trends in the data for NH listeners, it is useful for predicting data at lower frequencies, longer durations and normal levels of intensity within a sound field environment. The IRE model is therefore valid for predicting results for CI listeners within the same parameter ranges.

The predicted results for CI listeners using the COG decoder, shown in figure 4.5, are to be

expected considering the NH outcomes in the same range of frequencies (figure 4.1). There is a very slight oscillation in the relatively straight line, which is due to the filter peaks and cross-over points at which the filters were stimulated. Pretorius and Hanekom (2008) stimulated the filters at different positions relative to the points of highest activation on each filter, and found a relationship between the position of stimulation on a filter and the frequency discrimination performance of the subject. This effect cannot be observed to be occurring in these results to any substantial degree, but the oscillations in the COG data suggest that such an effect may have an influence on CI frequency discrimination outcomes. The Viterbi decoder output is much higher than the COG decoder output and is in the same order of magnitude as the experimental data, as is the case with predictions for NH listeners using the IRE model. Comparing trends in the data is difficult owing to the high standard deviations in the data of Pretorius and Hanekom (2008) although it is clear that the flat line yielded by the Viterbi algorithm decoder provides an accurate indication of the order of magnitude at which CI listeners can discriminate frequencies for lower frequencies, and the small oscillations do not appear to be visible. Rogers *et al.* (2006) tested fundamental frequency discrimination using words in a sound-field environment and found the mean discrimination threshold for CIs to be approximately 25 Hz. The IRE model underestimates this marginally (by approximately 13 Hz) at the lower frequencies, but is not very different from the performance of the better-performing subjects in the study of Rogers *et al.* (2006). Slight differences could be attributed to the increased level of difficulty in the detecting frequency changes in words as opposed to pure tones. The support provided by the two separate studies mentioned above validates the application of the model to predictions in the sound field.

The very slight rise in the threshold values over the lower frequency range (figure 4.5) can be attributed to the gradually changing representation of CFs in the ERB used while the area of activation by the electrodes remains relatively constant (as mentioned before, only changing with the point of activation on the filter). This is in contrast with what is observed in the NH listener results, where the λ-matrix changes similarly to the activation (owing to the fact that the activation is dependent on the λ-matrix) and therefore the frequency discrimination trend is sensitive to the changing shape of the λ-matrix over frequency.

The difference in trends between the COG decoder and the Viterbi decoder in figure 4.6 corresponds approximately to where the CI filters change from a linear distribution to a logarithmic one (above 1000 Hz). It is likely that the COG decoder provides a better representation of the trend in frequency discrimination at the higher frequencies, judging by the model's performance for NH, but no concrete conclusions can be drawn owing to a lack of

CI data with which to compare. However, Chen and Zeng (2004) also found an increasing trend over the same range of frequencies with implantees for frequency modulation detection using single-electrode stimulation. The trends they observed increased far more rapidly than those is observed in the IRE model predictions. Dorman *et al.* (1996) carried out frequency discrimination experiments using triads of frequencies and percentage correct scores of selecting the highest or lowest frequency in the triad. They also showed a monotonic decrease in the performance over frequency. Results such as these would suggest that the IRE model-predicted trend is correct.

The fact that the IRE model is able to predict frequency discrimination results satisfactorily for both hearing modalities suggests that the mechanism by which this is performed is a valid representation of the processing of information in the CANS. This implies that a central rate estimation mechanism utilising an internal model of the expected sounds attempts to create an internal "image" from which it decides what the input frequency is that it perceives. This process would need to be performed at regular intervals (millisecond intervals) instead of over long periods (hundreds of milliseconds intervals) of time, and the uncertainty in this prediction imposes a limit on the frequency discrimination performance levels.

4.5 SUMMARY

This chapter described the parameters and results used to predict frequency discrimination results using the IRE model. The predicted frequency discrimination thresholds for NH and CI listeners were predicted well over a limited range of frequencies and intensities using the IRE model. However, the duration dependence of frequency discrimination was not well accounted for. The limitations are probably due to the simplifying assumptions made in the model. The model is useful for predicting results for CI listeners, and these results were compared with data from Pretorius and Hanekom (2008) and Rogers *et al.* (2006). The results thus validate the functioning of the model for frequency discrimination. This suggests that the model used to describe the CANS is a valid representation of the processing that occurs in the auditory cortex. The IRE can be considered to be a valuable contribution to psychoacoustic models as it can predict the trends and magnitudes of frequency discrimination within a limited range for two hearing modalities. The next chapter investigates the application of the model to intensity discrimination tasks in order to see whether the model is able to make similar predictions for another auditory percept. Satisfactory results from such

predictions aid in confirming the hypothesis that the IRE model is a valid representation of the central auditory processes.

CHAPTER 5

INTENSITY DISCRIMINATION

5.1 CHAPTER OBJECTIVES

In this chapter, the model developed in chapter 3 is used to predict the outcomes of intensity discrimination experiments for NH and CI listeners. The aim of this chapter is to measure the usefulness of the model in predicting intensity discrimination results and thereby to support some of the findings of chapter 4. Most of the parameters used in this chapter are the same as those used in the previous chapter and are therefore not repeated, but a short introduction to the subtle differences commences the chapter. This is followed by the results of intensity discrimination for NH listeners, presented in the same way as the NH results for frequency discrimination. The predicted outcomes for the intensity discrimination of CI listeners is then provided. Finally, a discussion of the implications of these findings is presented.

5.2 METHODS

As with the frequency discrimination simulations, most of the methods used for intensity discrimination predictions are described in chapter 3, with the main difference between the two percepts being the λ-matrix used by the central estimator (see section 3.3.4.1). The parameters used for intensity discrimination simulations in this study were the same as those listed in tables 4.1, 4.2 and 4.3, and for this reason, parameter tables are not listed again. For

details, refer back to the above-mentioned chapters and tables.

In the generation of the λ-matrix for intensity discrimination, instead of a set of state frequencies, a set of state intensities was required. Given that the threshold was set to 49.1 dB as the standard, the set of state intensities was chosen to encompass the entire range of resulting rates (from spontaneous rate to maximum firing rate). Therefore, the standard intensities ranged between 20 dB SPL and 90 dB SPL divided over 100 states. Where this standard was not used, this is noted. Justifications for the parameters used are provided in section 4.2.4 (in the previous chapter).

Parameter consistency is maintained to test the model's ability in predicting a different percept without drastically changing the model. This could be considered analogous to switching the focus of the task in a psychoacoustic test: the physiological mechanisms remain the same, but the dimension analysed differs.

5.3 RESULTS

The results for intensity discrimination are presented here for both NH and CI listeners. The methods described in chapter 3, section 4.2 and section 5.2 detail how these results were obtained. Similarly to section 4.2, the NH results presented here are intended to provide validation of the model, followed by the CI results, presented for investigative purposes. In all cases, only the Viterbi decoder is used as the COG decoder cannot be applied in the absence of a tuning parameter (Snippe, 1996). Only objective observations of the results are provided in this section. Reasoning for the results obtained and insights into the model's performance are provided later, in the discussion section of this chapter (section 5.4). Further discussion is provided in chapter 6.

5.3.1 Normal hearing intensity discrimination

There is extensive data available in literature for intensity discrimination in NH listeners, as discussed in chapter 2. The model data is presented here along with some classic results and some more recent modelled results. As in the frequency discrimination section, intensity discrimination is presented as a function of frequency, intensity, and duration. These

three dimensions were used in Heinz *et al.* (2001a) to provide an indicator of their model's performance in predicting intensity discrimination data.

5.3.1.1 Intensity discrimination as a function of frequency

The results for modelled intensity discrimination data as a function of frequency are provided in figure 5.1. The data is compared with the experimental human data of Florentine *et al.* (1987) and the all-information model of Heinz *et al.* (2001a).

From figure 5.1, it can be observed that the IRE model using the Viterbi decoder predicts intensity discrimination levels of the same order of magnitude as those found by experiments on human subjects, but the general slightly upward trend observed in human data is not as clear in the IRE model data, which is mostly flat. This is in contrast with the data generated by Heinz *et al.* (2001a), where the trend is clearer and more similar to that of human data. However, their all-information model predicts far lower threshold values than those actually obtainable in humans.

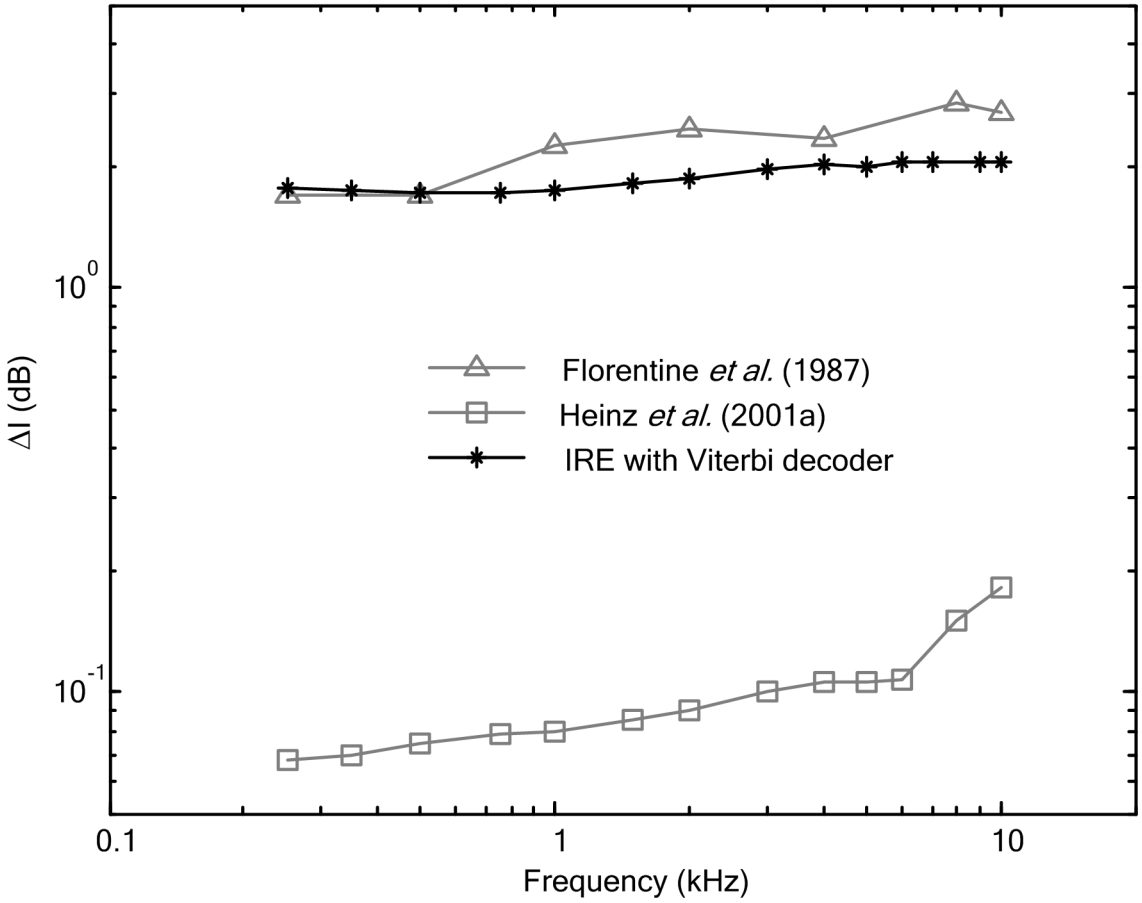

Figure 5.1.
Intensity discrimination data for NH listeners as a function of frequency. Data from literature is shown by the grey plots representing human experimental data from Florentine *et al.* (1987) (triangles, 40 dB SPL, 500 ms) and Heinz *et al.*'s all-information model (squares, 40 dB SPL, 200 ms, Heinz *et al.*, 2001a). Data from the IRE model with the Viterbi decoder is indicated by the black line punctuated by stars (40 dB SPL, 500ms). The data for this study is the mean of 50 simulations.

5.3.1.2 Intensity discrimination as a function of intensity

One of the fundamental psychoacoustic measures is intensity discrimination measured as a function of intensity, as described in chapter 2. This percept was simulated using the IRE model for NH listeners and compared with data found in the literature; specifically, human experimental data by Florentine *et al.* (1987) and modelled data by Heinz *et al.* (2001a),

Jepsen *et al.* (2008) and Dau, Kollmeier and Kohlrausch (1997a). This comparison can be seen in figure 5.2.

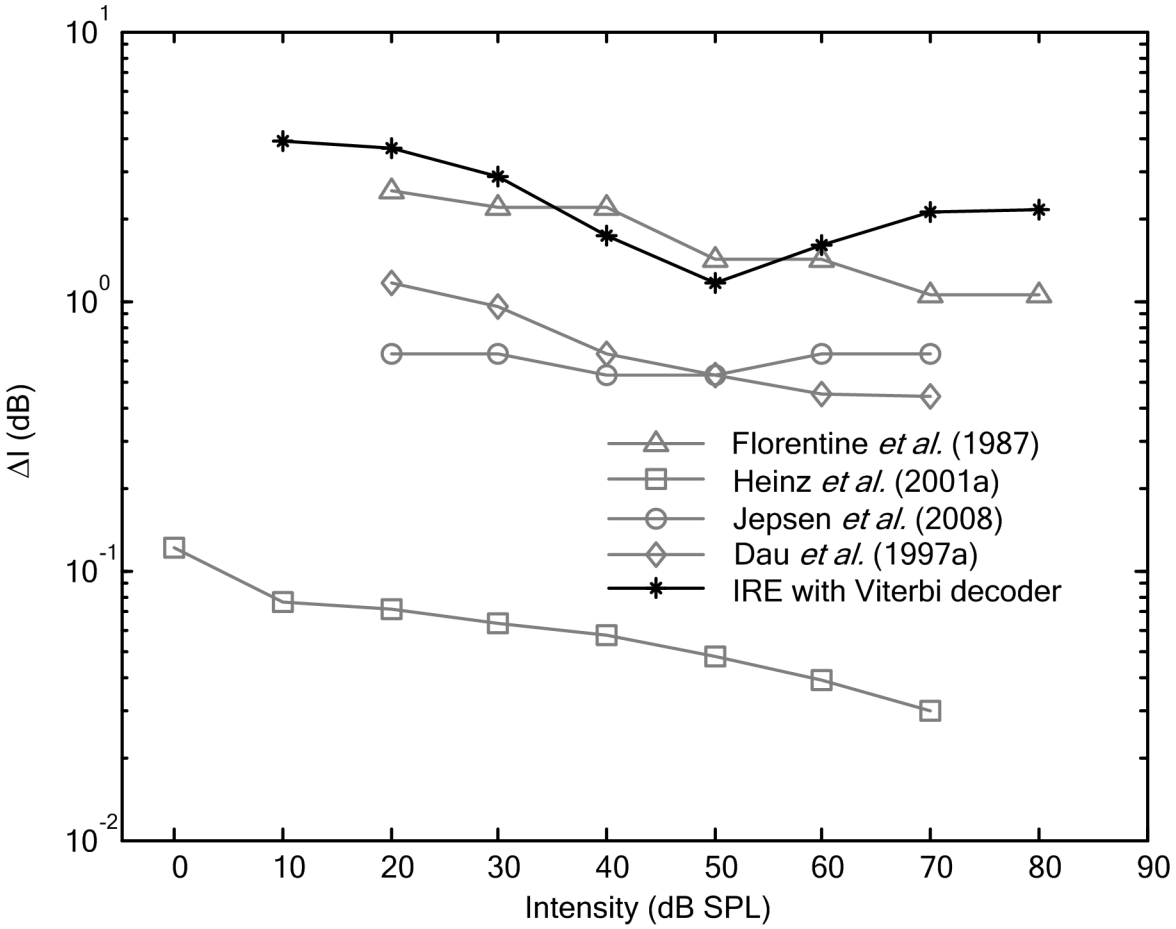

Figure 5.2.
Intensity discrimination data for NH listeners as a function of intensity. Data from literature is shown by the grey plots representing human experimental data from Florentine *et al.* (1987) (triangles, 1000 Hz, 500 ms), Heinz *et al.*'s all-information model (squares, 970 Hz, 500 ms, Heinz *et al.*, 2001a), Jepsen *et al.*'s model (circles, 1000 Hz, 800 ms, Jepsen *et al.*, 2008) and Dau *et al.*'s model (diamonds, 1000 Hz, 800 ms presented in Jepsen *et al.*, 2008). Data from the IRE model with the Viterbi decoder is indicated by the black line punctuated by stars (1000 Hz, 500ms). The data for this study is the mean of 50 simulations.

The model-predicted data for intensity discrimination using the IRE model with the Viterbi decoder shows once again that the predicted magnitudes are of the same order of magnitude as those of data taken from Florentine *et al.* (1987), but follow a slightly different trend.

At some point after the threshold value (49.1 dB SPL), the trend changes direction and the values of intensity discrimination begin to increase instead of continuing to decrease as can also be observed in the modelled data by Heinz *et al.* (2001a) and Dau *et al.* (1997a). When compared to Jepsen *et al.*'s (2008) data, the IRE model data is similar in trend but slightly different in magnitude.

Again, the modelled data by Heinz *et al.* (2001a) is two orders of magnitude lower than that of experimental data. The models by Dau *et al.* (1997a) and Jepsen *et al.* (2008) provide data that is close to that of experimental data, but Jepsen *et al.*'s does not yield the correct trend.

5.3.1.3 Intensity discrimination as a function of duration

The results for modelled intensity discrimination data as a function of duration are provided in figure 5.3. The data is compared with the experimental human data from Florentine (1986) and the all-information model of Heinz *et al.* (2001a).

The IRE model data shown in figure 5.3 is generally in the same order of magnitude as the measured human data from Florentine (1986), but follows a completely different trend. The trend is similar to that observed in the frequency discrimination as a function of duration for NH listeners for the IRE model. Where the trend appears to settle in the IRE data, the magnitude is higher than Florentine's data by a factor of approximately two (Florentine, 1986), but is also in the same order of magnitude as the experimental data. When a comparison is made between the data obtained by Florentine *et al.* (1987) and Florentine (1986) (as seen in figures 5.1 and 5.3 respectively) the magnitudes differ slightly, indicating that the difference observed between the IRE data and Florentine's data in figure 5.3 is not substantial. However, there is a substantial difference between the IRE model's data and the modelled data of Heinz *et al.* (2001a). As with frequency discrimination data, both the trend and the order of magnitude differ. However, Heinz *et al.* (2001a) managed to predict the correct trend over a range of durations, as they did with frequency discrimination data.

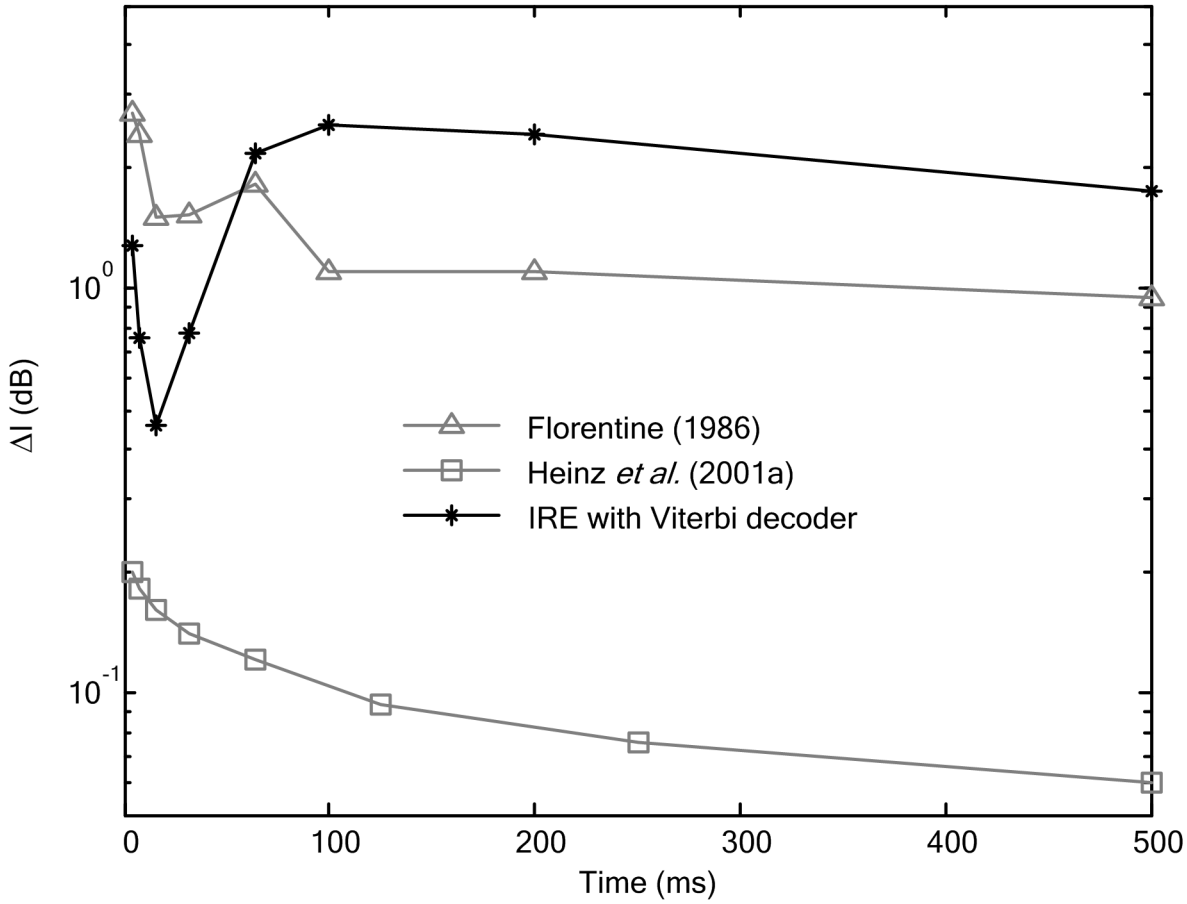

Figure 5.3.
Intensity discrimination data for NH listeners as a function of duration. Data from literature is shown by the grey plots representing human experimental data from Florentine (1986) (triangles, 40 dB SPL, 1000 Hz) and Heinz *et al.*'s all-information model (squares, 40 dB SPL, 970 Hz Heinz *et al.*, 2001a). Data from the IRE model with the Viterbi decoder is indicated by the black line punctuated by stars (40 dB SPL, 1000 Hz). The data for this study is the mean of 50 simulations.

5.3.2 Cochlear implant intensity discrimination

There is much data in literature that tests intensity discrimination at the single-electrode level, but very few studies have performed such tests in a sound field environment (Rogers *et al.*, 2006), making comparisons to existing data difficult. The IRE model-predicted data for CI listeners is illustrated in figure 5.4 and compared to IRE modelled NH data: experimental

data from Florentine *et al.* (1987) for NH listeners and experimental data from Rogers *et al.* (2006) for NH and CI listeners.

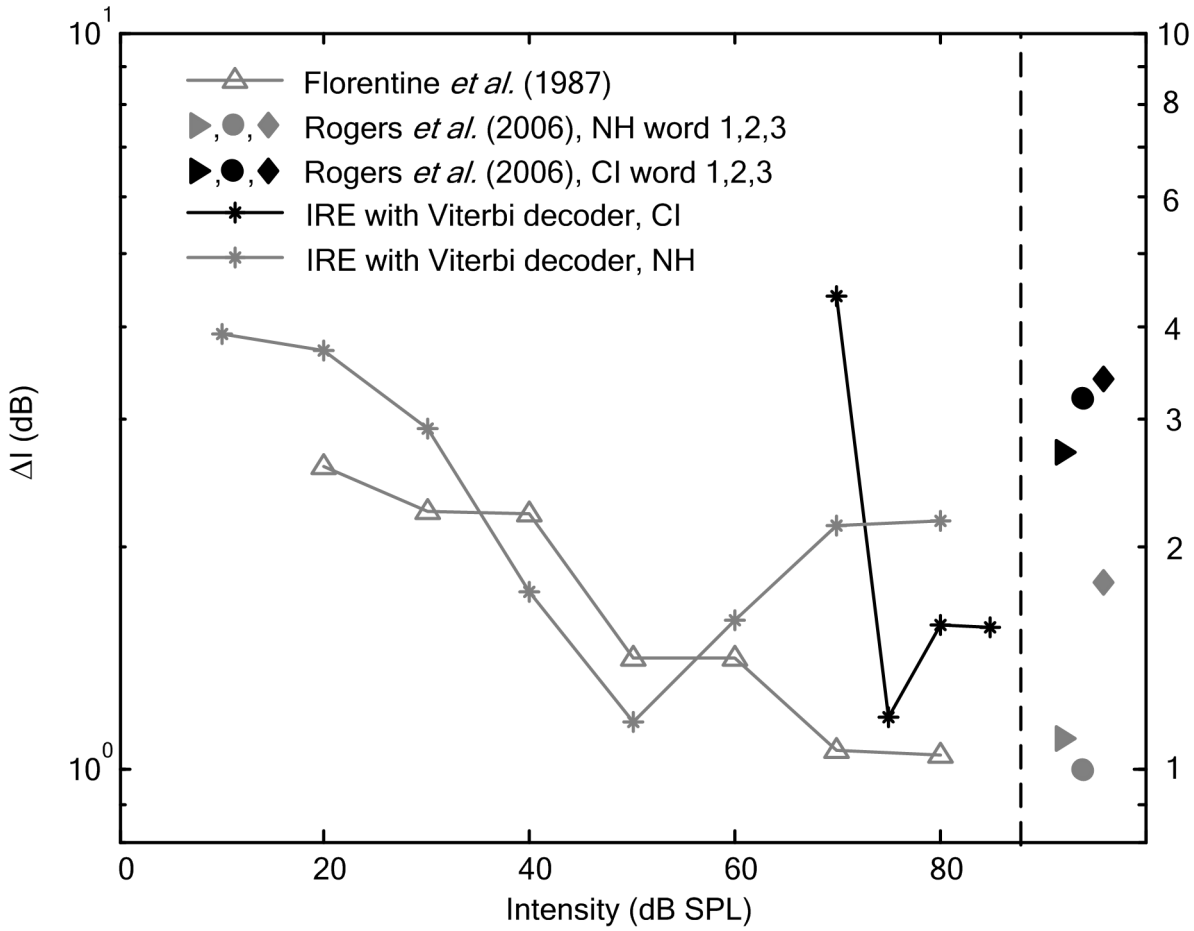

Figure 5.4.
Intensity discrimination data for NH and CI listeners as a function of intensity. NH data is shown in grey, while CI data is shown in black. The filled shapes represent the mean data from Rogers *et al.* (2006) for intensity difference limens obtained in a sound field environment (as part of a word at 70 dB SPL). The plots show NH experimental data from Florentine *et al.* (1987) (triangles, 1000 Hz, 500 ms) and IRE model data with the Viterbi algorithm for NH listeners (grey stars, 1000 Hz, 500 ms) and CI listeners (black stars, 1000 Hz, 500 ms). The data for this study is the mean of 50 simulations.

The intensity discrimination data in figure 5.4 for cochlear implantees from the IRE model is only represented over a small intensity range, much smaller than that shown for modelled and experimental data for NH listeners in the same figure. A decrease in threshold with

increasing intensity is initially apparent, but this is followed by a sharp increase and a slight plateau in values. This trend is similar to that observed in the IRE model-predicted data for NH listeners. Despite the smaller range, the two hearing modalities are within the same range of magnitudes of difference limens.

The experimental data from Rogers *et al.* (2006) shows a difference in the magnitudes of intensity discrimination, the values for CI listeners being slightly higher than for NH listeners. The magnitude difference between the predicted intensity discrimination values for NH and CI listeners does not appear to compare well with the relation between NH and CI listeners in Rogers *et al.*'s data, but both the model-predicted values and data from Rogers *et al.* (2006) are within the same order of magnitude for both NH and CI listeners.

5.4 DISCUSSION

Section 5.3 presented predictions from the IRE model for NH and CI listeners. Superficial observations on the predictions were made in comparison to results from experimental studies found in literature. The origins and implications of these results are discussed here.

5.4.1 Normal hearing intensity discrimination

The flat trend of the IRE model is to be expected, as other rate-place models yield the same trend (Heinz *et al.*, 2001a; Siebert, 1968); however, the magnitudes are of the same order as those found in human data, which is contrary to other rate-place data (an analysis showing the lower magnitudes obtained by rate-place codes was performed by Heinz *et al.*, 2001a). This is similar to the predictions obtained for frequency discrimination, except that the trend is consistent over the entire range of frequencies. Fibres in the model are not tuned to intensity, which supports the explanation that the poor performance of the model at high frequencies for frequency discrimination is related to the tuning. The superior performance of the IRE model relative to the all-information model of Heinz *et al.* (2001a), as well as the rate-place model by Siebert (1968) (which predicts an intensity difference limen of approximately 0.5 dB over the entire range of frequencies) is unexpected. This supports the idea that the central processor in the auditory system is well represented by the IRE in lieu of the Viterbi decoder. The trend matches that determined by Jesteadt *et al.* (1977), which was later found

by Florentine *et al.* (1987) to be an overly simple description. The discrepancy in trend is likely a result of the lack of timing information in the model. Such information is included in the model of Heinz *et al.* (2001a), which predicts a trend similar to that of the work by Florentine *et al.* (1987) up to 8000 Hz. The IRE model would not be able to predict all of the trends discussed in Florentine *et al.* (1987) considering that the low-intensity performance is uniform across frequencies. This is a result of the nonlinearities that relate to intensity which are not included in the model. This is discussed further below, in the discussion of the intensity dependence of intensity discrimination.

The model predictions of intensity discrimination trends are of the same order of magnitude as experimentally obtained results, but the trends are rather different. The model does not predict Weber's law or the near miss to Weber's law[1] for the entire range of intensities. The deviation from Weber's law begins just above the threshold and appears to be related to the threshold. This behaviour is expected as the model does not include a wide distribution of threshold levels as would be observed in actual neural populations (Kiang, 1965; Sachs and Abbas, 1974). Heinz *et al.* (2001a) include a spread of thresholds in their model and attribute the steady decrease in intensity discrimination levels partly to this feature. This conclusion is consistent with the demonstration by Winter and Palmer (1991) showing how different RI functions and thresholds can affect intensity discrimination data. The decrease in performance could also be related to the amount of information provided to the central processor. Many studies attribute the improvement in intensity discrimination ability to the increase in the number of neural channels recruited (with the increase in excitation pattern size) with increasing intensity, thereby providing more information to the central processor (Buus, 1990; Florentine and Buus, 1981; Heinz *et al.*, 2001a; Jepsen *et al.*, 2008; Siebert, 1968). The trend observed in the IRE model-predicted results is consistent with findings by Heinz *et al.* (2001c) for low spontaneous rate fibres for three different mechanisms (rate-place, all-information and coincidence detection) when using narrow bands of excitation, as well as with modelled findings by Winter and Palmer (1991). The IRE model assumes a constant ERB over the entire range of intensities and utilises only low spontaneous rate fibres; therefore, the inability of the IRE model to predict the experimentally observed trend is not surprising at higher intensities and is supported by the findings of Heinz *et al.* (2001c). This assumption of the CANS's utilisation of a constant ERB over all intensities was made for the sake of consistency between the model of frequency discrimination and the model of intensity discrimination, to minimise the changes made between them.

[1] See appendix A

In addition to the exclusion of multi-band influences on intensity discrimination, nonlinear growth of the excitation pattern (Nelson and Schroder, 1997) and compressive gain on the basilar membrane (Moore and Oxenham, 1998) were not taken into account in the model. The former contributes to the near miss to Weber's law (Moore, 2003) and the latter could contribute to the dynamic range of the model (Heinz, 2000). The absence of these effects from the model could influence the shape of the intensity discrimination graph. These limitations of the model might affect level-dependent effects on intensity discrimination as a function of frequency as investigated by Florentine *et al.* (1987). Heinz *et al.* (2001a) did not include the nonlinear effects listed above, but did include multi-band effects, and their model was able to predict Weber's law and the near miss to Weber's law. This, coupled with the IRE model's inability to reproduce experimental trends at high intensities, lends weight to the argument that multi-band effects are important in intensity discrimination tasks at high levels, as well as the importance of the contribution of high spontaneous rate fibres. The results from the models of Jepsen *et al.* (2008) and Dau *et al.* (1997a) indicate that inclusion of the compression at mid-levels might have a negative effect on the predicted intensity discrimination trends and that incorporating an across-frequency coincidence detection mechanism might assist in correcting such an effect (Heinz *et al.*, 2001c). However, a recent study by Jesteadt, Schairer, Nizami, Khaddam and Neely (2009) found that a quadratic-compression model is able to predict intensity discrimination trends in the presence of noise better than an excitation pattern model.

The duration dependence of intensity discrimination is not well predicted by the model. The results are very similar to those of frequency discrimination observed in section 4.3.1.3. These results are to be expected owing to the exclusion of any adaptive effects from the model (Elliot and Fraser, 1970), or onset and offset effects. As with the frequency discrimination results, these results are consistent with average rate-place models that are only useful for extended durations (> 100 ms according to Heinz *et al.*, 2001a; Siebert, 1970). This is not debilitating for the model as it validates the other results which are all performed at durations of 200 ms or longer. The factor of two difference between the settled IRE model and the experimental data from the study by Florentine (1986) can be attributed to experimental variation, such as variation between subject sets. When a comparison is made between experimental data obtained by Florentine *et al.* (1987) and Florentine (1986) (as seen in figures 5.1 and 5.3 respectively) the magnitudes differ slightly, indicating that the difference observed between the IRE data and Florentine's data in figure 5.3 is not substantial. The magnitudes of the IRE model predictions are still within the same magnitude order as experimental results, validating the model's usefulness at longer durations. This is in contrast

with Heinz *et al.* (2001a), who predict the trend better, but underestimated the intensity discrimination thresholds by an order of magnitude. The discrepancy can again be accounted for by the additional information used by their all-information model, i.e. all the timing and rate-place information along with adaptation effects and onset and offset effects.

All of the NH results for intensity discrimination do not provide very good trend indications, but these failings can be attributed to simplifying assumptions used in the model, largely on the side of the peripheral hearing. The magnitudes are well predicted for all parameters tested. This emphasises the usefulness of the IRE used in conjunction with a Viterbi decoder (or more generally, a maximum likelihood decoder) to predict intensity discrimination results.

5.4.2 Cochlear implant intensity discrimination

As with the frequency discrimination data, the predicted NH trends do not match those observed in literature for the entire range of intensities; however, this data can still be utilised as a basis for comparison, as the discrepancies with experimental data can largely be accounted for. Additionally, the comparisons made here are not only meant to judge the usefulness of the model, but also to investigate the functional mechanisms of the CANS. This can be done irrespective of the accuracy of the predictions.

The CI predicted data is not consistent with the monotonic downward trend as a function of increasing intensity found in studies that stimulate single electrodes (Cohen, Saunders and Clark, 2001; Kreft, Donaldson and Nelson, 2004; Nelson *et al.*, 1996; Pfingst, Burnett and Sutton, 1983) for the entire range of intensities tested. Given that the predicted NH trend as a function of intensity is not entirely similar to experimental data, it follows that the CI predicted thresholds would likely not be similar in trend either, so this behaviour is to be expected. If the IRE model were to include multi-band effects (as included in the models of Florentine and Buus, 1981; Heinz *et al.*, 2001a), then it is conceivable that the monotonic trend might become apparent. This appears to be consistent with the findings of Drennan and Pfingst (2005) that intensity discrimination in electrical hearing is indeed determined by growth in the excitation pattern along the cochlea.

The similarity in the predicted magnitudes of CI and NH listeners appears to contradict evidence that CI listeners are far more sensitive to intensity changes than NH listeners (Nelson

et al., 1996; Shannon, 1983a) for some subjects, but this is in comparison to single-electrode stimulation, as opposed to studies performed in sound field environments. To obtain a more representative comparison, sound field intensity discrimination results should be used as a benchmark. The slight inconsistency that the model shows with respect to the data of Rogers *et al.* (2006) is not important as the magnitudes of the modelled CI data compare well to the better-performing listeners in their study. Rogers *et al.* (2006) show that in natural conditions, intensity discrimination performance can decrease by a factor of up to 2.4. This is consistent with the discrimination results predicted by the IRE model. This is further supported by Drennan and Pfingst (2006), who showed by use of a model that a marked decrease in discrimination performance occurs as more electrodes are employed in a stimulation strategy. They attribute this to summative channel interaction effects, whereby the loudness elicited by individual electrodes would combine, making level discrimination more difficult. This is likely to be the case with the IRE model as well as in natural or sound field listening conditions.

It is immediately clear from figure 5.4 that the predicted intensity discrimination results for the CI listener yield a much smaller dynamic range than for the NH listener. The same trend can be observed over a much smaller range of intensities and below approximately 65 dB SPL the stimulation levels are insufficient to elicit a response. This is consistent with the numerous findings that show that CI listeners possess a much smaller dynamic range of approximately 10 to 20 dB (Galvin and Fu, 2009; Javel and Viemeister, 2000; Nelson *et al.*, 1996; Pfingst *et al.*, 1983; Pfingst, 1984; Zeng, Galvin and Zhang, 1998; Zeng and Shannon, 1999), than that of NH listeners, which is approximately 120 dB (Bacon, 2004). In fact, the IRE model-predicted data shows a dynamic range of between 10 and 20 dB, which would indicate that the CI model provides a good representation of the information conveyed to the CANS. It shows that the dynamic range limit applies in sound field environments as well as to direct electrode stimulation. The smallest intensity increment can be derived from the value of the dynamic range and the intensity discrimination value. At best, predictions of between eight and 20 steps should be possible with the IRE model. This is close to the mean number of steps found by Galvin and Fu (2009) at 500 pps stimulation rate (about 22 steps), and is similar to the number of steps found in some of the worse-performing subjects (about 10 steps). The small discrepancy could be explained by the single-electrode stimulation used by Galvin and Fu making the detection of small increments in intensity easier to detect.

The only important change to the model in predicting intensity discrimination as opposed to frequency discrimination is a change in the internal model. This assumption is justifiable,

as the internal model represents the internal processing, and the processing pathways of frequency discrimination and intensity discrimination are distinct. The hypothesis mentioned in the previous chapter regarding an internal representation being important in discrimination tasks in the auditory system is further reinforced by the use of such a mechanism in predicting the intensity discrimination outcomes for both electrical and acoustic hearing over a limited parameter range.

5.5 SUMMARY

This chapter provided the results generated by the IRE model for intensity discrimination. These model-predicted results were compared with those found in literature and comparisons between the experimental and modelled data were made. Similarities and discrepancies have been explained and the implications of this have been discussed. It appears that most of the deviations in the data can be explained by simplifying assumptions made in the model and not necessarily made in the central auditory model. This implies that the model provides a valuable contribution to predicting intensity discrimination outcomes for NH and CI listeners. Considering the predicted outcomes for frequency discrimination from chapter 4, it appears that the model can be applied to at least two psychoacoustic percepts for a limited range of parameters. The next chapter provides a discussion on the implications of the results, considering the limited success of the model for the two different percepts by changing very little in the model's function and maintaining the model structure. The general discussion will consider the lessons learnt from both chapters 4 and 5.

CHAPTER 6

GENERAL DISCUSSION

6.1 CHAPTER OBJECTIVES

The objective of this chapter is to provide a discussion of the implications of the findings from chapters 4 and 5. This discussion expands on those in the aforementioned chapters by considering all of the results, instead of just those relevant to the particular chapter. Specifically, this chapter explores the implications of the success of the model and the nature of the central estimation mechanism and the decoders.

6.2 DISCUSSION

It is important to consider the implications of the results from chapters 4 and 5. The model is successful in predicting perceptual thresholds for a limited range of frequencies and intensities. Moreover, the discrepancies can largely be explained by the many simplifying assumptions made in the model. Although simplified, the parameters that were used are biologically plausible and are well supported by the literature (see chapters 2 and 3). The points in the model where simplifications are used are mainly in the descriptions of the peripheral processes, especially the model of acoustic stimulation. An important consideration in the performance of the model is that it is idealised as discussed in the chapters 3, 4 and 5. The most important idealisations include: the constant ERB length over the entire cochlea

(section 3.3.3.1), the triangular tuning curve representation (section 3.3.2.2), the simplified representation of the RI functions (section 3.3.2), the lack of phase-locking information in the model (section 3.3.3), the simple model of electrical stimulation (section 3.3.7) and the black box nature of the model of the CANS (sections 3.3.4 and 3.3.5).

Many factors that may affect the discrimination performance are not taken into account and this could cause inaccuracies in the predictions of the model. In the case of the NH model, no synchronisation to the stimulating frequency is included and the neural responses are treated as Poisson processes with constant rates. The assumption that the spike trains have constant Poisson rates over time introduces inaccuracies in the neural firing rates at short durations as the onset rates and adaptation effects usually result in higher initial firing rates than sustained rates (Johnson, 1980). Additionally, a constant Poisson rate does not take the synchronisation to periodic stimuli observed at lower frequencies into account (Johnson, 1980; Joris *et al.*, 1994a; Joris *et al.*, 1994b). Thus, some temporal characteristics of neural spike trains are not included in the model (including: adaptation, phase locking and onset and offset effects), although information coded in these characteristics is considered to be important in the encoding of frequency (Goldstein and Srulovicz, 1977; Heinz *et al.*, 2001a; Heinz *et al.*, 2001b; Johnson, 1980). Nevertheless, the constant Poisson rate facilitates the use of the rate estimator in the model of the CANS (see section 3.3.4 and Rudemo, 1972). The effects of excluding the temporal characteristics of neural spike trains can be observed in the comparisons made by Heinz *et al.* (2001a) between an all-information model and an average rate-place model, where the trend in the frequency discrimination results decrease as a function of frequency instead of revealing a bowl-shaped curve.

Simple RI functions that are approximately constant in both the NH and CI models are assumed. However, substantial variation in these functions from neuron to neuron has been documented in the case of both acoustic stimulation (Kiang, 1965; Köppl and Yates, 1999; Sachs and Abbas, 1974; Yates, 1990) and electrical stimulation (Bruce, 1997; Javel *et al.*, 1987). Considering the dependence that the model has on the RI functions, including more detailed representations of the functions could affect both intensity and frequency discrimination predictions. The extent to which this would affect the model is cumbersome to quantify, but the inclusion of a realistic distribution of spontaneous rates (Jackson and Carney, 2005), maximum rates and thresholds (Kiang, 1965; Sachs and Abbas, 1974; Winter and Palmer, 1991) would result in a greater number of fibres firing in response to a specific stimulus. This is because fibres with high spontaneous rates fire continuously, even in the absence of stimuli, and fibres with lower thresholds would need lower stimulus intensities to

elicit neural firing. This would cause the excitation pattern to be spread over a larger area along the cochlea, and could increase the magnitude of error in the predicted results. This would need to be accounted for in the central estimation mechanism, which would need prior knowledge of the RI functions of each nerve fibre and to account for them in some way. It is plausible that some form of compensation for the wide variety of characteristics of neurons may be present in the CANS; e.g. it has been demonstrated by Sutter (2000) that central, rather than peripheral filtering, may have a substantial effect on some psychoacoustic phenomena. The inclusion of such mechanisms in the model are expected to severely affect the COG decoder as it relies on the representation of magnitude by firing rate to a large extent.

An important consideration is the fact that it is not possible to map acoustic intensity threshold to an equivalent current intensity, i.e. the RI functions for the two hearing modalities use two different and unrelateable units. However, this is unlikely to have greatly affected the data, as in both hearing modalities the stimulation levels were high enough to elicit suprathreshold stimulation of the neural channels (Javel, 1990).

The same neural properties describe the two different hearing modalities, which assumes that other than total hair cell loss, there is no other impairment to the auditory system. Cochlear pathology varies from individual to individual, depending on the severity of damage to the peripheral auditory structures (Nelson and Hinojosa, 2004; Schuknecht, 1993). This can result in psychophysical aberrations, such as the widening of critical bands (Nienhuys and Clark, 1979). Such abnormalities are not reflected by this model; therefore the model may overestimate the performance of implantees. Simplifying assumptions are made with regard to the placement of the electrode array and the anatomy of the cochlea. The array is assumed to be equidistant from all nerve channels. Cohen *et al.* (2001) show how the non-uniform placement of arrays affects a subject's psychophysical performance and Hatsushika, Shepherd, Tong, Clark and Funasaka (1990) show the extent of the variation in the scala tympani. If these considerations were included in the model they could have an effect on the predicted CI listener results. This might better predict the variation as observed in Pretorius and Hanekom (2008), because common abnormalities in the implantation of CIs and the variation in nerve survival in CI users affect individuals differently and can be responsible for the variation observed in experimental data (e.g. Nelson *et al.*, 1996).

The IRE model contains several simplifications of actual neural encoding. It is conceivable that processing mechanisms that might be essential in the encoding of intensity or frequency have been excluded from the IRE model. The discussion in the paragraphs above provides an

initial consideration of some of the factors that could influence the IRE model's predictions, but the discussion is not exhaustive and it must be recognised that many details could be refined, some of which are alluded to in this section. However, the model provides accurate predictions of discrimination thresholds for two different auditory percepts by using a basic model of peripheral stimulation, and where there are discrepancies with expected trends, it appears possible to account for these. In addition, there is a trade-off between conceptual simplicity and detailed complexity in any mathematical model. A good model should use the minimum amount of complexity to predict the data, which was one of the objectives of this study.

The goal of this study was to predict the outcomes of two perceptual experiments for acoustic and electrical hearing by varying as few parameters as possible in the central mechanism for the different tasks. The only differences between the central mechanisms used for intensity and frequency discrimination were in the internal representations describing the expected input. This approach has proven to be successful for both hearing modalities (acoustic and electrical hearing) and is supported by other successful models that have utilised internal representations to estimate incoming information (Hanekom and Krüger, 2001; Srulovicz and Goldstein, 1983). This implies that when performing two different discrimination tasks, the CANS need not drastically change its strategy in interpreting information provided to it by AP patterns; it need only know what is normally expected.

In the present model, the internal representation is the expected pattern for acoustic stimulation in the case of both hearing modalities. This implicitly assumes that the internal representation has not changed since the onset of the model subject's[1] deafness (and that the model subject was post-lingually deafened). The expression of neural plasticity in the brain plays a role in the adjustment of the CI to the difference in input (Møller, 2006; Wilson and Dorman, 2008b) or as a result of neural damage (Sanes and Bao, 2009), which implies that there could be some change to the internal representation utilised by the CANS (also suggested by Beitel, Snyder, Schreiner, Raggio and Leake, 2000). Owing to the fact that the model only matches experimental data over a limited range, it is not clear to what degree this has an effect, but it could certainly contribute to the differences observed between the two modalities. Age could also have an effect on this. Clinard, Tremblay and Krishnan (2010) show how age alters the perception and physiological representation of frequency, and this could conceivably have an effect on internal models.

[1]Model subject refers to the hypothetical listener that is represented by the model.

An important aspect of the predicted results for frequency discrimination is the difference that arises through the use of two different decoding mechanisms. The suggestion that the populations of neurons use a maximum likelihood-type mechanism is not unprecedented (Pouget *et al.*, 2000; Pouget, Dayan and Zemel, 2003). A study by Jazayeri and Movshon (2006) showed how a biologically plausible mechanism for a maximum likelihood mechanism could be implemented in the visual system. They applied this mechanism to detection, discrimination and identification tasks in motion detection. This suggests that the use of such a mechanism is plausible in the CANS, too, and the excellent performance of the model when a maximum likelihood decoder (the Viterbi decoder) is used for both percepts supports this. This does not necessarily mean that the human auditory system uses a Viterbi decoder specifically, but the idea that a maximum likelihood prediction is made based on a probability distribution by comparing the input with an internal representation is upheld (see equation 3.25). This is in contrast with other optimal models that require the addition of unspecified internal noise (Gresham and Collins, 1999b; Jepsen *et al.*, 2008) in order for the correct prediction of experimental data. The sub-optimal performance of the Viterbi decoder is likely due to the discrepancy between the estimated rate and the internal representation to which it is compared, creating a high level of uncertainty in the estimated instantaneous state. Because this sub-optimal performance yields performance levels similar to those of normal hearing, it appears that the discrepancy between the estimated rate and the internal model (brought on by the variability in the observation probabilty, see equation 3.28) is responsible for the correct magnitude prediction of frequency discrimination data.

It is conceivable that the poorer performance of the Viterbi decoder in comparison to the COG decoder is due to poor parameter choices, but this is not likely to be the case given the good magnitude predictions made by the model for two different percepts and the two different hearing modalities. This is further supported by the accurate magnitude predictions provided by other models that utilise an internal model in the decoding strategy (e.g. Hanekom and Krüger, 2001). It is striking that two decoders would yield such vastly different results based on the same information, which emphasises the importance of choosing the correct decoding strategy in perceptual models of this nature, as two strategies can differ greatly in the interpretation of data, resulting in the difference between experimentally comparative and vastly different results. Judging from the results in this study, it seems that maximum likelihood decoders provide a good method of decoding population-coded information in the CANS. Maximum likelihood decoding might be preferable to a COG decoding method, because the COG decoder may be more sensitive to noise sources that are not considered in this model. Snippe (1996) states that it is unlikely that a COG mechanism is used

in neural systems as they are very sensitive to nonlinearities (such as the case when sensor positions, in this case channel positions, are very irregular) and suggested that mechanisms such as maximum likelihood be used when interpreting population codes.

The population-coding approach of this model has important implications in the field of auditory perceptual modelling. It is explained in chapter 2 how a temporal coding mechanism is likely to be the main coding mechanism used at lower frequencies (further reviews are provided in Moller, 1999; Moore, 2003). The findings for both percepts suggest that the excitation pattern and the tuning have a large impact on the results. These are inherently place-code related mechanisms. It is therefore interesting that the model can reproduce the unique bowl-shaped trend of frequency discrimination at low frequencies using either decoder.

At this point, it is important to distinguish between temporal codes and the traditional "spike-counting" or average rate-place codes. Rieke *et al.* (1997) go to great lengths to emphasise that it is difficult to create clear boundaries between a rate-place code and a temporal code, and suggest that the distinction is rather between long and short windows of time. The IRE model is clearly a short-window model, and in this sense could be said to be closer to a temporal model that considers place. The idea of using a short-window approach is supported by Viemeister and Wakefield (1991), who concluded that auditory input is probably sampled at a fairly high rate.

The IRE mechanism takes some temporal information into account, as the timing between spikes affects the instantaneous estimate of the rate on a given channel (see figure 3.16 and equation 3.23), which directly affects the decoded signal for that instant. The use of instantaneous rate is in sharp contrast to rate-place codes that only use an average rate. This does not mean that the model considers phase locking or synchronization: this is impossible as these mechanisms are not built into the model. It does suggest, however, that the variation in spike timing is considered through the instantaneous representation of the spike rate, which directly affects the variation in the decoded state. This implies that even though the model has been classified as an instantaneous rate-place model in this study, it may be considered to reside in a class of its own, somewhere between rate-place code models and temporal code models.

It is mentioned by Heinz *et al.* (2001a) that a model that is intermediate to rate-place (e.g. Siebert, 1970) and all-information schemes (a restricted temporal model, e.g. Goldstein and

Srulovicz, 1977) should be used for predicting the percepts of frequency and intensity discrimination. The IRE model, or a model that utilises instantaneous rate instead of average rate, may be candidates for this gap. Oxenham, Bernstein and Penagos (2004) emphasise the important role that a code combining place and temporal information has to play in the context of complex pitch perception. They suggest that efforts to improve the temporal and fine structure information transmission through CIs may be fruitless if the correct tonotopic mapping is not provided. This also applies to intensity and frequency discrimination. Some have gone as far as to suggest that timing information may not be used at all by the brain (Shadlen and Newsome, 1994). Clark (2003) notes how the substantial jitter in the initiation of APs by a 1000 Hz tone suggests that a combined place and temporal coding mechanism is probably used in the auditory system. These examples support the relevance of the model and its mechanism of coding as a basis for future models of this kind.

A crucial aspect of this study is what the results say about the CANS. Considering that the model is successful in predicting perceptual magnitudes associated with both acoustic and electrical hearing, it is reasonable to assume that the strength of the model lies in the central processing mechanism. By assuming that the deviations from experimental data result from simplifications in the peripheral hearing models, the results suggest that the central estimation mechanism model provides a suitable representation of the processing that occurs in the CANS. This implies that because of noise on neural signals (Faisal, Selen and Wolpert, 2008), the brain uses an estimation mechanism to effectively obtain the signal through the noise. It obtains a best guess of the signal at each time instant (see Viemeister and Wakefield, 1991) for a population of fibres and attempts to find the most likely state of the input based on a probability distribution (see equation 3.27) and a learned, reference set of information.

The idea that the brain must address the issue of noise explicitly (as was emphasised by Eliasmith and Anderson, 2003), supports this notion, and the biological plausibility of probability mechanisms being present in our brain supports such an explanation (Jazayeri and Movshon, 2006). This explanation is based on the processing that occurs in the model (see sections 3.3.4 and 3.3.5) and is not intended to be definitive, but is rather a suggestion of the possible neural processing that occurs based on the strength of the model.

From the perspective of the central estimation model, the only differences between the acoustic and the electrical stimulation are in the spike train patterns. It has been suggested that the statistical independence of neural firing plays a role in perception (Zeng *et al.*, 2000).

Electrically stimulated neural fibres are statistically dependent because of the high level of synchronisation resulting from the wide spread of electrical stimulation. This is one of the main differences in the firing patterns in the IRE model. Therefore, the model seems to support the hypothesis that the statistical independence of spike trains plays a role in perception. This and the high level of sensitivity of the model to excitation patterns and tuning suggest that in the design of processing strategies for CIs, the statistical independence of neural firing should be mimicked in some way.

If the modelled CI replicated the neural firing of a NH listener perfectly, it would have predicted frequency discrimination results of NH listeners. Frequency discrimination ability has been linked to speech perception ability in CI listeners (Dorman *et al.*, 1996) and therefore, such mimicry may better facilitate CI hearing. Echoing the sentiments of Oxenham *et al.* (2004), such results highlight the importance of considering the central processing mechanisms in the design of next-generation CIs.

6.3 SUMMARY

This chapter provided a discussion of the work in a context relating to both intensity and frequency discrimination for acoustic and electrical hearing. The most important parameters used in the model, as well as possible improvements thereto, were explored. The implications for the coding mechanism used by the auditory system were presented in terms of the functioning of the central estimation mechanism. How these factors affect the field of CI research was also mentioned. The next chapter provides a brief overview of the work, including a discussion of the answers to the research questions, as well as directions for future work.

CHAPTER 7

CONCLUSION

A brief summary of the study is provided in this section and includes a short review of the study and the main findings. Detailed discussions of the study can be found in sections 4.4 and 5.4, as well as in chapter 6. This chapter is concluded with suggestions for future work.

7.1 RESEARCH OVERVIEW

The primary goal of this study was to investigate the feasibility of using a single model of the CANS (using an instantaneous rate-place coding mechanism) to predict the outcomes of perceptual experiments for both NH and CI listeners in a sound field environment. The following list shows how the research questions that were listed in chapter 1 were answered.

- A literature study was carried out to identify the simplest methods of modelling a single central estimator and peripheral hearing models for both acoustic and electrical hearing.
- A single central estimation model was developed and used to predict the outcomes of perceptual experiments for both NH and CI listeners, utilising the space-time AP patterns generated by models of peripheral hearing for acoustic and electrical stimulation of the cochlea (chapter 3).

- Predictions of frequency discrimination and intensity discrimination results were made for both NH and CI listeners (sections 4.3 and 5.3). These predictions were compared with experimental results available in literature (all figures 4.1 through to 5.4, except figure 4.2).
- The model was found to provide accurate indications of the trends and magnitudes observed in experimental data within a restricted set of parameters (see figures 4.3 and 4.5 for examples).
- The implications of the findings were considered and discussed within the context of available literature (sections 4.4 and 5.4 and chapter 6).

The main findings of the study are as follows.

- The IRE model was used to predict the outcomes of frequency discrimination experiments for NH and CI listeners and compared to data available in the literature that was obtained in sound field conditions (chapter 4). A COG decoder and a Viterbi decoder were used. The model showed that the frequency dependence of frequency discrimination for NH listeners could be predicted over a limited range of frequencies, up to 3000 Hz for the Viterbi decoder and up to approximately 5000 Hz for the COG decoder. In the region in which the correct trend was predicted, the Viterbi decoder yielded magnitudes similar to those in experimental data, whereas the COG decoder yielded magnitudes similar to those of other optimal detection models (figure 4.1). The same trends were observed for the dependence of frequency discrimination on intensity for NH listeners (figure 4.3). The difference between the decoders was attributed to the uncertainty introduced by the Viterbi algorithm's use of a probability distribution to yield a maximum likelihood prediction (section 4.4). This is in contrast with the COG decoder's predictions being based solely on the estimated maximum amplitude, which is clearly provided in the model by the constant spontaneous rate and threshold in the neurons. The discrepencies in the trend were attributed to simplifications made in the model (figure 4.2 and section 4.4).
- For the CI predictions of the frequency dependence of frequency discrimination, the discrepancy between the two decoders was similar to that found in NH conditions, and in the case of the Viterbi decoder, the predicted magnitudes were in the same order of magnitude as experimental values for CI listeners in a sound field environment

(figure 4.5). There was not as much variation in the modelled results as was observed in CI listeners, but this was attributed to the varying pathologies of human subjects (section 4.4).

- After the success of the frequency discrimination predictions, the Viterbi decoder was used to predict intensity discrimination results for NH and CI listeners, and these were compared with data from literature (chapter 5). Again, the model was able to predict the approximate order of magnitude of experimental intensity discrimination results for the frequency and intensity dependence of intensity discrimination for NH listeners (figures 5.1 and 5.2). The predicted trends were only correct over a limited range of intensities and frequencies, and this was attributed to simplifying assumptions made in the model (sections 3.3.2 and 5.4). Predictions were made for the intensity discrimination abilities of CI listeners as a function of intensity (figure 5.4). The model was able to predict the correct order of magnitude observed in literature for CI listeners as well.

- The dependence on duration was predicted for NH listeners for both frequency and intensity discrimination. The model was able to account for the approximate order of magnitude at longer durations when using the Viterbi decoder, but the predicted trend was very different from that observed in literature and in other models (figures 4.4 and 5.3). This was attributed to the lack of certain temporal features of neural firing, including neural adaptation and onset and offset effects, as well as phase locking (sections 4.4, 5.4 and 6.2).

- From the accurate predictions observed in the model under many conditions, it was suggested that the model provides a valid representation of the processing that occurs in the CANS. This would imply that the CANS makes an approximation of the instantaneous firing rate based on the input space-time AP patterns.

- The successful model predictions also suggest that the CANS makes a maximum likelihood prediction of the input frequency/intensity based on an internal representation and a probability distribution (section 6.2) using its estimations of the instantaneous firing rate. This finding is supported by studies of the central processing of motion detection, and supports the idea that more emphasis needs to be placed on the central processes in the auditory system in order to develop next-generation CIs.

7.2 FUTURE WORK

This study provides a basis for much future work, and suggestions for such work are listed here so that further studies can develop from the present work.

One of the most important suggestions of the results is that the CANS might work like an instantaneous estimator along with a maximum likelihood estimator. Though this has been suggested before in the more general context of neural coding, as well as in motion detection, it is important to attempt to verify this finding in the context of the CANS. This might be achieved by developing biologically plausible models of the CANS that implement maximum likelihood estimators. Psychological or behavioral tests could be designed to test whether the CANS makes use of this mechanism. If the hypothesis is found to be true, future CIs could be designed with this consideration in mind to provide better perception abilities to implantees.

The model provides a solid platform on which future models can be developed. The approach that was taken succeeded in representing two different percepts for both acoustic and electrical hearing modalities, within a limited set of parameters. Much can be added to the model to potentially provide better predictions of the data. Some of the more important improvements would be the inclusion of a range of spontaneous rates and thresholds in the data, as well as better representation of the RI functions. This would make the excitation patterns closer to those that appear biologically, and owing to the sensitivity of the data on these parameters, this might result in a more accurate model. Other improvements that could be made are the inclusion of phase-locking effects, neural adaptation and onset and offset effects. This would probably improve the duration dependence of the model and could affect the CI results, which are highly synchronised. There are already many studies that provide quantitative information on these aspects (see section 6.2), but inclusion of these features in the model would not be trivial as this may require modification of the central estimator to account for the changes. The expansion of the model to include other psychoacoustic percepts could be an option for advancement of the model. This would require the quantification of the effects of the percepts on firing rate, which could be made possible by utilising models such as that of Jepsen *et al.* (2008) as a basis.

During the development of the model, the quantitative definition of certain aspects of the RIs functions was conspicuously absent. Although the change in threshold has been well documented and modelled, there does not appear to be any quantification of the change in

the maximum firing rate as a function of change in frequency for RI functions. Moreover, the definition of the maximum firing rate for sloping-saturating and straight RI functions does not appear to be generally defined (rather, maximum firing rate has been fitted to the measured data in studies such as Köppl and Yates, 1999). As a result, a very rough approximation was used in this model, from data obtained in previous studies. If the RI functions could be generally defined over all characteristic frequencies, future studies of this nature would be able to make use of a more biologically plausible set of data.

The computing requirements of the model in this study were not emphasised as they were not the focus of the study. In fact, the central estimation model described in section 3.3.4.2 is extremely computationally expensive due to the processing requirements of computing matrix exponentials (equation 3.18). A mathematical study involving an efficient instantaneous rate estimator could be performed so that simulations for this type of model could be performed faster.

Because of the substantial dependence of the model on the internal representation in predicting the data, it would be interesting to investigate the effects of altering the internal representation on the predicted percepts. This is alluded to in section 6.2. Neural plasticity has an effect on CI listeners' abilities to perceive sounds, as do the varying pathologies among implantees (Møller, 2006; Schuknecht, 1993). If different pathologies or the changes resulting from neural plasticity could be quantified, this could be applied to the model and would potentially provide insight into the extent that these affect electrical hearing. If the effects of different pathologies were investigated, insight would also potentially be provided into the needs of individual subjects with CIs. This could contribute to advancements in custom signal processing techniques to improve the level of hearing ability based on the specific needs of an individual.

REFERENCES

Arndt, P., Staller, S., Arcaroli, J., Hines, A. and Ebinger, K. (1999). *Within-subject comparison of advanced coding strategies in the Nucleus 24 cochlear implant*, Cochlear Corporation, Englewood, CO. Technical report.

Arnoldner, C., Riss, D., Brunner, M., Durisin, M., Baumgartner, W.-D. and Hamzavi, J.-S. (2007). Speech and music perception with the new fine structure speech coding strategy: preliminary results, *Acta Oto-Laryngologica* **127**(12): 1298–1303.

Assefi, T. (1979). *Stochastic Processes and Information Theory with Applications*, Wiley-Interscience, John Wiley & Sons, Inc., New York.

Bacon, S.P. (2004). Overview of auditory compression, *in* S. Bacon, R. Fay and A. Popper (eds), *Compression: From Cochlea to Cochlear Implants*, Springer Handbook of Auditory Research, Spinger-Verlag, New York, chapter 1, pp. 1–17.

Barretto, R.L. and Pfingst, B.E. (1992). Electrical stimulation of the auditory nerve: effects of pulse width on frequency discrimination, *Hearing Research* **62**(2): 245–249.

Baskent, D. and Shannon, R.V. (2005). Interactions between cochlear implant electrode insertion depth and frequency-place mapping, *Journal of the Acoustical Society of America* **117**(3): 1405–1416.

Beitel, R.E., Snyder, R.L., Schreiner, C.E., Raggio, M.W. and Leake, P.A. (2000). Electrical cochlear stimulation in the deaf cat: comparisons between psychophysical and central auditory neuronal thresholds, *Journal of Neurophysiology* **83**(4): 2145–2162.

Bilger, R.C., Black, F.O. and Hopkinson, N.T. (1977). Implanted auditory prosthesis: an evaluation of subjects presently fitted with cochlear implants, *Transactions of the American Academy of Ophthalmology and Otolaryngology* **84**(4): 677–682.

Bingabr, M., Espinoza-Varas, B. and Loizou, P.C. (2008). Simulating the effect of spread of excitation in cochlear implants, *Hearing Research* **241**: 73–79.

Bleeck, S., Ives, T. and Patterson, R.D. (2004). Aim-mat: the auditory image model in MATLAB, *Acta Acustica united with Acustica* **90**(4): 781–787.

Boëx, C., De Balthasar, C., Kós, M.-I. and Pelizzone, M. (2003). Electrical field interactions in different cochlear implant systems, *Journal of the Acoustical Society of America* **114**(4): 2049–2057.

Bregman, A.S. (2001). *Auditory Scene Analysis: The Perceptual Organization of Sound*, 2nd edn, MIT Press, Massachusetts Institute of Technology.

Briaire, J.J. and Frijns, J.H.M. (2000). 3D mesh generation to solve the electrical volume conduction problem in the implanted inner ear, *Simulation Practice and Theory* **8**: 57–73.

Bruce, I.C. (1997). *Spatiotemporal coding of sound in the auditory nerve for cochlear implants*, PhD thesis, Department of Otolaryngology, The University of Melbourne, Melbourne.

Bruce, I.C., Irlicht, L.S., White, M.W., O'Leary, S.J. and Clark, G.M. (2000). Renewal-process approximation of a stochastic threshold model for electrical neural stimulation, *Journal of Computational Neuroscience* **9**(2): 119–132.

Bruce, I.C., Irlicht, L.S., White, M.W., O'Leary, S.J., Dynes, S., Javel, E. and Clark, G.M. (1999a). A stochastic model of the electrically stimulated auditory nerve: pulse-train response, *IEEE Transactions on Biomedical Engineering* **46**(6): 630–637.

Bruce, I.C., White, M.W., Irlicht, L.S., O'Leary, S.J. and Clark, G.M. (1999b). The effects of stochastic neural activity in a model predicting intensity perception with cochlear implants: low-rate stimulation, *IEEE Transactions on Biomedical Engineering* **46**(12): 1393–1404.

Bruce, I.C., White, M.W., Irlicht, L.S., O'Leary, S.J., Dynes, S., Javel, E. and Clark, G.M. (1999c). A stochastic model of the electrically stimulated auditory nerve: single-pulse response, *IEEE Transactions on Biomedical Engineering* **46**(6): 617–629.

Büchner, A., Nogueira, W., Edler, B, Battmer, R.-D. and Lenarz, T. (2008). Results from a psychoacoustic model-based strategy for the Nucleus-24 and Freedom cochlear implants, *Otology and Neurotology* **29**(2): 189–192.

Buechner, A., Brendel, M., Krüeger, B., Frohne-Büchner, C., Nogueira, W., Edler, B. and Lenarz, T. (2008). Current steering and results from novel speech coding strategies, *Otology and Neurotology* **29**(2): 203–207.

Busby, P.A. and Clark, G.M. (1996). Electrode discrimination by early-deafened cochlear implant patients, *Audiology* **35**(1): 8–22.

Buus, S. (1990). Level discrimination of frozen and random noise, *Journal of the Acoustical Society of America* **87**(6): 2643–2654.

Carney, L.H. (1993). A model for the responses of low-frequency auditory-nerve fibers in cat, *Journal of the Acoustical Society of America* **93**(1): 401–417.

Carney, L.H. (1994). Spatiotemporal encoding of sound level: models for normal encoding and recruitment of loudness, *Hearing Research* **76**: 31–44.

Chen, F. and Zhang, Y.-T. (2007). An integrate-and-fire-based auditory nerve model and its response to high-rate pulse train, *Neurocomputing* **70**(4-6): 1051–1055.

Chen, H. and Zeng, F.-G. (2004). Frequency modulation detection in cochlear implant subjects, *Journal of the Acoustical Society of America* **116**(4): 2269–2277.

Clark, G.M. (1996). Electrical stimulation of the auditory nerve: the coding of frequency, the perception of pitch and the development of cochlear implant speech processing strategies for profoundly deaf people, *Clinical and Experimental Pharmacology and Physiology* **23**: 766–776.

Clark, G.M. (2003). *Cochlear Implants: Fundamentals and Applications*, Modern Acoustics and Signal Processing, Springer-Verlag, New York.

Clark, G.M., Patrick, J.F. and Bailey, Q. (1979). A cochlear implant round window electrode array, *Journal of Laryngology and Otology* **93**(2): 107–109.

Clark, G.M., Tong, Y.C., Black, R., Forster, I.C., Patrick, J.F. and Dewhurst, D.J. (1977). A multiple electrode cochlear implant, *Journal of Laryngology and Otology* **91**(11): 935–945.

Clinard, C.G., Tremblay, K.L. and Krishnan, A.R. (2010). Aging alters the perception and physiological representation of frequency: evidence from human frequency-following response recordings, *Hearing Research* **264**: 48–55.

Cochlear Ltd (2002). *ACE™ and CIS DSP Strategies: Software Requirements Specification*. Part Number N95287F, Issue 1.

Cohen, L.T. (2009a). Practical model description of peripheral neural excitation in cochlear implant recipients: 1. growth of loudness and ecap amplitude with current, *Hearing Research* **247**: 87–99.

Cohen, L.T. (2009b). Practical model description of peripheral neural excitation in cochlear implant recipients: 2. spread of the effective stimulation field (esf), from ecap and fea, *Hearing Research* **247**: 100–111.

Cohen, L.T. (2009c). Practical model description of peripheral neural excitation in cochlear implant recipients: 3. ecap during bursts and loudness as function of burst duration, *Hearing Research* **247**: 112–121.

Cohen, L.T. (2009d). Practical model description of peripheral neural excitation in cochlear implant recipients: 4. model development at low pulse rates: General model and application to individuals, *Hearing Research* **248**: 15–30.

Cohen, L.T. (2009e). Practical model description of peripheral neural excitation in cochlear implant recipients: 5. refractory recovery and facilitation, *Hearing Research* **248**: 1–14.

Cohen, L.T., Saunders, E. and Clark, G.M. (2001). Psychophysics of a prototype perimodiolar cochlear implant electrode array, *Hearing Research* **155**: 63–81.

Cohen, M.A., Grossberg, S. and Wyse, L.L. (1995). A spectral network model of pitch perception, *Journal of the Acoustical Society of America* **98**(2): 862–879.

Colburn, H.S. (1973). Theory of binaural interaction based on auditory nerve data. I. General strategy and preliminary results on interaural discrimination, *Journal of the Acoustical Society of America* **54**(6): 1458–1470.

Collins, L.M., Zwolan, T.A. and Wakefield, G.H. (1997). Comparison of electrode discrimination, pitch ranking, and pitch scaling data in postlingually deafened adult cochlear implant subjects, *Journal of the Acoustical Society of America* **101**(1): 440–455.

Cox, D.R. (1962). *Renewal Theory*, Methuen, London.

Creelman, C.D. (1962). Human discrimination of auditory duration, *Journal of the Acoustical Society of America* **34**(5): 582–593.

Dau, T., Kollmeier, B. and Kohlrausch, A. (1997a). Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers, *Journal of the Acoustical Society of America* **102**(5): 2892–2905.

Dau, T., Kollmeier, B. and Kohlrausch, A. (1997b). Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration, *Journal of the Acoustical Society of America* **102**(5): 2906–2919.

Dau, T., Püschel, D. and Kohlrausch, A. (1996a). A quantitative model of the "effective" signal processing in the auditory system. I. Model structure, *Journal of the Acoustical Society of America* **99**(6): 3615–3622.

Dau, T., Püschel, D. and Kohlrausch, A. (1996b). A quantitative model of the "effective" signal processing in the auditory system. II. Simulations and measurements, *Journal of the Acoustical Society of America* **99**(6): 3623–3631.

Dawson, P.W., McKay, C.M., Busby, P.A., Grayden, D.B. and Clark, G.M. (2000). Electrode discrimination and speech perception in young children using cochlear implants, *Ear and Hearing* **21**(6): 597–607.

Dillier, N., Spillmann, T. and Guntensperger, J. (1983). Computerized testing of signal-encoding strategies with round-window implants, *Annals of the New York Academy of Sciences* **405**: 360–369.

Dorman, M.F., Loizou, P.C. and Rainey, D. (1997a). Simulating the effect of cochlear-implant electrode insertion depth on speech understanding, *Journal of the Acoustical Society of America* **102**(5): 2993–2996.

Dorman, M.F., Loizou, P.C. and Rainey, D. (1997b). Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs, *Journal of the Acoustical Society of America* **102**(4): 2403–2411.

Dorman, M.F., Smith, L.M., Smith, M. and Parkin, J.L. (1996). Frequency discrimination and speech recognition by patients who use the Ineraid and continuous interleaved sampling cochlear-implant signal processors, *Journal of the Acoustical Society of America* **99**(2): 1174–1184.

Drennan, W.R. and Pfingst, B.E. (2005). Current-level discrimination using bipolar and monopolar electrode configurations in cochlear implants, *Hearing Research* **202**: 170–179.

Drennan, W.R. and Pfingst, B.E. (2006). Current-level discrimination in the context of interleaved, multichannel stimulation in cochlear implants: effects of number of stimulated electrodes, pulse rate, and electrode separation, *JARO – Journal of the Association for Research in Otolaryngology* **7**(3): 308–316.

Dye, R.H. Jr and Hafter, E.R. (1980). Just-noticeable differences of frequency for masked tones, *Journal of the Acoustical Society of America* **67**(5): 1746–1753.

Eliasmith, C. and Anderson, C.H. (2003). *Neural Engineering: Computation, Representation, and Dynamics in Neurobiological Systems*, MIT Press, Massachusetts Institute of Technology.

Elliot, D.N. and Fraser, W.R. (1970). Fatigue and adaptation, *in* J. Tobias (ed.), *Foundations In Modern Auditory Theory*, Academic Press, New York.

Erell, A. (1988). Rate coding model for discrimination of simple tones in the presence of noise, *Journal of the Acoustical Society of America* **84**(1): 204–214.

Evans, E.F. (1978). Place and time coding of frequency in the peripheral auditory system: some physiological pros and cons, *Audiology* **17**(5): 369–420.

Faisal, A.A., Selen, L.P.J. and Wolpert, D.M. (2008). Noise in the nervous system, *Nature Reviews Neuroscience* **9**(4): 292–303.

Faulkner, A., Rosen, S. and Stanton, D. (2003). Simulations of tonotopically mapped speech processors for cochlear implant electrodes varying in insertion depth, *Journal of the Acoustical Society of America* **113**(2): 1073–1080.

Florentine, M. (1983). Intensity discrimination as a function of level and frequency and its relation to high-frequency hearing, *Journal of the Acoustical Society of America* **74**(5): 1375–1379.

Florentine, M. (1986). Level discrimination of tones as a function of duration, *The Journal of the Acoustical Society of America* **79**(3): 792–798.

Florentine, M. and Buus, S. (1981). An excitation-pattern model for intensity discrimination, *Journal of the Acoustical Society of America* **70**(6): 1646–1654.

Florentine, M., Buus, S. and Mason, C.R. (1987). Level discrimination as a function of level for tones from 0.25 to 16 kHz, *Journal of the Acoustical Society of America* **81**(5): 1528–1541.

Forney, G.D. Jr (1973). Viterbi algorithm, *Proceedings of the IEEE* **61**(3): 268–278.

Forrest, T.G. and Formby, C. (1996). Detection of silent temporal gaps in sinusoidal markers simulated with a single-channel envelope detector model, *Auditory Neuroscience* **3**: 21–33.

Frankenhaeuser, B. and Huxley, A.F. (1964). The action potential in the myelinated nerve fibre of *Xenopus laevis* as computed on the basis of voltage clamp data, *The Journal of Physiology* **171**: 302–315.

Friesen, L.M., Shannon, R.V., Baskent, D. and Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants, *Journal of the Acoustical Society of America* **110**(2): 1150–1163.

Galvin, J.J. III and Fu, Q.-J. (2005). Effects of stimulation rate, mode and level on modulation detection by cochlear implant users, *JARO – Journal of the Association for Research in Otolaryngology* **6**(3): 269–279.

Galvin, J.J. III and Fu, Q.-J. (2009). Influence of stimulation rate and loudness growth on modulation detection and intensity discrimination in cochlear implant users, *Hearing Research* **250**: 46–54.

Geldenhuys, T.A. (2007). *Temporal gap detection in electric hearing: modeling and experiments*, Master of engineering (bio-engineering), Faculty of Engineering, Built Environment and Information Technology, University of Pretoria, Pretoria.

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

Goldstein, J.L. and Srulovicz, P. (1977). Auditory-nerve spike intervals as an adequate basis for aural frequency measurement, *in* E. Evans and J. Wilson (eds), *Psychophysics and Physiology of Hearing*, Academic Press, London, pp. 337–347.

Green, D.M. and Swets, J.A. (1966). *Signal Detection Theory and Psychophysics*, John Wiley and Sons, Inc., New York.

Greenwood, D.D. (1961). Critical bandwidth and the frequency coordinates of the basilar membrane, *Journal of the Acoustical Society of America* **33**(10): 1344–1356.

Greenwood, D.D. (1990). A cochlear frequency-position function for several species – 29 years later, *Journal of the Acoustical Society of America* **87**(6): 2592–2605.

Gresham, L.C. and Collins, L.M. (1998). Analysis of the performance of a model-based optimal auditory signal processor, *Journal of the Acoustical Society of America* **103**(5): 2520–2529.

Gresham, L.C. and Collins, L.M. (1999a). A comparison using signal detection theory of the ability of two computational auditory models to predict experimental data, *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing – Proceedings* **2**: 933–936.

Gresham, L.C. and Collins, L.M. (1999b). Effect of a Poisson 'internal noise' process on theoretical acoustic signal detectability, *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, IEEE, Duke University, Durham, United States, pp. 219–222.

Hanekom, J.J. (1999). A model of frequency coding in the central auditory nervous system, *The South African Journal of Communication Disorders* **46**: 83–90.

Hanekom, J.J. (2000). What do cochlear implants teach us about the encoding of frequency in the auditory system?, *The South African Journal of Communication Disorders* **47**: 49–56.

Hanekom, J.J. and Krüger, J.J. (2001). A model of frequency discrimination with optimal processing of auditory nerve spike intervals, *Hearing Research* **151**: 188–204.

Hanekom, T. and Hanekom, J.J. (2000). Die bydrae van basiese navorsing in kliniese toepassings met verwysing na kogleêre inplantings, *The South African Journal of Communication Disorders* **47**: 41–47.

Hartmann, W.M. (1998). *Signals, Sound, and Sensation*, Modern Acoustics and Signal Processing, Springer Science+Business Media, LLC, New York.

Hatsushika, S., Shepherd, R.K., Tong, Y.C., Clark, G.M. and Funasaka, S. (1990). Dimensions of the scala tympani in the human and cat with reference to cochlear implants, *Annals of Otology, Rhinology and Laryngology* **99**(11): 871–876.

He, N.-J., Dubno, J.R. and Mills, J.H. (1998). Frequency and intensity discrimination measured in a maximum-likelihood procedure from young and aged normal-hearing subjects, *Journal of the Acoustical Society of America* **103**(1): 553–565.

Heinz, M.G. (2000). *Quantifying the effects of the cochlear amplifier on temporal and average-rate information in the auditory nerve*, Doctor of Philosphy in Speech and Hearing Sciences, Massachusetts Institute of Technology, Cambridge, MA.

Heinz, M.G., Colburn, H.S. and Carney, L.H. (2001a). Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve, *Neural Computation* **13**(10): 2273–2316.

Heinz, M.G., Colburn, H.S. and Carney, L.H. (2001b). Evaluating auditory performance limits: II. One-parameter discrimination with random-level variation, *Neural Computation* **13**(10): 2317–2338.

Heinz, M.G., Colburn, H.S. and Carney, L.H. (2001c). Rate and timing cues associated with the cochlear amplifier: level discrimination based on monaural cross-frequency coincidence detection, *Journal of the Acoustical Society of America* **110**(4): 2065–2084.

Heinz, M.G., Colburn, H.S. and Carney, L.H. (2002). Quantifying the implications of nonlinear cochlear tuning for auditory-filter estimates, *Journal of the Acoustical Society of America* **111**(2): 996–1011.

Helmerich, L.F. and Edgerton, B.J. (1982). Psychoelectric measurements and results from cochlear implant patients, *Annals of Otology, Rhinology and Laryngology* **Suppl. 91**(2): 35–40.

Henning, G.B. (1966). Frequency discrimination of random-amplitude tones., *Journal of the Acoustical Society of America* **39**(2): 336–339.

Hienz, R.D., Sachs, M.B. and Aleszczyk, C.M. (1993). Frequency discrimination in noise: comparison of cat performances with auditory-nerve models, *Journal of the Acoustical Society of America* **93**(1): 462–469.

Hill, A. V. (1936). Excitation and accommodation in nerve, *Proceedings of the Royal Society of London, Series B, Biological Sciences* **119**(814): 305–355.

Hochmair, I., Nopp, P., Jolly, C., Schmidt, M., Schößer, H., Garnham, C. and Anderson, I. (2006). MED-EL cochlear implants: state of the art and a glimpse into the future, *Trends in Amplification* **10**(4): 201–220.

Hodgkin, A.L. and Huxley, A.F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve, *The Journal of Physiology* **117**(4): 500–544.

House, W.F. and Edgerton, B.J. (1982). A multiple-electrode cochlear implant, *Annals of Otology, Rhinology and Laryngology* **Suppl. 91**(2): 104–116.

Houtsma, A.J., Durlach, N.I. and Braida, L.D. (1980). Intensity perception XI. Experimental results on the relation of intensity resolution to loudness matching, *Journal of the Acoustical Society of America* **68**(3): 807–813.

Huettel, L.G. and Collins, L.M. (1999). Using computational auditory models to predict simultaneous masking data: model comparison, *IEEE Transactions on Biomedical Engineering* **46**(12): 1432–1440.

Huettel, L.G. and Collins, L.M. (2002). A theoretical analysis of the effects of auditory impairment on intensity discrimination, *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing – Proceedings*, Vol. 2, Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708-0291, United States.

Huettel, L.G. and Collins, L.M. (2004). Predicting auditory tone-in-noise detection performance: the effects of neural variability, *IEEE Transactions on Biomedical Engineering* **51**(2): 282–293.

Imennov, N.S. and Rubinstein, J.T. (2009). Stochastic population model for electrical stimulation of the auditory nerve, *IEEE Transactions on Biomedical Engineering* **56**(10): 2493–2501.

ISO 8253-2 (2009). *Acoustics – Audiometric test methods – Part 2: Sound field audiometry with pure tone and narrow-band test signals*. International Organization for Standardization.

Jackson, B.S. and Carney, L.H. (2005). The spontaneous-rate histogram of the auditory nerve can be explained by only two or three spontaneous rates and long-range dependence, *JARO – Journal of the Association for Research in Otolaryngology* **6**(2): 148–159.

Javel, E. (1990). Acoustic and electrical encoding of temporal information, *in* J. M. Miller and F. A. Spelman (eds), *Cochlear Implants: Models of the Electrically Stimulated Ear*, Springer-Verlag, Berlin, Germany, pp. 245–295.

Javel, E. and Mott, J.B. (1988). Physiological and psychophysical correlates of temporal processes in hearing, *Hearing Research* **34**: 275–294.

Javel, E. and Viemeister, N.F. (2000). Stochastic properties of cat auditory nerve responses to electric and acoustic stimuli and application to intensity discrimination, *Journal of the Acoustical Society of America* **107**(2): 908–921.

Javel, E., Tong, Y.C., Shepherd, R.K. and Clark, G.M. (1987). Responses of cat auditory nerve fibers to biphasic electrical current pulses, *Annals of Otology, Rhinology and Laryngology* **96**: 26–30.

Jazayeri, M. and Movshon, J.A. (2006). Optimal representation of sensory information by neural populations, *Nature Neuroscience* **9**(5): 690–696.

Jepsen, M.L., Ewert, S.D. and Dau, T. (2008). A computational model of human auditory signal processing and perception, *Journal of the Acoustical Society of America* **124**(1): 422–438.

Jesteadt, W., Schairer, K.S., Nizami, L., Khaddam, S. and Neely, S.T. (2009). Effects of external noise on detection of intensity increments, *Journal of the Acoustical Society of America* **126**(4): 1941–1953.

Jesteadt, W., Wier, C.C. and Green, D.M. (1977). Intensity discrimination as a function of frequency and sensation level, *Journal of the Acoustical Society of America* **61**(1): 169–177.

Johnson, D.H. (1980). The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones, *Journal of the Acoustical Society of America* **68**(4): 1115–1122.

Johnson, D.H. and Swami, A. (1983). The transmission of signals by auditory-nerve fibers to single tones, *Journal of the Acoustical Society of America* **68**: 1115–1122.

Joris, P.X., Carney, L.H., Smith, P.H. and Yin, T.C.T. (1994a). Enhancement of neural synchronization in the anteroventral cochlear nucleus. I. Responses to tones at the characteristic frequency, *Journal of Neurophysiology* **71**(3): 1022–1036.

Joris, P.X., Smith, P.H. and Yin, T.C.T. (1994b). Enhancement of neural synchronization in the anteroventral cochlear nucleus. II. Responses in the tuning curve tail, *Journal of Neurophysiology* **71**(3): 1037–1051.

Kalman, R.E. (1960). A new approach to linear filtering and prediction problems, *Transactions of the ASME – Journal of Basic Engineering* **82**: 35–45.

Kessler, D.K., Loeb, G.E. and Barker, M.J. (1995). Distribution of speech recognition results with the Clarion cochlear prostheses, *Annals of Otology, Rhinology and Laryngology* **104**: 283–285.

Kiang, N.Y.S. (1965). *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve*, Reasearch Monograph No. 35, MIT Press, Cambridge, Massachusetts.

Kiefer, J., Hohl, S., Stürzebecher, E., Pfennigdorff, T. and Gstöettner, W. (2001). Comparison of speech recognition with different speech coding strategies (SPEAK, CIS, and ACE) and their relationship to telemetric measures of compound action potentials in the Nucleus CI 24M cochlear implant system, *Audiology* **40**(1): 32–42.

Kim, D.O. (1984). Functional roles of the inner- and outer-hair-cell subsystems in the cochlea and brainstem, *in* C. Berlin (ed.), *Hearing Science*, College-Hill Press, San Diego, CA.

Kim, D.O. and Parham, K. (1991). Auditory nerve spatial encoding of high-frequency pure tones: population response profiles derived from d' measure associated with nearby places along the cochlea, *Hearing Research* **52**(1): 167–180.

Klinge, A. and Klump, G.M. (2009). Frequency difference limens of pure tones and harmonics within complex stimuli in Mongolian gerbils and humans, *Journal of the Acoustical Society of America* **125**(1): 304–314.

Kluk, K. and Moore, B.C.J. (2004). Factors affecting psychophysical tuning curves for normally hearing subjects, *Hearing Research* **194**: 118–134.

Köppl, C. and Yates, G. (1999). Coding of sound pressure level in the barn owl's auditory nerve, *Journal of Neuroscience* **19**(21): 9674–9686.

Kreft, H.A., Donaldson, G.S. and Nelson, D.A. (2004). Effects of pulse rate and electrode array design on intensity discrimination in cochlear implant users, *Journal of the Acoustical Society of America* **116**(4): 2258–2268.

LePage, E.L. (1987). A spatial template for the shape of tuning curves in the mammalian cochlea, *Journal of the Acoustical Society of America* **82**(1): 155–164.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics, *Journal of the Acoustical Society of America* **49**(2): 467–477.

Liberman, M.C. (1978). Auditory-nerve response from cats raised in a low-noise chamber, *Journal of the Acoustical Society of America* **63**(2): 442–455.

Loizou, P.C. (1999a). Introduction to cochlear implants, *IEEE Engineering in Medicine and Biology Magazine* **18**(1): 32–42.

Loizou, P.C. (1999b). Signal-processing techniques for cochlear implants, *IEEE Engineering in Medicine and Biology* **18**: 34–46.

Loizou, P.C. (2006). Speech processing in vocoder-centric cochlear implants, *Advances in Oto-Rhino-Laryngology* **64**: 109–143.

Lou, H.-L. (1995). Implementing the Viterbi algorithm, *IEEE Signal Processing Magazine* **12**(5): 42–52.

Macherey, O., Carlyon, R.P., Van Wieringen, A. and Wouters, J. (2007). A dual-process integrator-resonator model of the electrically stimulated human auditory nerve, *JARO – Journal of the Association for Research in Otolaryngology* **8**(1): 84–104.

McDermott, H.J. and McKay, C.M. (1994). Pitch ranking with nonsimultaneous dual-electrode electrical stimulation of the cochlea, *Journal of the Acoustical Society of America* **96**(1): 155–162.

McKay, C.M., McDermott, H.J. and Clark, G.M. (1996). The perceptual dimensions of single-electrode and nonsimultaneous dual-electrode stimuli in cochlear implantees, *Journal of the Acoustical Society of America* **99**(2): 1079–1090.

Miller, G.A. (1947). Sensitivity to changes in the intensity of white noise and its relation to masking and loudness, *Journal of the Acoustical Society of America* **19**(4): 609–619.

Møller, A.R. (2006). Physiological basis for cochlear and auditory brainstem implants, *Advances in Oto-Rhino-Laryngology* **64**: 206–223.

Moller, A.R. (1999). Review of the roles of temporal and place coding of frequency in speech discrimination, *Acta Oto-Laryngologica (Stockholm)* **119**: 424–430.

Moore, B.C.J. (1973). Frequency difference limens for short duration tones, *Journal of the Acoustical Society of America* **54**(3): 610–619.

Moore, B.C.J. (1974). Relation between the critical bandwidth and the frequency difference limen, *Journal of the Acoustical Society of America* **55**(2): 359.

Moore, B.C.J. (2003). *An Introduction to the Psychology of Hearing*, 5th edn, Academic Press, San Diego, CA.

Moore, B.C.J. and Glasberg, B.R. (1986). The role of frequency selectivity in the perception of loudness pitch and time, *in* B. C. Moore (ed.), *Frequency Selectivity in Hearing*, Academic Press, London, chapter 5, pp. 251–308.

Moore, B.C.J. and Glasberg, B.R. (1989). Mechanisms underlying the frequency discrimination of pulsed tones and the detection of frequency modulation, *Journal of the Acoustical Society of America* **86**(5): 1722–1732.

Moore, B.C.J. and Oxenham, A.J. (1998). Psychoacoustic consequences of compression in the peripheral auditory system, *Psychological Review* **105**(1): 108–124.

Moore, B.C.J. and Raab, D.H. (1974). Pure tone intensity discrimination: some experiments relating to the "near miss" to Weber's law, *Journal of the Acoustical Society of America* **55**(5): 1049–1054.

Moore, B.C.J., Glasberg, B.R. and Hopkins, K. (2006a). Frequency discrimination of complex tones by hearing-impaired subjects: evidence for loss of ability to use temporal fine structure, *Hearing Research* **222**: 16–27.

Moore, B.C.J., Glasberg, B.R., Flanagan, H.J. and Adams, J. (2006b). Frequency discrimination of complex tones; assessing the role of component resolvability and temporal fine structure, *Journal of the Acoustical Society of America* **119**(1): 480–490.

Nelson, D.A. and Schroder, A.C. (1997). Linearized response growth inferred from growth-of-masking slopes in ears with cochlear hearing loss, *Journal of the Acoustical Society of America* **101**(4): 2186–2201.

Nelson, D.A., Donaldson, G.S. and Kreft, H. (2008). Forward-masked spatial tuning curves in cochlear implant users, *Journal of the Acoustical Society of America* **123**(3): 1522–1543.

Nelson, D.A., Schmitz, J.L., Donaldson, G.S., Viemeister, N.F. and Javel, E. (1996). Intensity discrimination as a function of stimulus level with electric stimulation, *Journal of the Acoustical Society of America* **100**(4): 2393–2414.

Nelson, E.G. and Hinojosa, R. (2004). Questioning the relationship between cochlear otosclerosis and sensorineural hearing loss: a quantitative evaluation of cochlear structures in cases of otosclerosis and review of the literature, *Laryngoscope* **114**(7): 1214–1230.

Nienhuys, T.G.W. and Clark, G.M. (1979). Critical bands following the selective destruction of cochlear inner and outer hair cells, *Acta Oto-Laryngologica* **88**(5-6): 350–358.

Nilson, M.S., Soli, D. and Sullivan, J.A. (1994). Development of the hearing in noise test for the measurement of speech recognition thresholds in quiet and in noise, *Journal of the Acoustical Society of America* **95**: 1085–1099.

Nobbe, A., Schleich, P., Zierhofer, C. and Nopp, P. (2007). Frequency discrimination with sequential or simultaneous stimulation in MED-EL cochlear implants, *Acta Oto-Laryngologica* **127**(12): 1266–1272.

Nobili, R., Mammano, F. and Ashmore, J. (1998). How well do we understand the cochlea?, *Trends in Neurosciences* **21**(4): 159–167.

Nogueira, W., Büchner, A., Lenarz, T. and Edler, B. (2005). A psychoacoustic "NofM"-type speech coding strategy for cochlear implants, *EURASIP Journal of Applied Signal Processing* **18**: 3044–3059.

Nogueira, W., Kátai, A., Harczos, T., Klefenz, F., Buechner, A. and Edler, B. (2007). An auditory model based strategy for cochlear implants, *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, Hoerzentrum Hannover, Medical University of Hannover, Germany, pp. 4127–4130.

Novitski, N., Tervaniemi, M., Huotilainen, M. and Näätänen, R. (2004). Frequency discrimination at different frequency levels as indexed by electrophysiological and behavioral measures, *Cognitive Brain Research* **20**(1): 26–36.

O'Leary, S.J., Black, R.C. and Clark, G.M. (1985). Current distributions in the cat cochlea: a modelling and electrophysiological study, *Hearing Research* **18**(3): 273–281.

O'Leary, S.J., Clark, G.M. and Tong, Y.C. (1995). Model of discharge rate from auditory nerve fibers responding to electrical stimulation of the cochlea: identification of cues for current and time-interval coding, *Annals of Otology, Rhinology and Laryngology. Supplement* **166**: 121–123.

Oxenham, A.J. and Moore, B.C.J. (1995). Overshoot and the "severe departure" from Weber's law, *Journal of the Acoustical Society of America* **97**(4): 2442–2453.

Oxenham, A.J., Bernstein, J.G.W. and Penagos, H. (2004). Correct tonotopic representation is necessary for complex pitch perception, *Proceedings of the National Academy of Sciences of the United States of America* **101**(5): 1421–1425.

Papoulis, A. (1984). *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill Series in Electrical Engineering, McGraw-Hill, New York.

Patrick, J.F., Busby, P.A. and Gibson, P.J. (2006). The development of the Nucleus Freedom cochlear implant system, *Trends in Amplification* **10**(4): 175–200.

Patterson, R.D. (2000). Auditory images : how complex sounds are represented in the auditory system, *Acoustical Science and Technology* **21**(4): 183–190.

Patterson, R.D., Allerhand, M.H. and Giguere, C. (1995). Time-domain modeling of peripheral auditory processing: a modular architecture and a software platform, *Journal of the Acoustical Society of America* **98**(4): 1890–1894.

Pfingst, B.E. (1984). Operating ranges and intensity psychophysics for cochlear implants. Implications for speech processing strategies, *Archives of Otolaryngology* **110**(3): 140–144.

Pfingst, B.E. (1988). Comparisons of psychophysical and neurophysiological studies of cochlear implants, *Hearing Research* **34**(3): 243–252.

Pfingst, B.E., Burnett, P.A. and Sutton, D. (1983). Intensity discrimination with cochlear implants, *Journal of the Acoustical Society of America* **73**(4): 1283–1292.

Plonsey, R. and Barr, R.C. (2000). *Bioelectricity: A Quantitative Approach*, 2nd edn, Kluwer Academic/Plenum Publishers, 233 Spring Street, New York.

Pouget, A., Dayan, P. and Zemel, R.S. (2000). Information processing with population codes, *Nature Reviews Neuroscience* **1**(2): 125–132.

Pouget, A., Dayan, P. and Zemel, R.S. (2003). Inference and computation with population codes, *Annual Review of Neuroscience* **26**: 381–410.

Pretorius, L.L. and Hanekom, J.J. (2008). Free field frequency discrimination abilities of cochlear implant users, *Hearing Research* **244**: 77–84.

Proakis, J.G. and Salehi, M. (2002). *Communication Systems Engineering*, 2nd edn, Prentice-Hall, Upper Saddle River, New Jersey.

Rabiner, L.H. and Juang, B.H. (1993). *Fundamentals of Speech Recognition*, Prentice-Hall, Englewood Cliffs, NJ.

Rattay, F. (1990). *Electrical Nerve Stimulation: Theory, Experiments and Applications*, Spinger-Verlag, Wien.

Rattay, F., Leao, R.N. and Felix, H. (2001a). A model of the electrically excited human cochlear neuron. II. Influence of the three-dimensional cochlear structure on neural excitability, *Hearing Research* **153**: 64–79.

Rattay, F., Lutter, P. and Felix, H. (2001b). A model of the electrically excited human cochlear neuron I. Contribution of neural substructures to the generation and propagation of spikes, *Hearing Research* **153**: 43–63.

Rieke, F., Warland, D., de Ruyter van Steveninck, R. and Bialek, W. (1997). *Spikes: Exploring the Neural Code*, MIT Press, Cambridge, Massachusetts.

Riesz, R.R. (1928). Differential intensity sensitivity of the ear for pure tones, *Physical Review* **31**(5): 867–875.

Rogers, C.F., Healy, E.W. and Montgomery, A.A. (2006). Sensitivity to isolated and concurrent intensity and fundamental frequency increments by cochlear implant users under natural listening conditions, *Journal of the Acoustical Society of America* **119**(4): 2276–2287.

Rose, J.E., Brugge, J.F., Anderson, D.J. and Hind, J.E. (1967). Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey, *Journal of Neurophysiology* **30**(4): 769–793.

Rosenblith, W.A. and Stevens, K.N. (1953). On the DL for frequency, *Journal of the Acoustical Society of America* **25**(5): 980–985.

Rubinstein, J.T., Wilson, B.S., Finley, C.C. and Abbas, P.J. (1999). Pseudospontaneous activity: stochastic independence of auditory nerve fibers with electrical stimulation, *Hearing Research* **127**: 108–118.

Rudemo, M. (1972). Doubly stochastic Poisson processes and process control, *Advances in Applied Probability* **4**(2): 318–338.

Ruggero, M.A. and Rich, N.C. (1991). Furosemide alters organ of Corti mechanics: evidence for feedback of outer hair cells upon the basilar membrane, *Journal of Neuroscience* **11**(4): 1057–1067.

Sachs, M.B. and Abbas, P.J. (1974). Rate versus level functions for auditory nerve fibers in cats: tone burst stimuli, *Journal of the Acoustical Society of America* **56**(6): 1835–1847.

Sachs, M.B., Winslow, R.L. and Sokolowski, B.H.A. (1989). A computational model for rate-level functions from cat auditory-nerve fibers, *Hearing Research* **41**(1): 61–70.

Sanes, D.H. and Bao, S. (2009). Tuning up the developing auditory CNS, *Current Opinion in Neurobiology* **19**(2): 188–199.

Sanger, T.D. (2003). Neural population codes, *Current Opinion in Neurobiology* **13**(2): 238–249.

Schindler, R.A., Kessler, D.K. and Barker, M. A. (1995). Clarion patient performance: an update on the clinical trials, *Annals of Otology, Rhinology and Laryngology* **104**: 269–272.

Schroder, A.C., Viemeister, N.F. and Nelson, D.A. (1994). Intensity discrimination in normal-hearing and hearing-impaired listeners, *Journal of the Acoustical Society of America* **96**(5): 2683–2693.

Schuknecht, H.F. (1993). *Pathology of the Ear*, 2nd edn, Lea and Febiger, Philadelphia.

Schwarz, J.R., Reid, G. and Bostock, H. (1995). Action potentials and membrane currents in the human node of Ranvier, *Pflugers Archiv European Journal of Physiology* **430**(2): 283–292.

Sek, A. and Moore, B.C.J. (1995). Frequency discrimination as a function of frequency, measured in several ways, *Journal of the Acoustical Society of America* **97**(4): 2479–2486.

Shadlen, M.N. and Newsome, W.T. (1994). Noise, neural codes and cortical organization, *Current Opinion in Neurobiology* **4**(4): 569–579.

Shannon, R.V. (1983a). Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics, *Hearing Research* **11**(2): 157–189.

Shannon, R.V. (1983b). Multichannel electrical stimulation of the auditory nerve in man. II. Channel interaction, *Hearing Research* **12**(1): 1–16.

Shepherd, D. and Hautus, M.J. (2009). Evidence of stochastic resonance in an auditory discrimination task may reflect response bias, *Attention, Perception, and Psychophysics* **71**(8): 1931–1940.

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

Shofner, W.P. and Sachs, M.B. (1986). Representation of a low-frequency tone in the discharge rate of populations of auditory nerve fibers, *Hearing Research* **21**(1): 91–95.

Shower, E. G. and Biddulph, R. (1931). Differential pitch sensitivity of the ear, *Journal of the Acoustical Society of America* **3**(2A): 275–287.

Siebert, W.M. (1965). Some implications of the stochastic behavior of primary auditory neurons, *Kybernetik* **2**(5): 206–215.

Siebert, W.M. (1968). Stimulus transformation in the peripheral auditory system, *in* P. Kolers and M. Eden (eds), *Recognizing Patterns*, MIT Press, Cambridge, MA, pp. 104–133.

Siebert, W.M. (1970). Frequency discrimination in the auditory system: place or periodicity mechanisms?, *Proceedings of the IEEE* **58**(5): 723–730.

Sinnott, J.M., Brown, C.H. and Brown, F.E. (1992). Frequency and intensity discrimination in Mongolian gerbils, African monkeys and humans, *Hearing Research* **59**(2): 205–212.

Skinner, M.W., Holden, L.K., Whitford, L.A., Plant, K.L, Psarros, C. and Holden, T.A. (2002). Speech recognition with the Nucleus 24 SPEAK, ACE and CIS speech coding strategies in newly implanted adults, *Ear and Hearing* **23**: 207–223.

Smit, J.E., Hanekom, T. and Hanekom, J.J. (2009a). Modelled temperature-dependent excitability behaviour of a generalised human peripheral sensory nerve fibre, *Biological Cybernetics* **101**(2): 115–130.

Smit, J.E., Hanekom, T. and Hanekom, J.J. (2009b). Modelled temperature-dependent excitability behaviour of a single Ranvier node for a human peripheral sensory nerve fibre, *Biological Cybernetics* **100**(1): 49–58.

Snippe, H.P. (1996). Parameter extraction from population codes: a critical assessment, *Neural Computation* **8**(3): 511–529.

Snyder, D.L. and Miller, M.I. (1991). *Random Point Processes in Time and Space*, 2nd edn, Springer-Verlag, New York.

Srulovicz, P. and Goldstein, J.L. (1983). A central spectrum model: a synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum, *Journal of the Acoustical Society of America* **73**(4): 1266–1276.

Sutter, M.L. (2000). Shapes and level tolerances of frequency tuning curves in primary auditory cortex: quantitative measures and population codes, *Journal of Neurophysiology* **84**(2): 1012–1025.

Swanson, B., Van Baelen, E., Janssens, M., Goorevich, M., Nygard, T. and Van Herck, K. (2008). Cochlear implant signal processing ics, *Proceedings of the Custom Integrated Circuits Conference*, NXP, Leuven, Belgium, pp. 437–442.

Teich, M.C. and Lachs, G. (1979). A neural-counting model incorporating refractoriness and spread of excitation. I. Application to intensity discrimination, *Journal of the Acoustical Society of America* **66**(6): 1738–1749.

Vandali, A.E., Whitford, L.A., Plant, K.L. and Clark, G.M. (2000). Speech perception as a function of electrical stimulation rate: using the Nucleus 24 cochlear implant system, *Ear and Hearing* **21**(6): 608–624.

Verveen, A.A. (1961). *Fluctuation in excitability*, PhD thesis, Faculty of Mathematics and Natural Sciences, University of Amsterdam, The Netherlands.

Viemeister, N.F. (1972). Intensity discrimination of pulsed sinusoids: the effects of filtered noise, *Journal of the Acoustical Society of America* **51**(4): 1265–1269.

Viemeister, N.F. and Bacon, S.P. (1988). Intensity discrimination, increment detection, and magnitude estimation for 1-kHz tones, *Journal of the Acoustical Society of America* **84**(1): 172–178.

Viemeister, N.F. and Wakefield, G.H. (1991). Temporal integration and multiple looks, *Journal of the Acoustical Society of America* **90**(2): 858–865.

Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm, *IEEE Transactions on Information Theory* **13**(2): 260–269.

Vondrášek, M., Tichý, T. and Sovka, P. (2006). Virtual electrodes discrimination in Nucleus 24 cochlear implant, *International Conference on Applied Electronics 2006, AE*, Department of Circuit Theory, Faculty of Electrotechnical Engineering, Czech Technical University, Prague, pp. 229–232.

Wakefield, G.H. and Nelson, D.A. (1985). Extension of a temporal model of frequency discrimination: intensity effects in normal and hearing-impaired listeners, *Journal of the Acoustical Society of America* **77**(2): 613–619.

Wesselink, W.A., Holsheimer, J. and Boom, H.B.K. (1999). A model of the electrical behaviour of myelinated sensory nerve fibres based on human data, *Medical and Biological Engineering and Computing* **37**(2): 228–235.

Wever, E.G. (1949). *Theory of Hearing*, Wiley Publications in Psychology, Wiley, New York.

Wier, C.C., Jesteadt, W. and Green, D.M. (1977). Frequency discrimination as a function of frequency and sensation level, *Journal of the Acoustical Society of America* **61**(1): 178–184.

Wilson, B.S. and Dorman, M.F. (2008a). Cochlear implants: a remarkable past and a brilliant future, *Hearing Research* **242**: 3–21.

Wilson, B.S. and Dorman, M.F. (2008b). Cochlear implants: current designs and future possibilities, *Journal of Rehabilitation Research and Development* **45**(5): 695–730.

Wilson, B.S., Finley, C.C., Lawson, D.T., Wolford, R.D., Eddington, D.K. and Rabinowitz, W.M. (1991). Better speech recognition with cochlear implants, *Nature* **352**: 236–238.

Wilson, B.S., Lawson, D.T., Zerbi, M. and Finley, C.C. (1994). Recent developments with the CIS strategies, *in* I. Hochmair-Desoyer and E. Hochmair (eds), *Advances In Cochlear Implants, Proceedings Third International Cochlear Implant Conference*, Innsbruck, Austria, pp. 103–112.

Winter, I.M. and Palmer, A.R. (1991). Intensity coding in low-frequency auditory-nerve fibers of the guinea pig, *Journal of the Acoustical Society of America* **90**(4): 1958–1967.

Xu, Y. and Collins, L.M. (2004). Predicting the threshold of pulse-train electrical stimuli using stochastic auditory nerve model: the effects of stimulus noise, *IEEE Transactions on Biomedical Engineering* **51**(4): 590–603.

Xu, Y. and Collins, L.M. (2005). Predicting dynamic range and intensity discrimination for electrical pulse-train stimuli using a stochastic auditory nerve mode: the effects of stimulus noise, *IEEE Transactions on Biomedical Engineering* **52**(6): 1040–1049.

Xu, Y. and Collins, L.M. (2007). Predictions of psychophysical measurements for sinusoidal amplitude modulated (SAM) pulse-train stimuli from a stochastic model, *IEEE Transactions on Biomedical Engineering* **54**(8): 1389–1398.

Yates, G.K. (1990). Basilar membrane nonlinearity and its influence on auditory nerve rate-intensity functions, *Hearing Research* **50**: 145–162.

Yates, G.K., Winter, I.M. and Robertson, D. (1990). Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range, *Hearing Research* **45**(3): 203–219.

Young, E.D. and Barta, P.E. (1986). Rate responses of auditory nerve fibers to tones in noise near masked threshold, *Journal of the Acoustical Society of America* **79**(2): 426–442.

Young, E.D. and Calhoun, B.M. (2005). Nonlinear modeling of auditory-nerve rate responses to wideband stimuli, *Journal of Neurophysiology* **94**(6): 4441–4454.

Zeng, F.-G. (2002). Temporal pitch in electric hearing, *Hearing Research* **174**: 101–106.

Zeng, F.-G. and Shannon, R.V. (1999). Psychophysical laws revealed by electric hearing, *NeuroReport* **10**(9): 1931–1935.

Zeng, F.-G., Fu, Q.-J. and Morse, R. (2000). Human hearing enhanced by noise, *Brain Research* **869**: 251–255.

Zeng, F.-G., Galvin, J.J. III and Zhang, C. (1998). Encoding loudness by electric stimulation of the auditory nerve, *NeuroReport* **9**(8): 1845–1848.

Zwicker, E. (1970). Masking and psychological excitation as consequences of the ear's frequency analysis, *Frequency Analysis and Periodicity Detection in Hearing*, Sijthoff, Leiden.

Zwicker, E. and Fastl, H. (1999). *Psychoacoustics: Facts and Models*, Springer Series in Information Sciences, 2nd edn, Springer-Verlag.

Zwolan, T.A., Collins, L.M. and Wakefield, G.H. (1997). Electrode discrimination and speech recognition in postlingually deafened adult cochlear implant subjects, *Journal of the Acoustical Society of America* **102**(6): 3673–3685.

APPENDIX A

GENERAL CONCEPTS IN AUDITORY FUNCTION AND PSYCHOACOUSTICS

The mechanisms of hearing are very briefly described in section 1.1.1. This description does not provide anything more than an elementary picture of the processing that occurs in the auditory system. A few more specific concepts are therefore described in this appendix to provide the necessary background on some of the principles used in this study. These descriptions are not intended to be exhaustive, however, and should simply be used as a glossary for readers not familiar with the field of auditory neuroscience. References are given where appropriate, if further elaboration on the subject is required.

A.1 NEURAL CODING IN THE AUDITORY SYSTEM

The nerve impulses that travel from the cochlea to the CANS are necessarily ordered in some manner to facilitate the fine hearing ability that humans and animals possess. This order manifests itself in the form of a neural code.

A.1.1 Rate-intensity functions

RI functions are what characterise the behaviour of nerves. They provide a representation of a nerve's firing rate as a function of its input stimulus. In the case of cochlear neurons, this is input SPL (in dB SPL) when stimulated by acoustic means and current intensity (in dB *re* 1 μA) when stimulated by electrical means. The RI function is further characterised by its spontaneous rate, the rate at which the particular nerve fires when it is not stimulated at all; its maximum firing rate, the saturated firing rate; and the threshold. The threshold of cochlear neurons is usually defined in two ways: it can be defined as the point at which the firing rate is 10% above the spontaneous rate; alternatively, it can be defined as the rate at which the firing rate is 50% of the maximum firing rate minus the spontaneous rate. The choice of how to measure the threshold is arbitrary, but this study uses the latter definition for the sake of consistency with the model of electrical stimulation of Bruce *et al.* (1999a; 1999b; 1999c) and it is simply referred to as "the threshold" throughout this document. A typical rate-intensity function is illustrated in figure A.1.

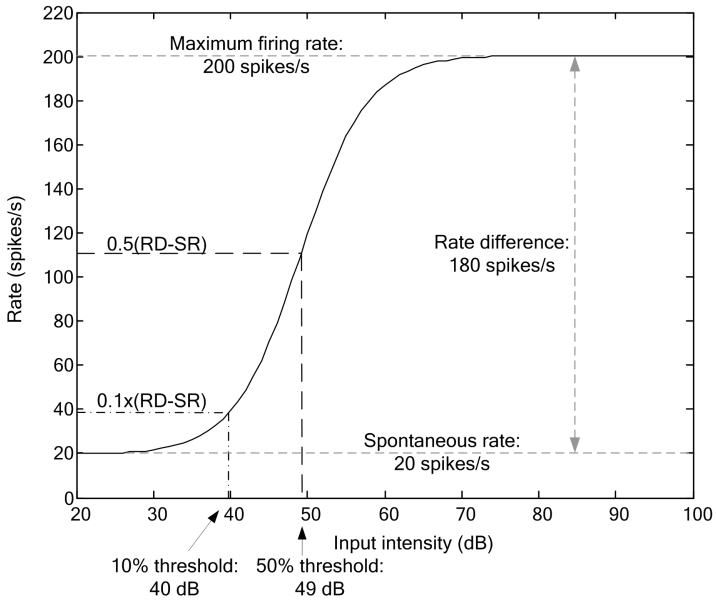

Figure A.1.
An example rate-intensity function with the maximum firing rate, spontaneous rate, rate difference, 10% threshold and 50% threshold indicated. The horizontal axis is given in dB.

A.1.2 Tuning curves

Every neuron is tuned to a specific parameter. A neuron will tend to fire at an average rate when a particular stimulus that the neuron is sensitive to is at a certain magnitude. This can be represented by means of a tuning curve, which shows at what rate a specific neuron will fire when stimulated by a certain magnitude of stimulus (Eliasmith and Anderson, 2003). In the context of cochlear neurons, the tuning parameter is the threshold defined as a function of frequency, an example of which is shown in figure A.2. The tuning curve has been the subject of many studies and continues to be used in research. Recently, this concept has been extended to cochlear implantees, where the tuning parameter is space (Nelson *et al.*, 2008).

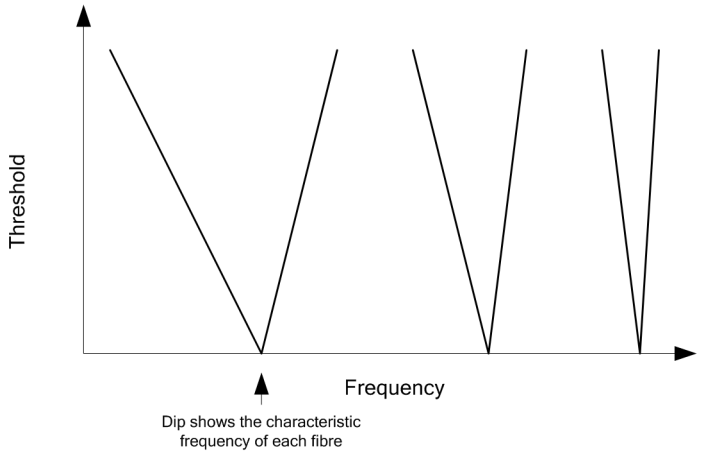

Figure A.2.
Example tuning curves of neurons in the cochlea, as illustrated by the firing threshold matched to increasing frequency. As frequency increases in the cochlea, the tuning sharpens up to a point, as demonstrated by Kiang (1965) using cat data. The CF of a fibre is determined by the point at which its threshold is the lowest.

A.1.3 Tonotopicity of the cochlea

The frequency tuning of the auditory neurons is central to the idea of the tonotopic arrangement of the cochlea. Tonotopicity refers to the way in which the cochlea is tuned to different frequencies at different places along the length of the cochlea, i.e. the cochlea is differentially sensitive to frequency depending on the location in the cochlea. This concept was originally

quantified by Greenwood (1961) and revised in Greenwood (1990). This implies that it is possible to determine the CF of the nerves at a specific point on the cochlea, or conversely, the approximate position on the cochlea of a nerve can be determined by knowing its CF.

A.1.4 The critical band and equivalent rectangular bandwidth

It has been posited that the peripheral auditory system acts as a bank of overlapping band-pass filters. Each section of the cochlea responds differently to certain frequencies, and each of these filters is considered to be an auditory filter. The name hints at the underlying concept: the critical band is the bandwidth in which signals outside of the passband of the auditory filter have no influence on the hearing process. This implies that nerve firing outside of the critical band will not influence the perception of the signal that falls within the critical band. This tenet does not hold true for all circumstances (especially in the case of complex tones), but it is useful in the modelling of the filtering process carried out in the peripheral system when dealing with pure tones, or pure tones in noise.

The size and shape of the filters have been estimated in a couple of ways, namely by using psychophysical tuning curves (not to be confused with the concept of neural tuning curves) or notched noise measurements, both of which are described and explained in Moore (2003). The ERB is derived from the latter measurement, and is similar to the critical band, but obtained in a different way. It represents the bandwidth of an ideal rectangular filter that is roughly equivalent to the auditory filter bandwidth measured using the notched noise method. ERB actually varies according to frequency. This can be approximated by an equation proposed by Glasberg and Moore (1990), shown in equation A.1:

$$\mathrm{ERB_N} = 24.7 \cdot (4.37 f + 1) \,, \tag{A.1}$$

where $\mathrm{ERB_N}$ is given as a frequency range in Hz, and f is given in kHz.

A.1.5 Place coding

There are a number of ways in which neural encoding and decoding can be performed (Rieke *et al.*, 1997; Eliasmith and Anderson, 2003), and one of these methods, which applies specifically to the auditory system, is place coding. There is not a great degree of certainty as to

exactly how the auditory system encodes frequency information, but place coding is one theory that has been used to explain it. The basic idea is that the fundamental mechanism of hearing is related to the pattern of excitation that is elicited along the cochlea as a result of stimulation. This is generally accepted as the main method of coding in the auditory system for frequencies above 5 kHz (Moore, 2003). It has also been used as one of the ways in which intensity discrimination can be explained, and can in fact account for the near miss to Weber's law (see section 2.5).

A.1.6 Temporal coding

Temporal coding is another coding method that has been used to explain certain percepts, such as intensity discrimination and frequency discrimination. The concept of temporal coding suggests that the brain uses the temporal properties of the spike train patterns to determine the input stimulus. Temporal coding can rely on different temporal properties of the spike trains, namely the phase-locking properties of the neurons and the time intervals between individual spikes. This is generally accepted as the main method of coding in the auditory system for frequencies below 5 kHz (Moore, 2003).

The reason that temporal coding is thought to no longer play a role at higher frequencies is attributed to the roll-off to phase locking to the stimulus by the neural firings. Johnson (1980) quantified this phenomenon and found that the phase-locking properties of auditory neurons gradually become weaker as frequency increases, starting at about 2 kHz.

A.2 PSYCHOACOUSTICS

Psychophysics is the study of how stimuli are perceived by the brain. Psychoacoustics is a sub-field of psychophysics dealing specifically with the auditory system, and there are a large number of publications devoted to specific aspects of this well-established field (Moore and Glasberg, 1986; Zwicker and Fastl, 1999; Bregman, 2001; Moore, 2003). This study focuses mainly on the psychoacoustic measures of frequency discrimination and intensity discrimination.

A.2.1 Perceptual measures of frequency

There are a number of perceptual measures that relate to frequency. These include pure-tone frequency discrimination, frequency discrimination of change, frequency modulation discrimination, frequency selectivity, frequency discrimination in noise and frequency masking, among others. This study investigates pure-tone frequency discrimination, a measure of a subject's ability to detect changes in frequency over time. Frequency discrimination is usually measured by exposing a subject to two consecutive frequencies, one being a reference frequency, and the other being a frequency different to the first. The smallest detectable change in these two frequencies (Δf) by the subject is the frequency discrimination threshold, also known as the frequency difference limen. This can also be measured in terms of Weber fractions, discussed below in section A.2.4.

Frequency selectivity, a concept that could be confused with frequency discrimination, is the ability to discern specific frequencies from complex sounds, i.e. select specific frequencies from a sound consisting of multiple tones simultaneously. Frequency modulation discrimination is also only slightly different from frequency discrimination. In the case of this percept, the subject would be required to detect which of two consecutive frequencies is modulated by another frequency, in a similar fashion to frequency discrimination. It is measured by frequency modulation difference limens. Frequency discrimination of change involves presenting the listener with two successive pairs of frequencies. In one set, the two frequencies will be the same; in the other set, the two will be different. The listener has to choose which of the two pairs differed in frequency. These percepts do not necessarily use the same perceptual mechanisms as frequency discrimination, and are not the subject of this study, but are noteworthy for the distinctness in the mechanisms used in comparison to frequency discrimination.

A.2.2 Perceptual measures of intensity and loudness

There are fewer perceptual measures related to intensity than to frequency. Intensity discrimination is the main measure, and loudness perception is another. Intensity discrimination refers to the measure of smallest detectable change in intensity over time (ΔI, the intensity difference limen) which can also be represented by Weber fractions (see section A.2.4). Intensity discrimination is measured in a similar way to frequency discrimination. However, it

is also possible to measure intensity discrimination using other techniques, such as modulation detection or increment detection.

Loudness perception, although not the subject of this study, is important as there is a nonlinear relationship between loudness and intensity. In NH listeners the relationship is approximated by a power function. Loudness also varies as a function of frequency. The units of loudness are called phons.

A.2.3 Dynamic range

Dynamic range refers to the range of comfortable, audible sound levels that a listener can perceive. The lower end of dynamic range is represented by the threshold of audibility, the lowest level of loudness that can be perceived by a listener. The upper end is the maximum comfortable loudness level, which is marked by the point at which the listener begins to experience discomfort from sound at that level. This is not a linear scale for any NH listener, as is implied by the nonlinear relationship between sound intensity and loudness. Cochlear implantees have a much smaller dynamic range than NH listeners. The signal processors of implants are designed to account for this difference by compressing a large intensity range into the small dynamic range that implantees are capable of hearing.

A.2.4 Weber's law and Weber fractions

Ernst Heinrich Weber (1795–1878) studied many psychophysical properties including intensity discrimination. Weber's law, one of the fundamental principles in psychophysics, states that the smallest detectable change in a stimulus is proportional to the magnitude of the stimulus; in the case of intensity discrimination, this could be represented as

$$C \propto \Delta I / I \ ,$$

where C is a constant and ΔI is the smallest detectable change in intensity at intensity I, also called the intensity difference limen. The term $\Delta I / I$ is also known as the Weber fraction. Although this law was widely regarded as true for many years, it has been shown by numerous studies that it does not hold for pure tones for all intensities in hearing (Riesz, 1928; Viemeister, 1972; Jesteadt *et al.*, 1977; Florentine, 1983; Florentine *et al.*, 1987; Oxenham and

Moore, 1995). Nonetheless, Weber's law is still used as a basis for comparison in psychophysics, and the Weber fraction is used to represent psychophysical measures, including frequency discrimation, where the Weber fraction is $\Delta f / f$. In this case Δf is the frequency difference limen at frequency f.

APPENDIX B

RATE-INTENSITY DATA

The data from figure 3.6 in chapter 3, on which curve fitting was performed, was largely sourced from Sachs and Abbas (1974), but Evans (1978) was also used. This data was read off of the graphs provided in those papers and as such is merely a rough estimate that was intended only as an approximation of the behaviour of RI functions. Owing to the large variability and inconsistency that is observed in RI functions, this rough approximation is adequate and serves its purpose. However, it is proposed for future work that more studies be performed on the frequency variability of RI functions.

The data that was read off of the figures is presented here in its entirety although not all of the data was used. Only data for a percentage change in frequency of less than 50% was considered, as frequency changes greater than this were never needed in this model. In table B.1, the percentage difference in frequency was determined by equation B.1, and the percentage difference in rate was determined by equation B.2.

$$\Delta f = \left| 100 \cdot \frac{CF - F}{CF} \right| \tag{B.1}$$

$$\Delta r = \left| 100 \cdot \frac{MR_{CF} - MR_F}{MR_{CF} - SR} \right| \tag{B.2}$$

Table B.1.
Data read off from RI functions in literature along with important nerve fibre parameters. SA refers to data from Sachs and Abbas (1974) and EV refers to data from Evans (1978). F refers to the stimulus frequency, MR_{CF} refers to the maximum firing rate at the CF and MR_F refers to the maximum firing rate at the stimulus frequency. Percentage changes in frequency and rate are also displayed.

Source/ Unit	CF **(kHz)**	SR **(spikes/s)**	MR_{CF} **(spikes/s)**	F **(kHz)**	MR_F **(spikes/s)**	Δf **(%)**	Δr **(%)**
SA 2-12	0.72	30	167	0.72	167	0	0
				1.1	150	53	12
				1.3	130	81	27
				1.4	105	94	45
				1.5	80	108	64
SA 11-01	1.3	21	175	1.3	175	0	0
				1.8	150	38	16
				2.0	110	54	42
EV Page 22	2.2	50	148	2.2	148	0	0
				2.4	148	9	0
				2.6	126	18	22
				2.8	112	27	37
				3.0	88	36	61
SA 2-02	2.3	1	200	2.3	200	0	0
				2.6	167	13	17
				2.8	152	22	24
				3.0	133	30	34
SA 10-06	11.3	30	275	11.3	275	0	0
				13.5	267	19	3
				14.0	233	24	17
				14.5	175	28	41
				15.0	122	33	63
SA 9-01	14.5	7	244	14.5	244	0	0
				16.0	207	10	16
				16.5	160	14	35

The data for the breakpoint parameter (A_3), was sourced in a similar way and is shown in table B.2. The abbreviations and references apply as they did in table B.2.

Table B.2.
Approximate breakpoint parameter data read from RI functions in literature, along with important nerve fibre parameters.

Source/ unit	*CF* **(kHz)**	**Threshold at** *CF* **(dB SPL)**	**Approximate** A_3 **(dB SPL)**
SA/ 2.12	0.72	27	73
SA/ 11.01	1.3	47	76
EV/ page 22	2.2	38	52
SA/ 2.02	2.3	45	74
SA/ 5.15	4.0	63	75
SA/ 4.06	6.75	60	100
SA/ 5.13	7.5	42	70
SA/ 10.06	11.3	30	70
SA/ 12.07	11.7	59	90
SA/ 8.01	13.5	72	83
SA/ 9.01	14.5	26	70

These were plotted on a figure and an average was calculated, as shown in figure B.1.

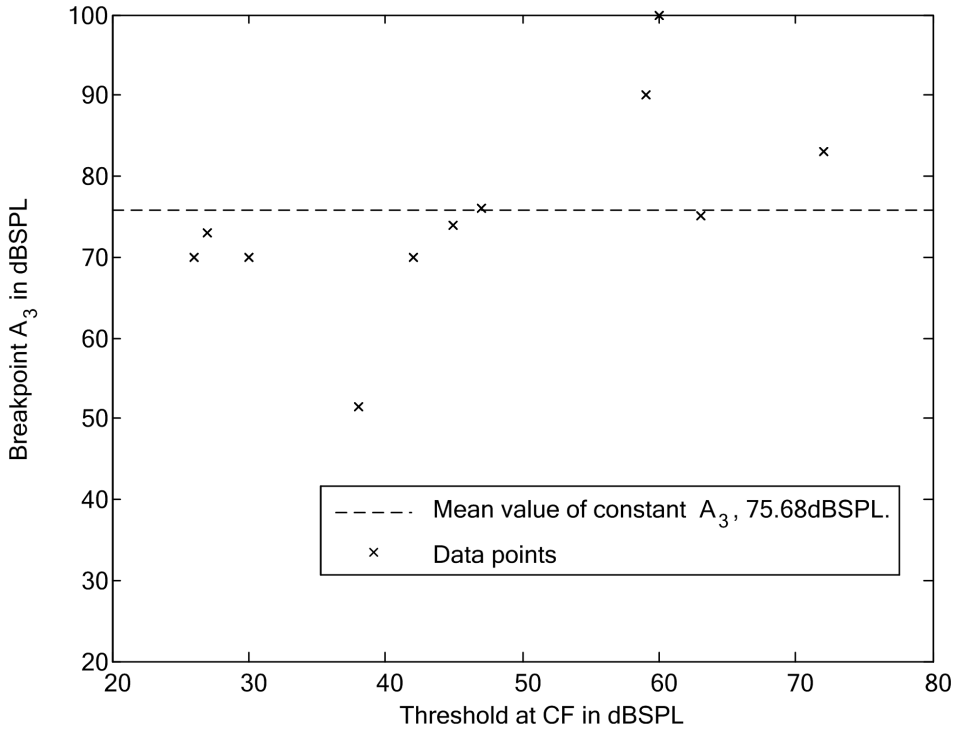

Figure B.1.
Scatter diagram of the threshold at CF against the approximate A_3 value. The dotted line shows the mean value plotted across the figure, 76 dB SPL.

APPENDIX C

MARKOV PROCESSES

This appendix provides background information on Markov Processes. The details are not intended to be extensive, but rather to provide readers without a background in the subject with some basic information necessary for the understanding of the mathematics used.

According to Papoulis (1984), a Markov process is a stochastic process in which a past event does not affect any future events if the present state is known. A Markov chain defined in discrete time is a Markov process (η_n) that has a finite number of states (in space K). It is specified by its state probabilities, shown in equation C.1, and its transition probabilities, shown in equation C.2.

$$p_k[t] = P(\eta_t = k) \tag{C.1}$$

$$\pi_{ij}[t_1, t_2] = P(\eta_{t_2} = k_j | \eta_{t_1} = k_i) \tag{C.2}$$

The transition probability matrix is given by

$$\Pi = \begin{bmatrix} \pi_{11} & \pi_{12} & \dots & \pi_{1T} \\ \pi_{21} & \pi_{22} & \dots & \pi_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{T1} & \pi_{T2} & \dots & \pi_{T} \end{bmatrix}, \tag{C.3}$$

where

$$\sum_j \pi_{ij} = 1. \tag{C.4}$$

This matrix (equation C.3) provides a quantitative description of the probabilities of one state transitioning to another. The application of these equations and matrices can easily be visualised by the state diagram shown in figure C.1. For more information and detail on the various properties of Markov processes, please refer to Assefi (1979) or Papoulis (1984).

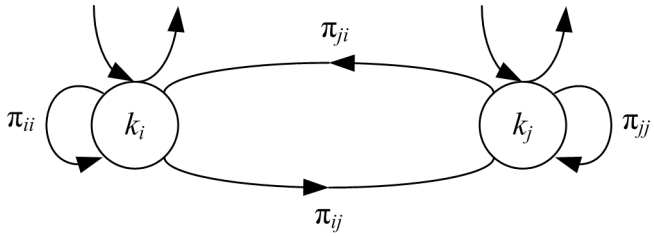

Figure C.1.
State diagram illustrating a discrete-time Markov process. Two different states and the associated probabilities are shown. Adapted from Papoulis (1984).