UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Models and psychophysics of acoustic and electric hearing

by

Johannes Jurgens Hanekom

Submitted as partial fulfilment of the requirements for the degree

Philosophiae Doctor

in the

Faculty of Engineering, Built Environment and Information Technology

University of Pretoria, Pretoria

November 2001

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Models and psychophysics of acoustic and electric hearing

by

Johannes Jurgens Hanekom

Advisor : Prof Johann J Krüger

Department : Electrical, Electronic and Computer Engineering

Degree : Philosophiae Doctor

KEY WORDS

Cochlear implants, electrical stimulation, modelling, phase-lock coding, frequency discrimination, gap detection, channel interaction, central auditory nervous system

ABSTRACT

Especially important in developing improved cochlear implants is to develop a deeper understanding of the processing of sound in the central auditory nervous system, for both acoustic and electrical stimulation of the auditory system. This thesis contributes to this objective through cochlear implant psychoacoustic research and modelling of auditory system sound processing.

The primary hypothesis of the thesis was that the same underlying mechanisms are responsible for sound perception in both electric and acoustic hearing. Thus, if appropriate models are created for normal acoustic hearing, they should be able to predict psychoacoustic data from electric hearing when the model input is changed from acoustic to electrical stimulation. A second hypothesis was that electrode interaction could be measured by gap detection and that predictions of current spread in the cochlea could be obtained from gap detection data.

Measured gap detection thresholds in three cochlear implant users were a function of the physical separation of electrode pairs used for the two stimuli that bound the gap, resulting in

a U-shaped "tuning curve" for this across-channel condition. Models of gap detection in acoustic and electric hearing were created to explain these U-shaped curves. A technique was developed to obtain estimates of cochlear current spread from gap detection data. Predictions of electrode discrimination were obtained from the current spread estimates, and these were compared to data measured in cochlear implant users.

The model for acoustic hearing could predict the U-shaped curves found in acoustic hearing, and when the input spike train statistics were adapted appropriately, the same model could also predict gap detection data for electric hearing. Predictions of current spread exhibited current peaks close to the electrodes and had length constants between 0.5 mm and 3 mm, similar to measured data quoted in literature. Predictions of electrode discrimination correlated well with measured data in one subject, but not in two others.

The primary conclusion from the modelling results is that if the mechanisms of central auditory nervous system signal processing of acoustic stimulation are understood, these same mechanisms may be applied to understand the signal processing in auditory electrical stimulation and to predict psychoacoustic data for electrical stimulation. A second conclusion is that spatial mechanisms, as opposed to temporal mechanisms, may determine gap detection thresholds in the across-channel condition. This is important in cochlear electrical stimulation, where spike trains are strongly phase-locked to the stimulus and temporal mechanisms cannot predict gap detection thresholds. A third conclusion is that gap detection can be used to measure channel interaction and to predict current distributions in the cochlea, although there is still uncertainty about the accuracy of these predictions. However, the gap detection data and predictions for current distributions indicate that electrodes are not discriminable when they are closer than 1.5 mm. The implication of these last two conclusions taken together is that research should focus on obtaining better spatial resolution in cochlear implants.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Modelle en psigofisika van akoestiese en elektriese gehoor

deur

Johannes Jurgens Hanekom

Promotor	:	Prof Johann J Krüger
Departement	:	Elektriese, Elektroniese en Rekenaar-Ingenieurswese
Graad	:	Philosophiae Doctor

SLEUTELWOORDE

Kogleêre inplantings, elektriese stimulasie, modellering, fase-sluit kodering, frekwensiediskriminasie, gapingsdeteksie, kanaalinteraksie, sentrale ouditiewe senuweestelsel

OPSOMMING

Tydens die ontwikkeling van beter kogleêre inplantings is dit besonder belangrik om 'n dieper begrip van die verwerking van klank in die sentrale ouditiewe senuweestelsel te ontwikkel, beide vir akoestiese en elektriese stimulasie van die ouditiewe stelsel. Hierdie proefskrif dra by tot hierdie oogmerk deur psigoakoestiese navorsing en modellering van klankverwerking deur die ouditiewe stelsel.

Die primêre hipotese is dat dieselfde onderliggende meganismes verantwoordelik is vir klankpersepsie in beide elektriese en akoestiese gehoor. Dus, indien toepaslike modelle vir normale akoestiese gehoor geskep word, behoort hulle psigoakoestiese data van elektriese gehoor te kan voorspel indien die modelinset verander word vanaf akoestiese na elektriese stimulasie. 'n Tweede hipotese is dat elektrode-interaksie gemeet kan word met gapingsdeteksie en dat voorspellings van die stroomverspreiding in die koglea verkry kan word uit gapingsdeteksiedata.

Gemete gapingsdeteksiedrempels in drie gebruikers van kogleêre inplantings was 'n funksie van die fisiese afstand tussen elektrodepare wat gebruik is vir die twee stimuli aan weerskante van die gaping en het gelei tot 'n U-vormige "stemkromme" vir hierdie inter-kanaal geval. Modelle van gapingsdeteksie in akoestiese en elektriese gehoor is geskep om hierdie U-vormige krommes te verduidelik. 'n Tegniek is ontwikkel om skattings van stroomverspreiding in die koglea te verkry vanuit gapingsdeteksiedata. Voorspellings van elektrodediskriminasie is verkry uit die stroomverspreidingskattings en is met gemete data van kogleêre inplantinggebruikers vergelyk.

Die model vir akoestiese gehoor kan die U-vormige gapingsdeteksiekrommes van akoestiese gehoor voorspel, en met gepaste aanpassing van die statistieke van die senuwee-impulsreeks op die modelinset kan dieselfde model ook gapingsdeteksiedata vir elektriese gehoor voorspel. Voorspellings vir stroomverspreidings toon pieke naby die elektrodes en het lengtekonstantes van 0.5 mm tot 3 mm, soortgelyk aan data waarna in die literatuur verwys word. Voorspellings van elektrodediskriminasie korreleer goed met gemete data in een proefpersoon, maar dieselfde geld nie in twee ander proefpersone nie.

Die primêre gevolgtrekking uit die resultate van die modellering is dat indien die meganismes van seinverwerking in die sentrale ouditiewe senuweestelsel verstaan word, kan hierdie meganismes toegepas word om seinprosessering in die elektries-gestimuleerde gehoorstelsel te verstaan en om psigoakoestiese data vir elektriese gehoor te voorspel. 'n Tweede gevolgtrekking is dat ruimtelike meganismes eerder as temporale meganismes verantwoordelik kan wees vir gapingsdeteksiedrempels in die inter-kanaal geval. Dit is belangrik in kogleêre elektriese stimulasie waar senuwee-impulsreekse sterk fasegesluit is aan die stimulus en waar temporale meganismes nie gapingsdeteksiedrempels kan voorspel nie. 'n Derde gevolgtrekking is dat gapingsdeteksie gebruik kan word om kanaalinteraksie te meet en om stroomverspreidings in die koglea te voorspel, alhoewel onsekerheid bestaan oor die akkuraatheid van hierdie voorspellings. Nietemin, die gapingsdeteksiedata en voorspellings vir stroomverspreiding wys dat elektrodes nie diskrimineerbaar is as hulle nader as 1.5 mm aan mekaar is nie. Die implikasie van die laaste twee gevolgtrekkings is dat dit nodig is dat navorsing fokus op die verkryging van beter ruimtelike resolusie in kogleêre inplantings.

CONTENTS

LIST OF ABBREVIATIONS

2IFC	two interval forced choice
Δf	just noticeable difference in frequency
ALSR	average localized synchronized rate
ANOVA	analysis of variance
AR	apical reference
BP	bipolar
CF	characteristic frequency
CIS	continuous interleaved sampling
CN	cochlear nucleus
CRLB	Cramer-Rao lower bound
dc	direct current
DCN	dorsal cochlear nucleus
ISI	inter-spike interval
jnd	just noticeable difference
jndf	just noticeable difference in frequency
pdf	probability density function
RS	relative spread
SL	sensation level
SNR	signal-to-noise ratio
SPEAK	spectral peak
SPL	sound pressure level
SR	spontaneous rate

Chapter 1
INTRODUCTION

Cochlear implants have been developed to stimulate the auditory systems of deaf individuals to create a hearing sensation (Clark, 1993;Clark, 1996;Clark, Tong, and Patrick 1990). Cochlear implants are intended to elicit firing patterns on the auditory nerve that are faithful replicas of firing patterns found in normal hearing. Many thousands of deaf people worldwide use cochlear implants with varying degrees of success (Kou, Shipp, and Nedzelski, 1994).

1 ISSUES IN COCHLEAR IMPLANT RESEARCH

A primary problem with cochlear implants is the large variability in speech recognition performance across users (Clark, 1996). This is also reflected in inter-subject performance variability in psychoacoustic experiments. Most implant listeners cannot appreciate music (Fujita and Ito, 1999) and many users cannot use a telephone or struggle to follow speech in noisy environments (Hirsch, 1993). Auditory sensations vary as a result of complex interactions between several patient and device related factors, among others electrode distance from nerve fibres, surviving nerve fibre distribution (Zimmerman, Burgess, and Nadol, 1995), and current distribution in the cochlea due to non-homogeneous cochlear impedance (Ifukube and White, 1987).

Initially, cochlear implant research was driven by clinical objectives. Safety and reliability issues, like biocompatibility of electrodes and implantable electronics, needed to be solved. Design and development of the external speech processors, the implantable stimulator, electrode arrays and communication protocols between the stimulator and the speech processor required attention. Stimulation parameters (currents and waveforms) and stimulation strategies

needed to be developed.

Medical-ethical aspects, including animal experiments to prove safety and functionality, proving that deaf people would benefit from cochlear implants and determining which segment of the deaf population (including children) would benefit needed to be addressed and regulatory body approval had to be obtained.

After this initial phase, research focus shifted to more fundamental issues. Many of the issues identified in chapter 25 of (Miller and Spelman, 1990) are still topics of active research. Progress has been made in finite element modelling to predict current flow through excitable tissue in the cochlea (Frijns, de Snoo, and Schoonhoven, 1995; Hanekom, 2001). New and more selective electrode designs have been introduced. Methods to limit channel interaction and current spread is a topic of current research (Kral et al., 1998). Techniques are being researched to determine channel interaction in individual cochlear implant users (Fu, 1997; Hanekom and Shannon, 1998; Shannon, 1985), to enable optimum setting of stimulation parameters. New encoding strategies have been developed (Loizou, 1999), based on improving either spectral or temporal presentation of sound, or both.

Especially important in developing improved cochlear implants is to develop a deeper understanding of the processing of complex speech signals in the central auditory nervous system. This needs to be achieved through neurophysiological animal experiments, psychoacoustic experiments with cochlear implant users, and the development of models of the neurophysiology of the coding and processing of information in the central auditory nervous system in both acoustic and electric hearing. Relating psychoacoustic results to the underlying neurophysiology is essential. Appropriate models can greatly assist in this effort.

2 MODEL BASED RESEARCH

This thesis promotes a model-based research approach. Therefore, it is appropriate to give a more detailed description of the process and value of modelling in cochlear implant research, before the objectives of this thesis are described. The objective of this section is to place the models created in this thesis into context.

Engineers and scientists frequently create mathematical models to represent processes and predict the effects of changes on a system. A model is a mathematical description of a natural phenomenon, in this case a mathematical description of the functioning of the central auditory nervous system under acoustic or electrical stimulation.

Neurophysiological research, psychoacoustic research and modelling research are interrelated and cannot exist in isolation (figure 1.1). Without models to aid interpretation, psychoacoustics research may follow an empirical or trial and error approach, without a clear route to follow to improve cochlear implant designs.

2.1 Types of models

The researcher creates a system of mathematical equations that describes the behaviour of the normal or electrically stimulated auditory system. Modelling equations may be solved with analytical methods (e.g. Siebert, 1970), or with numerical methods (e.g. Finley, Wilson, and White, 1990) or a combination of both (e.g. Bruce et al., 1999b). Numerical solutions can be obtained with computer simulations. It is a requirement that any type of model produce numerical predictions.

Furthermore, models can vary from biology-faithful models to black box models. Biology-faithful models incorporate as much as possible of the existing knowledge about the anatomy and physiology. These models may, for example, represent the function of a specific system in the brain by modelling a large number of nerve fibres and combining these nerve fibres into a

larger structure (Bower, 1990). More specifically for the current application, these models may describe each nerve fibre in the model by a full-blown Hodgkin-Huxley model (Rattay and Motz, 1986). The Hodgkin-Huxley model is one of a number of biology-faithful nerve fibre models and uses a number of differential equations to describe the functioning of a single nerve cell. Biology-faithful models can be accurate, but may require substantial computer time for simulation.

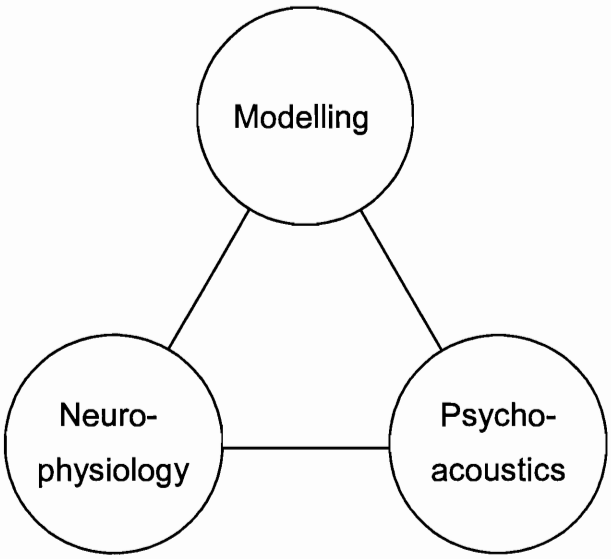

Figure 1.1.
Relationship between neurophysiological research, psychoacoustic research and modelling research.

Black box models are phenomenological models that simply describe input-output data with a mathematical equation without consideration for the biology. Some nerve fibre models model the nerve fibre as a simple generator of neural spikes (e.g. Gabbiani and Koch, 1996). In the simplest form black box models simply fit a curve to the data and assumes no knowledge of the functioning of the system that is modelled. Black box models tend to be very limited in terms of the data sets that can be predicted.

A whole continuum of models that incorporate existing knowledge of the biology to a greater or lesser extent exists between biology-faithful models and black box models. Some models deliberately ignore existing knowledge (thus are more idealized than biology-faithful models) with the objective to exclude extraneous factors (e.g. variations between people or noise in signals) to uncover the core calculations performed by the system that is investigated. Therefore, models sometimes deliberately simplify the real situation to get rid of "noise" that hides the true function that the system under consideration performs. Examples of this approach will be evident in this thesis.

Models are characterized by the model structure and the parameters. Modelling is frequently an iterative process. Based on the researcher's understanding of the processing performed in the system that is modelled, the researcher determines a model structure. Parameters for the model must subsequently be measured and can sometimes be found in the literature. Estimating model parameters is frequently necessary and sometimes it is adequate if the estimate is only roughly correct. With the structure and parameters known, computer simulations can be performed with the model and numerical predictions can be obtained.

If the model is well-behaved, i.e. the model can reproduce measured data accurately (it can predict magnitudes and trends), the choice of structure may be presumed to be appropriate. Sometimes trends are predicted correctly, but the magnitudes of the measured and predicted data do not correspond. This may be because the structure of the model is correct, but parameters have been guessed or measured incorrectly. Or it may happen that the model cannot predict the measured data at all, in which case the model structure may be incorrect.

2.2 The value of models

An important question that should be answered about each model is: can the model explain measured data outside the data set for which the model was created? If not, the model probably requires further development. The wider the data set that a model can explain, the more confident the researcher may be that the functionality of the underlying system is

understood and that the model structure is correct. For example, if a model were created to predict frequency discrimination of pure tones, but it can also predict discrimination of complex tones, it gives confidence in the validity of the model. As models are developed, understanding of the system that is modelled grows.

Models cannot predict the effect of stimulation or the behaviour of the auditory system perfectly, since they represent a simplification of the real system. However, an exact representation of reality is not necessarily required, or available, to describe the functioning of a system. The cochlear implant is a good example of this, because development of and improvements to implants have been successful, based on, among others, modelling studies, although the functioning of the auditory system is not fully understood (Clark, 1996). If a model can predict the most important characteristics of a system accurately enough, the model may be useful.

Finally, models may be regarded as part of the measuring tools in psychoacoustic and neurophysiological research. Models enable the researcher to test some concepts in computer simulations rather than performing experiments directly on humans or animals. Models summarize knowledge and show gaps in knowledge, but perhaps most important, they build understanding. This will be demonstrated in the various models discussed in this thesis. Many conclusions of this thesis (see the ends of the chapters) are based on discoveries made during the process of deriving numerical predictions that mimic measured data.

3 OBJECTIVES OF THIS THESIS

The absence of comprehensive, anatomically and physiologically faithful models to predict the effect of electrical stimulation on the auditory system is a primary deficiency in cochlear implant research. The creation of new, more comprehensive models, to incorporate a broader base of data is crucial for the further development of cochlear implants. Specifically, many gaps exists in our knowledge of the relationship between electrical stimuli and the perceived sound. It is

unclear why certain stimuli elicit certain perceived sounds (perception is measured psychoacoustically).

The primary objective of this thesis is to build understanding of the functioning of the peripheral and central auditory system in acoustic and electric hearing, through measurement and modelling. The study of auditory electrical stimulation provides a window on the auditory system that provides new perspectives on how the auditory system functions.

4 HYPOTHESES OF THIS THESIS

Four hypotheses are investigated in this thesis, as described here. Although not all of these can be proven conclusively, they are proven qualitatively by examples.

4.1 Hypothesis 1

Psychoacoustic data for electric hearing can be predicted by models for acoustic hearing

A lack of an integrated interpretation of different psychoacoustic results exists in literature. Throughout, the models in this thesis aim to reconcile psychoacoustic and neurophysiological data. An important assumption in this thesis is that the same underlying processes (in the central auditory nervous system) are responsible for the perception of electrical and normal acoustic stimulation. The primary hypothesis of this thesis is that appropriate models for normal acoustic hearing should be able to predict psychoacoustic data from electric hearing when the model input is changed from acoustic to electrical stimulation (figure 1.2). Chapters 3 and 4 (that describe models of acoustic and electric gap detection) and 6 and 7 (that describe models of acoustic and electric frequency discrimination) investigate this hypothesis.

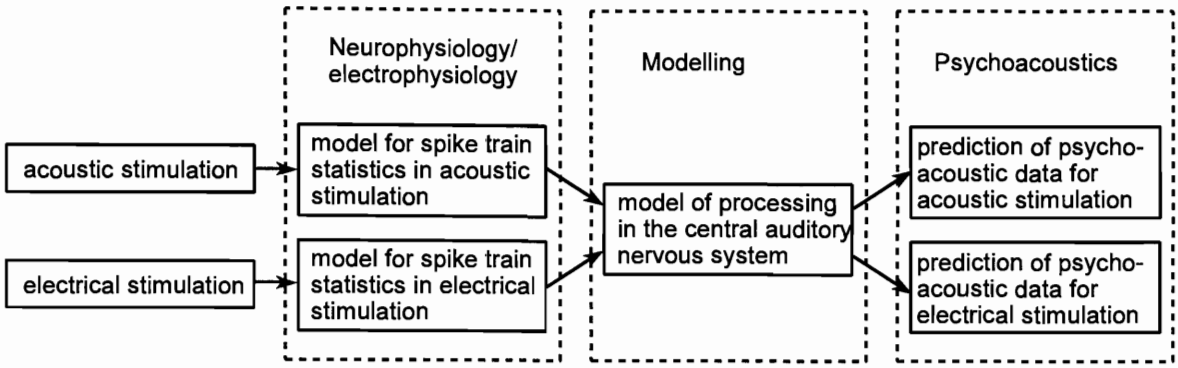

Figure 1.2.

The same underlying model of the central auditory nervous system is used to predict psychoacoustic data for both acoustic and electric hearing.

4.2 Hypothesis 2

The auditory system employs an internal model to interpret sounds

Classical detection and estimation theory as applied to auditory research asserts that the auditory system has no prior knowledge of the type of signals (speech, music, environmental sounds) that it receives. But evidence exists that the brain does use its prior knowledge to aid it in estimating input signals. A secondary hypothesis of this thesis is that the auditory system employs an internal model or an analysis-by-synthesis mechanism to estimate sounds impinging on the ear. This hypothesis is investigated in chapters 5 and 6.

4.3 Hypothesis 3

Temporal and spatial mechanisms can be explained by the same underlying neurophysiology

Mechanisms that neurophysiologists and psychophysicists interpret as "temporal mechanisms", and "frequency" or "spatial mechanisms", are based on and should therefore be explained by the same underlying neuroanatomy and neurophysiology. This hypothesis is investigated in this thesis. This is important in cochlear implants, where a trade-off between temporal and spectral

resolution exists. If fewer electrodes are stimulated, higher stimulation rates may be achieved. Thus it is necessary to know what information transmitted to the electrically stimulated cochlear nerve is perceptually significant. Strategies used in current cochlear implant systems reflect different approaches. In the Spectral Peak (SPEAK) strategy (Skinner et al., 1994; Loizou, 1999), which relies on the existence of a rate-place mechanism, spectral peaks are extracted and presented to electrodes that are arranged tonotopically. In contrast, the Continuous Interleaved Sampling (CIS) strategy (Wilson et al., 1991; Loizou, 1999) assumes that it is important to conserve temporal waveform information and therefore employs high pulse-rate stimulation.

Chapters 3 and 4 employ a spatial model to predict psychoacoustic data for temporal gap detection. Chapters 5 and 6 employ spatial and temporal models respectively to obtain predictions for frequency discrimination.

4.4 Hypothesis 4

Electrode interaction in cochlear implants can be measured with gap detection

Multiple-electrode stimulation is preferred in cochlear implants, because it is generally accepted that the tonotopic organization found along the length of the cochlea in the healthy auditory system is retained to some degree for electrical hearing. Many research studies, including earlier work by Eddington (1980) and more recent studies by Nelson et al. (1995) and Donaldson and Nelson (2000), have shown that electrodes stimulating the more basal areas of the cochlea result in higher perceived pitches (or sharper tonal quality) and stimulation closer to the apex results in lower perceived pitches.

Assuming that tonotopic organization in electrical hearing is retained by multiple electrodes selectively stimulating discrete neural populations would be natural. However, the assumption that discrete neural populations can be excited is not always true. When electrodes are closely spaced, considerable overlap occurs in the neural populations excited by the stimulation current. This electrode interaction problem was addressed by (among others) Townshend and

White (1987). This is the result of spread of electrical current in the biological medium of the cochlea.

The implication of electrode interaction is that, if two electrodes stimulate the same neural population or overlapping neural populations, sound sensations elicited by the two stimuli might be confused or might even be indistinguishable. This reduces the number of independent channels of information that can be conveyed to the cochlear implant user's auditory system. Realizing that the number of independent channels of information is not equal to the number of electrodes is important in the design of processors for cochlear implants. Chapters 2 and 4 investigates the hypothesis that electrode interaction can be measured with gap detection, and shows that estimates can be obtained for the current distribution in the cochlea.

5 THESIS OUTLINE

The hypotheses summarized in 4.1 to 4.4 above is reflected in the layout of this thesis. The chapters are entirely self-contained and some of them have been published.

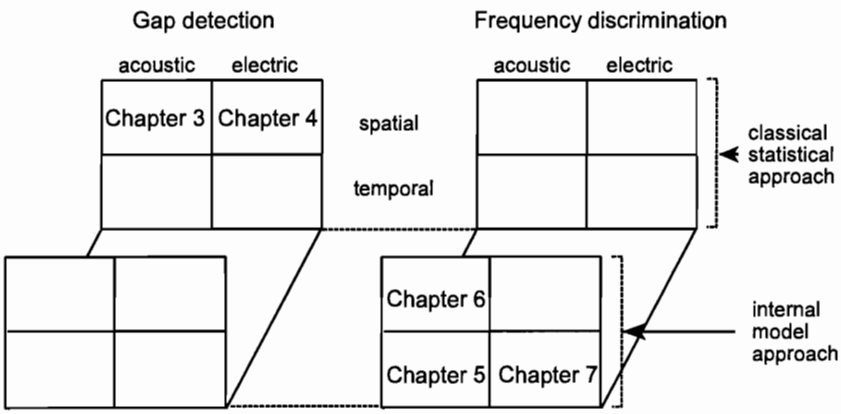

Figure 1.3.
Illustration of the multidimensional problem addressed in this thesis, and an indication of where each chapter fits in.

Figure 1.3 summarizes the layout of the thesis. The figure shows that in some chapters model are created using a classical statistical approach, while in other chapters an internal model approach is followed. Temporal and spatial models of acoustic and electric hearing are investigated, and psychoacoustic experiments that use frequency and time signals are modelled.

Chapter 2 describes psychoacoustic data for gap detection, a technique usually used for probing the temporal ability of the auditory system, but used here to measure spatial characteristics in auditory electrical stimulation. The material in this chapter has been published in Hanekom and Shannon (1998).

Chapter 3 creates a model of gap detection in acoustic hearing and shows that gap detection data can be explained in terms of spatial mechanisms.

Chapter 4 expands the model of chapter 3 for electric hearing to predict the data of chapter 2. This chapter shows that the same model used for acoustic hearing may be applied to electric hearing if the inputs are appropriate. The chapter also shows that current distribution in the cochlea influences channel interaction.

In addition, chapter 4 shows how predictions for current distributions in the cochlea may be obtained from gap detection data. Predicted current distributions (based on the gap detection data in chapter 2) are then used to predict electrode discriminability. These predictions are compared with electrode discriminability data measured in the same subjects (previously published in Hanekom and Shannon, 1996).

Chapter 5 develops a spatial model for frequency discrimination in acoustic hearing. The model is built on the concept of an internal model. A priori knowledge about spike train statistics and possible frequencies that may exist in the "external world" is used in a Markov model for the signal. A non-linear frequency estimator that observes a spatial spike train pattern then estimates the frequency of a pure tone impinging on the auditory system. The material in this chapter has been published in Hanekom (1999).

Chapter 6 develops a model for frequency discrimination in acoustic hearing, based on temporal mechanisms. This model uses a Kalman filter to estimate the pure tone auditory input frequency from temporal information in spike trains. The Kalman filter incorporates an internal model of the system that generates the spike trains and uses an analysis-by-synthesis mechanism to arrive at frequency estimates. The material in this chapter has been published in Hanekom and Krüger (2001).

Chapter 7 extends the temporal model for frequency discrimination in acoustic hearing (chapter 6) to electric hearing. The chapter then interprets psychoacoustic data obtained with cochlear implants in terms of spatial and temporal mechanisms for the coding of frequency information in the auditory system. The material in this chapter has been published in Hanekom (2000).

The thesis is concluded in chapter 8 with a summary and discussion of the main results and findings. It is shown there that the four hypotheses have been proven in a qualitative fashion.

Chapter 2

GAP DETECTION AS A MEASURE OF ELECTRODE INTERACTION IN COCHLEAR IMPLANTS

The results in this chapter have previously been published: Hanekom, J.J. & Shannon, R.V. 1998, "Gap detection as a measure of electrode interaction in cochlear implants", *Journal of the Acoustical Society of America*, vol. 10 no. 4, pp. 2372-2384.

1 INTRODUCTION

The individual electrodes in a modern multi-electrode cochlear implant are intended to selectively stimulate discrete neural populations. However, the assumption that discrete neural populations can be activated is not always true. It is widely assumed that stimuli applied between closely-spaced or adjacent bipolar electrode pairs lead to the localized activation of neurons, whereas widely spaced bipolar electrode pairs (including monopolar stimulation) will lead to broad electrical fields and wide areas of neural activation (van den Honert and Stypulkowski, 1987b; Busby et al., 1994). Even for a closely spaced electrode pair, a broad region can be activated at high stimulation current levels (van den Honert and Stypulkowski, 1987b). The consequence is that when two sets of bipolar electrode pairs are stimulated, and these two sets are closely spaced, overlap can occur in the neural populations excited by the stimulation currents. This overlap of neural populations can occur regardless of whether the stimuli are non-simultaneous or simultaneous. Simultaneous stimuli give rise to direct electrical field interactions, which pose additional problems for electrical stimulation, but even non-simultaneous stimuli may produce activation of overlapping neural regions.

If two electrode pairs stimulate the same neural population or overlapping neural populations, the implication is that sound sensations elicited by the two stimuli might be confused or might

even be indistinguishable. This may reduce the number of independent channels of information that can be conveyed to the cochlear implant user's auditory system, presumably resulting in a deterioration of speech recognition ability. If two electrode pairs stimulate the same population of neurons and are perceptually indistinguishable, they probably cannot convey two separate channels of information.

1.1 The number of information channels in an implant

Studies by Fishman et al. (1997) and Lawson et al. (1993, 1996) indicate that increasing the number of electrodes does not necessarily lead to better speech recognition. In fact, very slight or no improvement was evident when the number of electrodes used was increased from 7 to 20. For some speech recognition tasks, no improvement in performance was found when the number of electrodes used increased from 4 to 20. In these experiments, all spectral information that is usually presented across all 20 electrodes, was applied to a limited number of electrodes, i.e. no spectral information was discarded. In a more recent study, Friesen et al. (2001) tested speech recognition as a function of the number of electrodes used in noisy conditions. Nineteen implant users of two different implants (Nucleus-22 and Clarion) participated. A general finding was that, for all noise levels, consonant and vowel recognition scores improved up to seven electrodes, and that speech recognition improved up to ten electrodes, irrespective of the implant used. A reduction in the number of channels was equivalent to a reduction in the signal-to-noise ratio at low signal to noise ratios.

These results suggest that the actual number of information channels available to these patients was not a function of the number of electrodes, and that the actual number of information channels might be limited to somewhere between 4 and 7. Interestingly, in a study with normal-hearing listeners, Shannon et al. (1995) used 4 channel processors and found that listeners achieved near-perfect speech recognition, implying that 4 information channels might be adequate, at least in quiet listening conditions. The study of Friesen et al. (2001) included normal-hearing listeners that listened to a noise-band simulation of a CIS-like processor. (CIS is a stimulation strategy used in the Clarion implant. See Wilson et al., 1991). It was found that

speech recognition continued to improve up to 20 channels under similar noise conditions than used for the cochlear implant listeners.

In another study, on patients with the Nucleus cochlear implant device, Hanekom and Shannon (1996) showed that for several different seven-electrode speech processors, speech recognition performance was a function of which set of 7 electrodes were used. This indicates that different choices of which electrodes are used in a processor might lead to different numbers of information channels. A reduced number of electrodes, including only discriminable electrodes, were also used in the speech processors of eleven Nucleus cochlear implant users who participated in a study by Zwolan et al. (1997). While some subjects showed significant improvement in specific speech recognition tasks, others showed a decline in speech recognition performance. Although no strong relationship between electrode discrimination performance and speech recognition was observed, this study again indicates that the choice of electrodes in a reduced electrode processor influences speech recognition ability in some implant users. This supports the suggestion that the number of information channels is a function of the choice of electrodes in a reduced electrode processor. Lawson et al. (1996) measured a larger difference in performance between two different selections of six electrodes than between six and 20 electrodes. This suggests that there should be a way to maximize the number of information channels used for a specific subject by judicious choice of electrodes. Further maximization may be possible using electrical field focussing (Townshend et al., 1987), or by compensating for a missing patch of nerve, or by shifting the speech analysis filters to better match the electrode location (Fu and Shannon, 1999). No maximization of this sort is presently done in implant programming strategies, partly because measurement tools are not yet available and partly because the relation between the electrode interaction and information channel capacity is not well understood.

1.2 Physical factors affecting electrode interaction

To fully account for the effects of electrode interaction we must (1) identify the factors in the patterns of speech that are most important for speech recognition (Shannon et al., 1995), (2)

be able to measure the electrode interaction pattern in an individual implant patient, and (3) use the information from both (1) and (2) to optimize the reception of the most important speech pattern information for an individual patient.

A number of variables can influence the interaction of electrodes in a cochlear implant user. These include the electrode placement within the cochlea and nerve survival at the cochlear level and also at the central auditory level.

1.2.1 Electrode placement.

The proximity of the electrode to the surviving neurons, as well as the impedance and paths of current flow between the electrode and the neural population, will determine the spatial selectivity of the stimulation. The impedance and the current pathways could be influenced by new bone formation in the implanted cochlea and encapsulation tissue around the electrode (Grill and Mortimer, 1994). Broad spread of activation will occur if the electrode is physically distant from the excitable neurons (along the lateral wall of the cochlea for example, rather than next to the modiolus) or if nerve survival is poor immediately adjacent to the electrode. Although not routinely used, techniques such as spiral tomography (Wang et al., 1996) are available to measure the exact placement of the electrodes inside the scala tympani. The absolute electrode location can then be used to deduce which nerve fibers will be activated. Finley et al. (1990) modeled nerve fiber activation in a finite element model with idealized electrode placement, but no work has been reported using real electrode placement data.

Although it is clear that placement of electrodes further from the modiolus requires higher stimulus levels to reach threshold and consequently leads to larger current spread, the influence of electrode placement is not yet quantified regarding the interaction or independence of information channels. It is generally assumed that placement of electrodes close to the modiolus is preferable because more focused stimulation can be achieved (Rebscher et al., 1994). Unfortunately, very few tools are available for perceptually assessing and quantifying electrode absolute location and spatial selectivity and their exact influence on speech recognition.

1.2.2 Nerve survival.

A second factor that should affect electrode interaction is the nerve survival pattern in an individual patient. Several anatomical post-mortem studies (Zimmermann et al., 1995; Linthicum et al., 1991; Fayad et al., 1991) have shown from human temporal bones that nerve survival patterns vary greatly among subjects, even for the same disease. It is not clear how the amount and pattern of neuron survival affects implant performance. However, with fewer neurons, the distance between the stimulating electrode and neurons might be larger. Certainly, the further the neurons are distant from the electrode, the larger the current required for activation and the broader the spread of activation. This, in turn, may reduce the number of independent information channels.

It is clear that it is necessary to quantify the available auditory abilities and to optimize the use of the available information channels, i.e. to optimize information transfer in the current generation of implants. Tools are needed to establish the number, the location, and the characteristics of information channels available in each individual cochlear implant user. In this chapter, gap detection is proposed as one such tool.

1.3 Gap detection as a measure of tonotopic spread

Gap detection has traditionally been used as a measure of temporal processing (Plomp, 1969). At moderate levels and higher, normal-hearing listeners can detect 3-5 ms gaps in a stimulus when identical stimuli are bounding the gap, irrespective of the frequency of the stimuli (Penner, 1976; Fitzgibbons, 1983; Florentine and Buus, 1984; Hall et al., 1996; Shailer and Moore, 1983). This results is characterized as the "within-channel" temporal resolution. However, when the frequencies or levels of the two stimuli that bound the gap are different, gap detection thresholds increase about an order of magnitude - to 30-50 ms (Divenyi and Danner, 1978; Divenyi and Sachs, 1979; Formby and Forrest, 1991, Formby et al., 1992). In this case, even the standard stimulus with no gap is perceived as having a discontinuity. The discontinuity that identifies the actual gap must be long enough to be distinctive from this no-gap, standard condition. This temporal comparison must be done centrally "across channels"

in that the two stimuli bounding the gap are processed through largely independent neural pathways. A simple model of peripheral frequency resolution can largely explain these results, indicating that gap detection can indicate the degree of neural population overlap between two stimuli (Heinz et al., 1996).

In cochlear implant users, Chatterjee et al. (1998) observed that "within-channel" gap detection thresholds increase when the stimuli marking the gap were of unequal amplitude or unequal pulse rate. They concluded that the perceptual discontinuity caused by dissimilar markers complicated the gap detection task, and suggested that under these conditions gap detection thresholds may be a function both of limitations caused by peripheral mechanisms and a central perceptual distance detector. Their results also emphasize the importance of loudness balancing the stimuli marking the gap.

Shannon (1989) measured gap detection thresholds in cochlear implant users as a function of stimulus level, for both closely spaced (bipolar) and widely spaced (monopolar) electrode configurations, using sinusoids and pulsatile stimuli. He found that gap detection thresholds were a strong function of stimulus level, with the shortest gap thresholds in the order of 1.5 to 3.1 ms regardless of the separation between the active and reference electrodes. He concluded that the temporal resolution for implant subjects was as good as or better than for normal-hearing listeners. However, all measures were made with the stimuli marking the gap on a single electrode pair, i.e., no cross-channel gap detection was done.

The present study measures gap detection thresholds as an indicator of the characteristics of the available neural channels, i.e. the number of channels available, the position of these channels (which electrodes provide independent channels) and the width of the channels. A simple conceptual model is hypothesized which relates gap detection thresholds to neural excitation. When the two stimuli that bound the gap are presented on different electrode pairs, it is expected that gap thresholds will be short if the two electrode pairs stimulate the same neural population. Gap detection in this case is presumably determined by a “within-channel” temporal mechanism and so is determined by the time constant of the peripheral auditory

system. As the electrode pairs are separated and the amount of neural overlap decreases, temporal information is carried in separate neural pathways, the stimuli sound more dissimilar and the gap thresholds are expected to increase. Gap detection in this case is presumably limited by the time constant of the centrally located auditory integration because the comparison is made "across-channels". As the two electrodes defining each of the electrode pairs are separated, moving from BP stimulation mode (bipolar between adjacent electrodes) toward BP+3 stimulation mode (bipolar between nonadjacent electrodes with three electrodes separating the stimulation pair), the amount of neural overlap between the two electrode pairs is also expected to increase, resulting in reduced gap thresholds. Using the same argument, the gap thresholds should presumably also be higher for lower levels of stimulation, as there would be less spread of excitation. According to this model, gap detection thresholds can be used to infer the amount of overlap in neural populations stimulated by two pairs of electrodes.

2 METHODS

2.1 Subjects

Three users of the Nucleus cochlear implant participated in this study. All were users of the Nucleus Spectra speech processor, which implements the SPEAK speech processing strategy (McDermott, 1989; McDermott et al., 1991). They were highly trained in various psychoacoustic experiments, having participated in many similar experiments over a period of months. Table 2.1 contains detailed demographic information on the three subjects.

2.2 Electrode parameters

All three subjects used the Nucleus 22 electrode array (Clark et al., 1990), implanted into the scala tympani. The electrodes are numbered from 1 at the basal end to 22 at the apical end. Adjacent electrodes were separated by 0.75 mm. Electrode pairs are referenced by their basal-most member (the *active* electrode); the *reference* electrode is the apical-most member of an electrode pair.

Table 2.1.

Subject information for the three subjects who participated in this study. Insertion depth refers to the number of electrode bands inside the cochlea. The first twenty-two electrodes are active, but eight additional inactive electrode rings aid in placing the electrode and measuring the insertion depth. Speech recognition scores for these subjects were obtained in a previous study (Fishman et al., 1997). Recognition of words from sentences was measured with the CUNY everyday sentences. For consonant and vowel recognition tests, sixteen medial consonants in a v/C/v context and eight vowels in a h/V/d context were used.

Subject	Age	Gender	Age of onset of profound hearing loss	Time of implant use	Processor type	Insertion depth	Cause of deafness	Sentence recognition	Vowel recognition	Consonant recognition
N3	55	Male	45	6 years	SPEAK	27	trauma	61	58	46
N4	39	Male	35	4 years	SPEAK	26	trauma	95	92	95
N7	54	Male	47	6 months	SPEAK	22	unknown; progressive hearing loss	71	98	75

The Nucleus speech processor allows different stimulation modes. Stimuli were presented either in bipolar mode between adjacent electrodes (BP); bipolar between nonadjacent electrodes for electrode separations up to 3 mm (BP+1: 1.5 mm separation; BP+2: 2.25 mm separation; BP+3: 3mm separation); or in pseudo-monopolar mode, using the apical-most electrode as reference electrode. Pseudo-monopolar mode is not a true monopolar mode, as the reference electrode is not located remotely, but inside the scala. In this mode, which will be called AR (apical reference) mode for simplicity, the actual mode of stimulation varies with the active electrode position, so that, for example, when electrode 20 is used as active electrode, the mode is BP+1. The spread of the current field should be larger for larger spacing

between the active and reference electrodes of the pair.

2.3 Stimulus parameters

All stimuli were charge-balanced, 200 μs/phase biphasic pulses, with anodic phase first, and were presented at a stimulation rate of 1000 pulses per second. Stimuli were presented at a comfortable level of stimulation. The stimuli were loudness balanced across electrodes before the start of the experiment, using a bracketing loudness balance procedure. First, thresholds and upper loudness levels were obtained in each stimulation mode. Then the subjects were asked to choose a comfortable level of stimulation on electrode 10. All subjects chose comfort levels somewhere between 50% and 85% of their dynamic ranges in the various stimulation modes. All other electrodes were then loudness balanced to this electrode by instructing the subject to adjust the loudness of an adjustable stimulus to be just louder than, then just softer than and finally equal to the reference stimulus. Loudness was adjusted by adjusting pulse amplitude. This was repeated as many times as was necessary to obtain consistent decisions about the relative loudnesses. Loudness balancing was repeated for all conditions (each level of stimulation in each stimulation mode).

Gaps were presented between two 200 ms stimuli. These two stimuli were presented on the same electrodes in the baseline condition and on different electrodes otherwise. Gap thresholds were measured as a function of the separation of the two electrodes. In a single run, the first electrode position was held constant, and gap thresholds were measured for different positions of the second electrode. The experiment was performed in BP, BP+1, BP+2, BP+3 and a pseudo monopolar mode as described above.

A computer program generated the appropriate stimuli and recorded the subject responses. The stimuli were encoded in the correct format to enable presentation directly to the internal receiver of the Nucleus device (without using the subjects' processors), via a custom interface (Shannon et al., 1990).

2.4 Psychophysical procedure

Gap thresholds were collected using an adaptive, two-interval, forced-choice procedure. The gap was initially 100 ms and two consecutive correct decisions led to a decrease in gap size, and one error increased gap size. This procedure estimates the gap size required for 70.7% correct responses (Levitt,1971). Initially the increase or decrease was by a factor of two, but after four reversals this factor was 1.3. Data collection was for twelve reversals and the mean of the last eight reversals was used to estimate the gap threshold.

Gap detection thresholds were obtained in BP+1 mode for all three subjects using all the even-numbered electrodes as standard (the first stimulus). For each standard, the gap thresholds were measured as a function of probe electrode (just even numbered or both even and odd numbered) separation from the standard. Three repetitions were made for each measurement, which resulted in six measurements of gap threshold for each combination of stimulation electrodes when using both orderings of electrodes. (That is, when electrode i was used as standard, three measurements were obtained for probe electrode j, and when j was the standard, another three measurements were obtained with i as probe). Also, gap detection thresholds were obtained in BP, BP+2, BP+3 and AR modes for all three subjects using electrodes 6, 10 and 14 as standard. Again, gap thresholds were measured on even numbered electrodes as a function of probe electrode separation from the standard. In this task two to six measures were taken at each probe electrode location.

3 RESULTS

3.1 Gap threshold as a function of electrode separation

Figure 2.1 compares gap threshold data from Formby et al. (1996) (in normal hearing listeners) to gap threshold data from cochlear implant patients. Formby et al. (1996) measured gap thresholds as a function of marker frequency separation and the study described in this chapter

measured gap thresholds as function of electrode separation in cochlear implant patients. The electrode axis in figure 2.1 is scaled to match the approximate location of the linearly spaced electrodes to the cochlear frequency-position function of Greenwood (1990). There is good agreement between the two sets of data in the shape of the gap threshold curves and in the absolute values of gap thresholds for implant subjects N3 and N7. Gap thresholds for implant listener N4 were consistently lower than those from Formby et al. at every comparison point.

Gap thresholds were measured as a function of electrode separation for ten standard electrodes (all the even numbered electrodes) for each of the three subjects (figures 2.2 to 2.4). The lowest gap thresholds were always achieved when the two stimuli that bound the gap were presented on the same electrode. The minimum values of gap threshold were near 1 ms for most electrodes for N4 and 3 to 4 ms for the other two subjects. This is consistent with the range of gap thresholds reported by Shannon (1989).

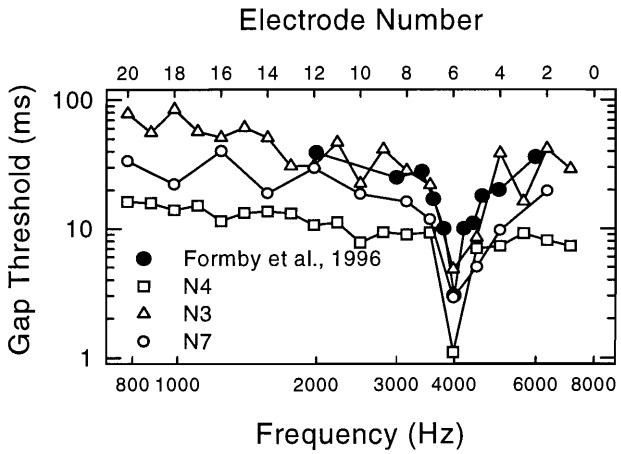

Figure 2.1.

Comparison of gap detection as a function of electrode separation with comparable results from Formby et al. (1996) on gap detection as a function of frequency separation between marker stimuli. The electrode number axis (top) has been reversed and scaled to match the approximate location and extent of the electrode according to Greenwood's (1990) formula. Data for N3, N4 and N7 was obtained in BP+1 stimulation mode.

Gap thresholds increased considerably as electrode separation increased. In general, gap thresholds increased by almost a factor of 10 as the two electrodes were separated. The absolute values and ranges of gap thresholds varied considerably among the subjects, particularly when electrodes were widely separated. Subject N4 had gap thresholds of between 10 and 20 ms for widely separated electrodes, while subject N7 had maximum gap thresholds of 20-70 ms, and N3 had maximum gap thresholds of 100-200 ms.

For most electrodes towards the basal end of the array, the spatial selectivity of the gap threshold curves was sharpest for N4, while N7 had broader selectivity, and N3 had broad "spatial tuning" that covered most of the length of the electrode array. For simplicity, the gap threshold curves will be referred to as "tuning curves". Many of the gap detection tuning curves have two portions: a sharply tuned "tip" region in the vicinity of the standard electrode, and a shallow, bowl-shaped portion for electrodes distant from the standard. These two sections may reflect two different mechanisms relating electrode similarity to gap detection.

Figures 2.2 to 2.4 show a general tendency for the slopes of the bowl-shaped portion to become steeper on the apical side of the gap threshold tuning curves (i.e. towards electrode 20) and shallower on the basal side as the standard moved from base to apex. For electrodes near the base, asymmetry was towards the apex (slopes were shallower on the apical side). This is consistent with measurements of electrode interaction in the same three subjects, using forward masking (Chatterjee and Shannon, 1998). The shallower slopes towards the base (for apical electrodes) suggest larger current flow towards the basal region, but the shallower slopes towards the apex (for basal electrodes) suggest larger current flow towards the apex. Previous measures of electrode interaction using forward masking (Lim et al., 1989) suggested larger current flow towards the basal region for stimulated electrodes at all cochlear locations, but the data presented here does not confirm this observation.

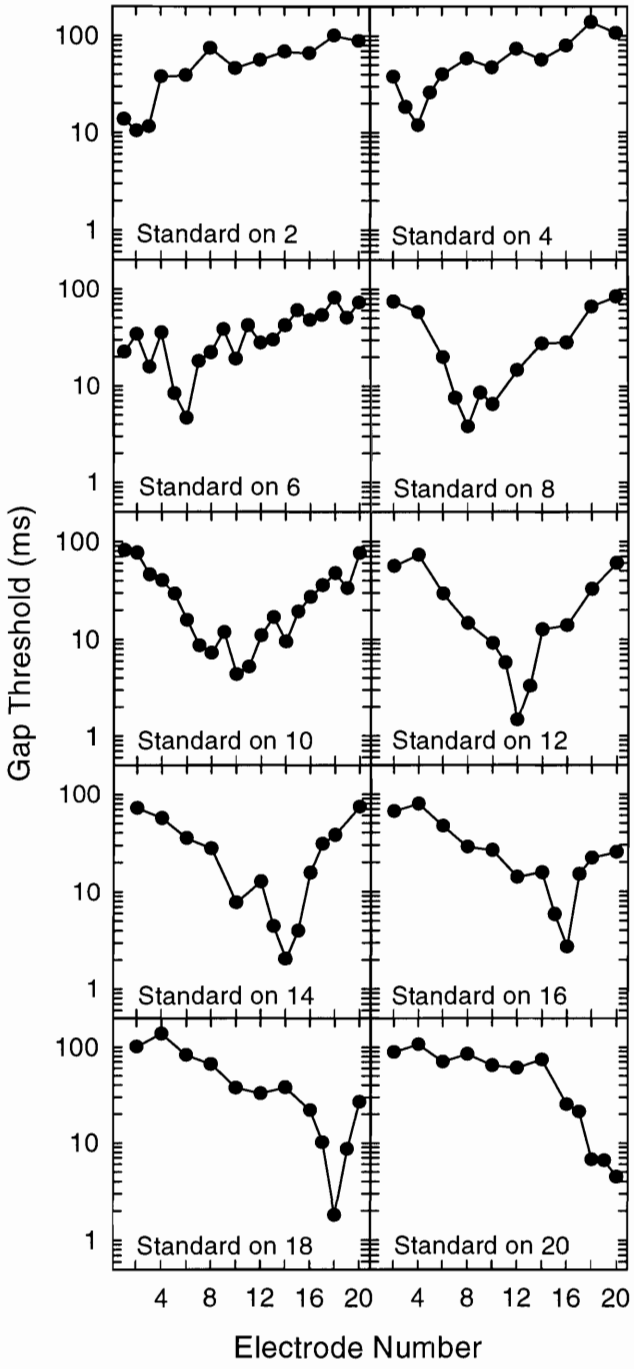

Figure 2.2.

Gap detection for N3 in BP+1 stimulation mode as a function of separation between electrodes. One of the marker bursts was presented to the standard electrode pair and the other to another electrode pair. Gap detection "tuning curves" are shown for all even numbered electrodes as standard. Gap detection thresholds were measured on all even-numbered electrodes.

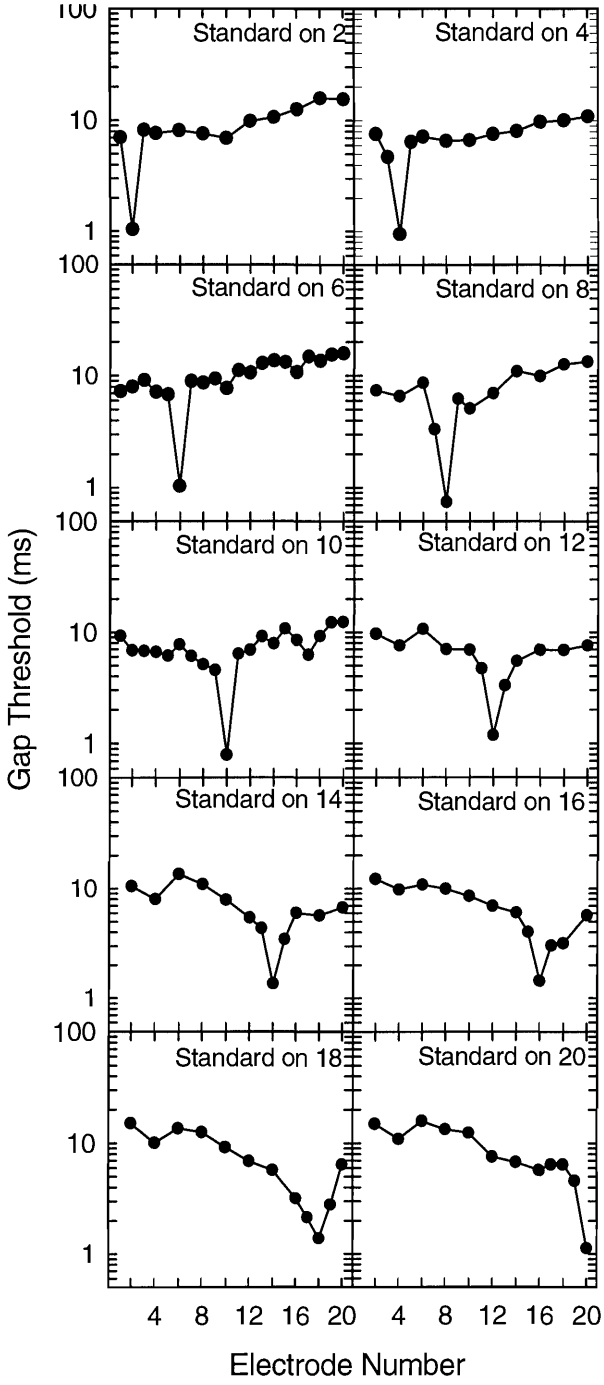

Figure 2.3.

Same as figure 2.2, but for subject N4, using BP+1 stimulation mode.

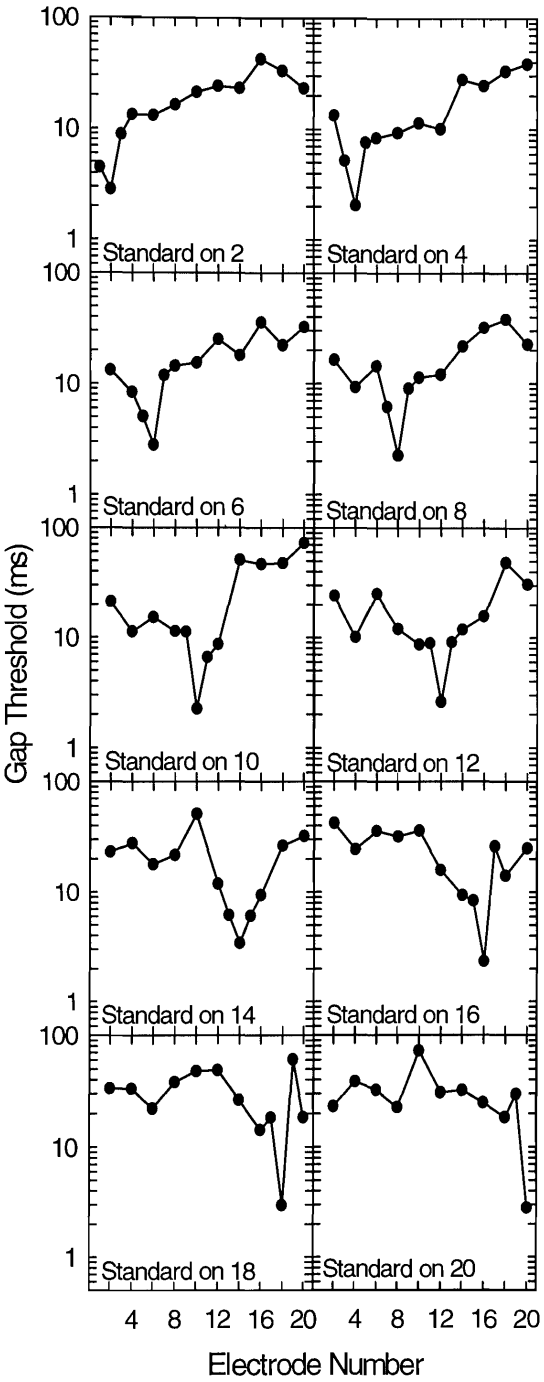

Figure 2.4.

Same as figure 2.2, but for subject N7, using BP+1 stimulation mode.

3.2 Gap thresholds as a function of level of stimulation

Figures 2.5 to 2.7 show the gap thresholds for two levels of stimulation for each of the three subjects, for three standard electrodes. Stimulation was either at a relatively loud level (comfort level was at 81% of the dynamic range for N3, 84% for N4 and 69% for N7) or a softer level (around 30% of the dynamic range for all subjects). BP+1 stimulation mode was used throughout. The most striking difference between the gap threshold tuning curves at high

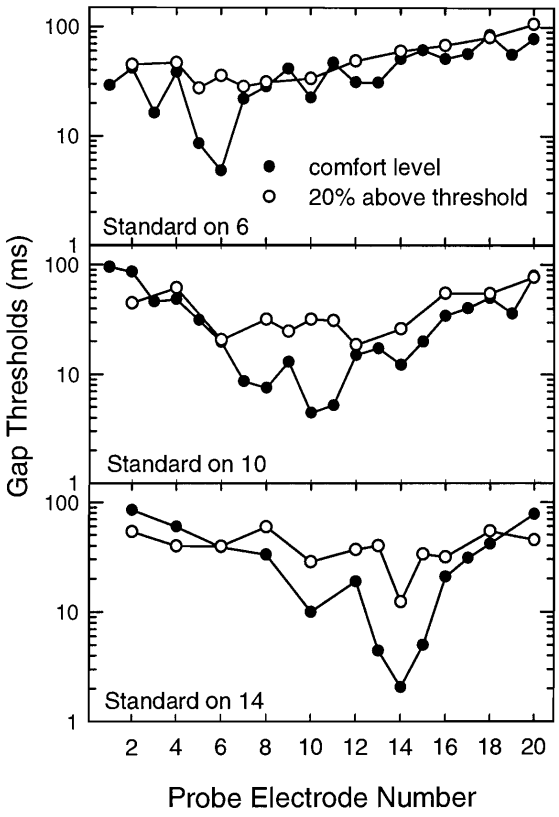

Figure 2.5.

Gap detection tuning curves for N3 for two stimulation levels. The comfortable stimulation level was at 81% of the dynamic range. Each panel represents a different standard electrode location.

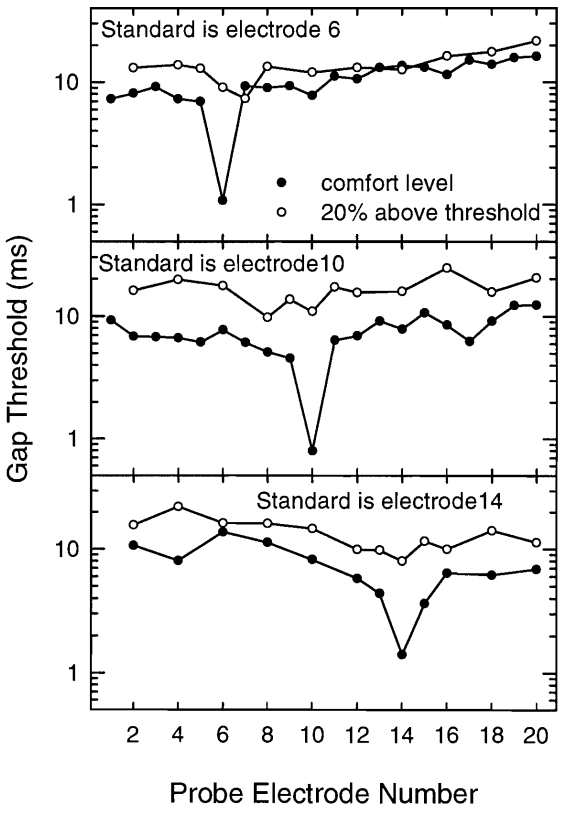

Figure 2.6.

Gap detection tuning curves for N4 for two stimulation levels at three standard electrode locations. The comfortable stimulation level was at 84% of the dynamic range.

and low levels is the absence of the sharp tip region in most cases at low levels. In the region of the tuning curve tips gap thresholds increased at the softer levels in every case. In addition to the loss of the sharp tips, for N4 there was also a 5 to 10 ms increase in gap thresholds across the whole pattern at the lower level. Subjects N3 and N7 did not show a clear shift in gap thresholds with level away from the tip region.

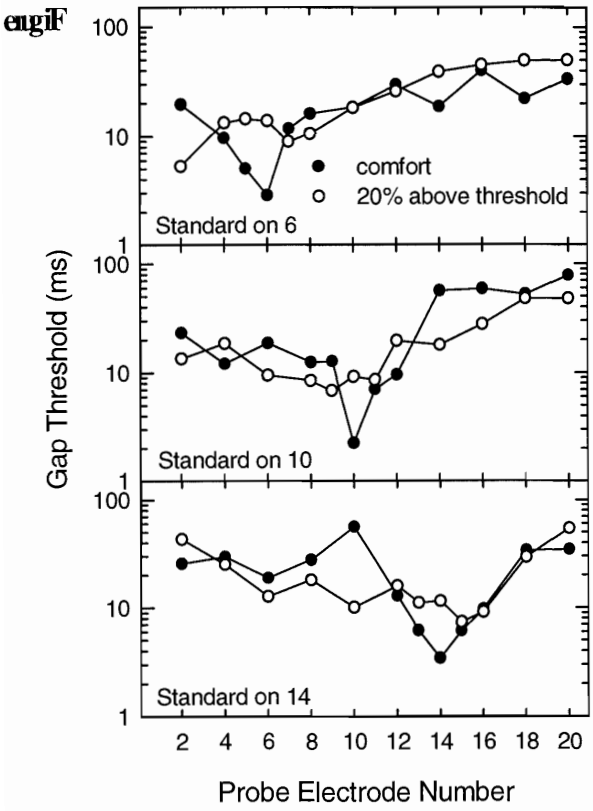

Figure 2.7.

Gap detection tuning curves for N7 for two stimulation levels at three standard electrode locations. The comfortable stimulation level was at 69% of the dynamic range.

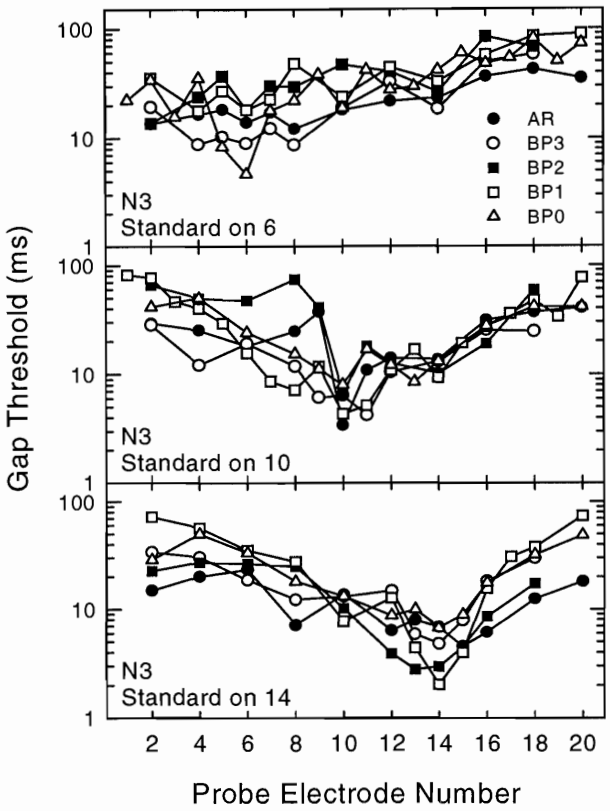

Figure 2.8.

The five curves in each panel are gap detection tuning curves for N3 for five different spacings between the two electrodes of a bipolar pair (i.e., five stimulation modes). The three panels represent measurements at three standard electrode locations.

3.3 Gap threshold as a function of mode of stimulation

N3 (figures 2.8 and 2.11) exhibited the sharpest tuning in BP+1 mode and poorer tuning in all other modes, with the exception of sharp tuning in AR mode when the standard was on electrode 6. Surprisingly, his poorest tuning occurred in BP mode, which should produce the most localized current field. Tuning curves were so flat in BP+2 mode and in AR mode when the standard was on electrode 6 that tuning curve widths could not be calculated. Gap thresholds were greater than 10 ms even at the tip of the tuning curve. There was a 10 ms difference between the AR and the BP gap threshold curves near the tip and a difference as large as 50 ms across stimulation modes away from the standard electrode.

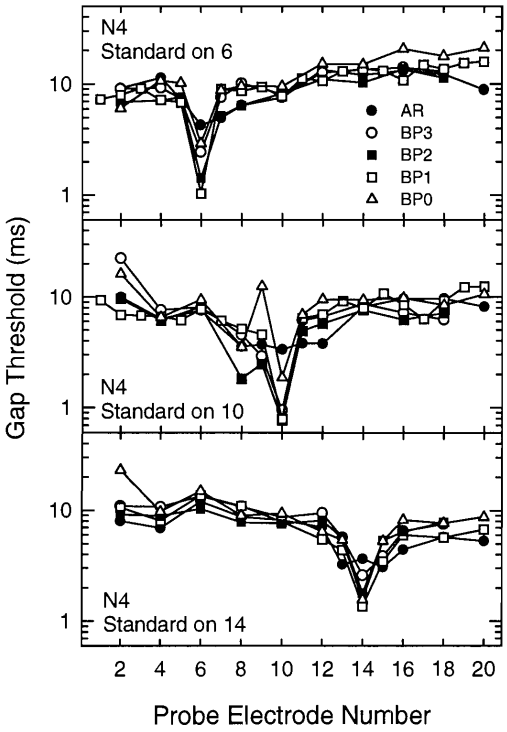

Figure 2.9.

Same as figure 2.8, but for subject N4. Gap detection tuning curves for five stimulation modes at three standard electrode locations.

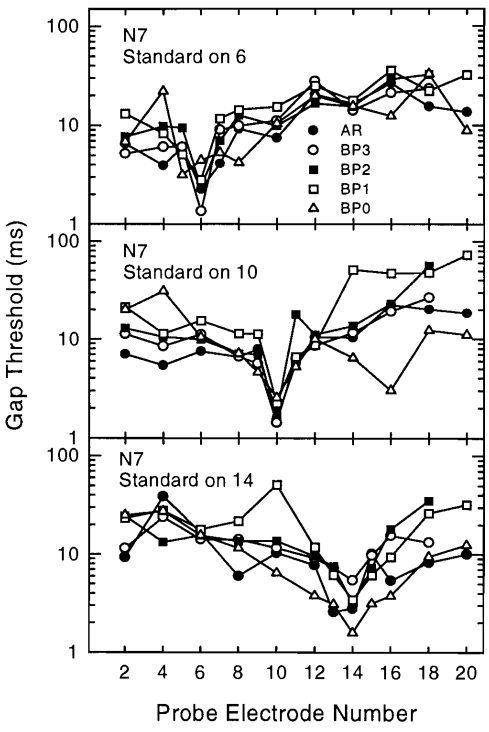

Figure 2.10.

Same as figure 2.8, but for subject N7. Gap detection tuning curves for five stimulation modes at three standard electrode locations.

N4 (figures 2.9 and 2.11) demonstrated almost the same sharp tuning in most stimulation modes, with decidedly broader tuning in AR mode. The difference between the stimulation modes was most pronounced for electrode 6, with BP having a significantly steeper slope than AR. At the tip of the tuning curve, BP+1 produced the lowest gap threshold among the stimulation modes.

For N7 (figure 2.10), tuning did not change dramatically across all conditions. Overall, BP exhibited the widest tuning - even wider than AR mode. Lowest gap threshold also did not change for N7 across all stimulation modes and standard electrode locations.

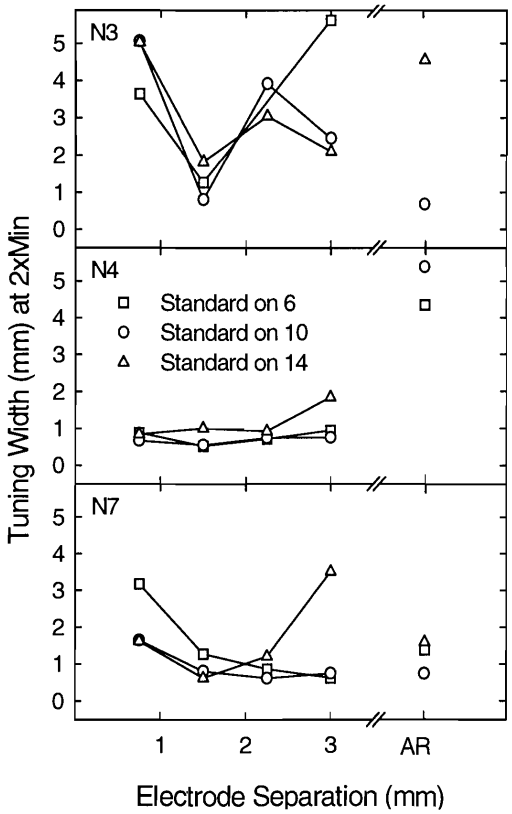

Figure 2.11.

Width of the gap detection tuning curves at twice the gap threshold value at the tip as a function of the separation of active and reference electrodes. The three curves in each panel represent the tuning width measures for three standard electrode locations.

Overall, these three implant listeners showed quite different patterns of tuning as a function of stimulation mode. Based simply on electrical field spread it would have been expected that tuning curve widths would have broadened as the separation between active and reference electrodes in each pair increased. It was expected that AR mode would produce the poorest tuning. No subject showed this expected pattern, although N4 at least showed the poorest tuning in AR mode. N3 showed the lowest gap thresholds and the sharpest tuning in BP+1 mode, which was the stimulation mode used in his normal speech processor. There was no clear change in the pattern of tuning for different standard electrode locations - similar tuning was generally observed for a given listener whether the standard electrode was 6, 10 or 14.

4 DISCUSSION

4.1 Relation between spatial selectivity and gap detection thresholds

The hypothesized relationship between gap detection thresholds and neural activation is reflected in the graphs of gap detection as a function of electrode separation (figures 2.2 to 2.4). Short gap detection thresholds were found where neural interaction was assumed to be large (zero or small electrode separation) and larger gap detection thresholds were found as the separation between the two electrode pairs increased. The conclusion is that a narrow "tuning" in the gap detection thresholds is an indication of good neural selectivity.

The data in figure 2.11 was used to calculate a two-factor ANOVA, to test the statistical relationship between tuning width and two factors: stimulation mode and subject. The ANOVA indicated a statistically significant difference between the tuning widths obtained for the three subjects ($F(2,30)=4.74$, $p=0.016$). For these data, the sentence recognition scores (table 2.1) were higher in subjects with smaller tuning widths, although this cannot be stated as a general rule as the statistical sample was too small.

The shortest gap thresholds observed at the tip of the tuning curves were similar across

subjects, 1-4 ms. This is consistent with the results of Shannon (1989) who saw similar gap thresholds across patients at the highest stimulation levels. The two studies thus indicate that there is little relation between the best gap thresholds and speech recognition performance. Both studies found similar gap thresholds across subjects and included subjects with a wide range of speech recognition performance.

However, the longest gap threshold, generally observed for widely separated electrodes, may be related to speech recognition performance. Gap thresholds in this case differed by an order of magnitude between the best and poorest implant user. It is not clear what factor might underlie such a large difference in gap thresholds. The two curve segments (sharp tip and shallow bowl) may indicate the selectivity of two mechanisms of similarity between electrodes (figure 2.12). The sharp tip may reflect a peripheral/neural process that indicates the amount of overlap in the neural populations excited by the two electrodes. The "shoulder" of the gap threshold tuning curves may indicate a point of transition to a condition where electrodes do not stimulate overlapping neural populations. When electrodes are moved even further apart, a further increase in gap detection thresholds would not be expected according to our conceptual model. However, the shallow bowl portion of the function may indicate a weak effect of perceptual similarity for two electrodes that do not activate overlapping neural populations. Although the gap detection must be performed centrally in this case the shallow bowl-shaped function may indicate that there is also a mild effect of overall perceptual similarity on gap detection. The transition that is heard between electrodes that are highly distinctive complicates the gap detection task. Electrodes that are highly distinctive require longer gaps for detection than two electrodes that are less distinctive, even if no neural populations are in common in either case.

The models developed in chapters 3 and 4 shows that the gap detection tuning curves can also be predicted by peripheral mechanisms alone. In the model in chapter 3 (for gap detection in acoustic hearing), the shoulder of the gap detection tuning curve (where the sharp tip transforms into the shallow bowl) is speculated to be a transition from a gap detection task to a gap discrimination task. In the model in chapter 4 it is shown that the sharp tip is obtained

when entrainment is close to 100% (i.e. spike occur on every cycle of the stimulus waveform), while the shallow bowl results when entrainment is less than 100%. This notion is supported by the data in figures 2.5 to 2.7, where the sharp tips disappear at lower stimulation levels where entrainment levels are expected to be below 100%.

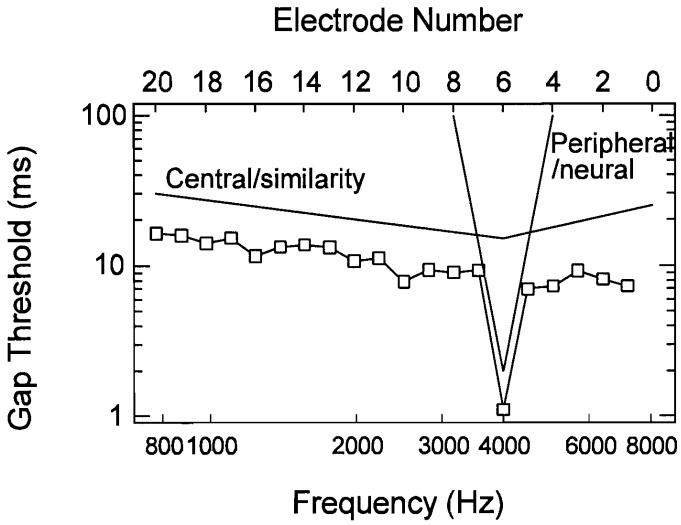

Figure 2.12.

Schematic representation of a conceptual model of gap detection. When the two markers defining the gap excite overlapping neural populations gap thresholds are lowest, but this mechanism is not useful if the neural populations do not overlap. When the neural populations of the two markers do not overlap the gap threshold is determined by a slower, central mechanism. This hypothesized mechanism is only broadly tuned in that markers that are more similar produce slightly lower gap thresholds than markers that are highly dissimilar.

Our original hypothesis was that the longer gap detection time was indicative of the time constant of a central mechanism comparing outputs from different peripheral neural channels. If this is the case, then the long gap thresholds exhibited by N3 may be long enough to interfere with recognition of speech transitions across channels. Long gap thresholds can presumably lead to the misinterpretation or missing of important temporal cues for the identification of consonants (e.g. voice onset time; Divenyi and Sachs, 1978), resulting in poorer speech

recognition performance. In the case of N4, poorest gap detection was still 10-20 ms, which may be rapid enough to allow processing of all relevant cross-channel speech transitions. It may be that within-channel gap detection is not a limiting factor for speech recognition, but that long, central cross-channel comparison times can interfere with speech temporal distinctions. It is not clear what might cause the large difference observed in these gap thresholds between different subjects. It appears that there is a larger range of individual differences in the implant results than in similar conditions with acoustic hearing in normal listeners (Formby et al., 1991).

4.2 Relation between stimulation level and spatial selectivity

Gap thresholds are a strong function of stimulus level in cochlear implant users (Shannon, 1989). All of Shannon's measurements were made with identical markers before and after the gap, a condition similar to the tips of the tuning curves in the present data. The data in figures 2.5 to 2.7 also show longer gap thresholds at softer stimulus levels around the tip region. However, gap thresholds did not change as much with level away from the tip region, resulting in the loss of the sharp tips of the tuning curves at lower levels. Less interaction would be expected at lower stimulus levels, because the region of neurons activated should be smaller. Our conceptual model would predict narrower tips at low stimulus levels rather than no tips. However, the interpretation of the present data are confounded by the strong change in gap thresholds with level. In conditions where the markers before and after the gap were on different electrodes (away from the tip region), little change in gap thresholds with level was observed in two of the three subjects. According to our conceptual model this suggests that central mechanisms of gap detection are less dependent on stimulus level than peripheral mechanisms of gap detection. Clearly, more data is needed to validate this suggestion.

4.3 Relation between stimulation mode and spatial selectivity

As the active and reference electrodes in a bipolar pair are separated the electric field becomes more diffuse and spatial selectivity decreases. In general, broader tuning in the gap detection threshold curves was not found as the active and reference were separated. What was not

expected was the inconsistency across subjects - gap thresholds varied much more across subjects than across stimulation modes. The ANOVA on the data in figure 2.11 indicated a statistically significant difference between the tuning widths obtained for the three subjects, but no statistical difference between tuning widths for different stimulation modes.

Three possible explanations are proposed. The first is that current spread is already so large that the effect of using stimulation modes with widely separated electrodes has little additional effect (see Lim, Tong and Clark, 1985). A second explanation is that stimulation modes with larger electrode separation do not increase current spread as much as expected. The electrical field model of Finley et al. (1990) predicted a broadly spreading field around banded electrodes, such as those used in the Nucleus implant. Other electrode designs with more localized current distributions might produce more significant and consistent variations in gap detection thresholds in different stimulation modes. A third explanation is that the effects of changing stimulation mode were confounded by the fact that stimulation current level also changed with stimulation mode. Lower currents were used in the stimulation modes with larger spacing between the active and reference electrodes, because these modes produce lower thresholds and uncomfortable loudness levels. The original goal was to change the extent of the neural population activated by using more widely spaced electrodes in a stimulation pair. However, widely spaced electrodes in a stimulation pair result in lower stimulation currents, presumably exhibiting less current spread, and so may partially offset the increased neural extent due to the electrode separation. The net observed result was that sharpness of the gap threshold tuning curves remained effectively unchanged. So it is possible that the effect of stimulation modes with anticipated larger current spread was offset by using lower currents in these stimulation modes.

An additional unanticipated result is that in AR stimulation mode gap thresholds were not the very low values that should correspond to wide spread of stimulation. For wide spread of stimulation, all neural channels receive similar input and there are more channels to aid in gap detection, so a flat tuning curve with very low values of gap thresholds was expected for all separations of electrode pairs. In fact, although AR mode had a slightly flattened curve, the

curve was not entirely flat and in most instances the lowest gap threshold values were higher than the tip region of the other stimulation modes. Because true monopolar stimulation was not used, the current paths were directed towards the apical region (where the reference electrode was situated). When detecting the gap between, for example, stimuli on electrode pair (10,22) and electrode pair (19,22), gap detection is performed for an electrode pair with wide current spread activating a large neural population and an electrode pair activating a subset of the same neural population. This complete neural overlap would have been expected to produce low values of gap thresholds. This observation again suggests that the amount of neural overlap is not the only factor in determining the gap detection thresholds. Perceptual dissimilarity between the two stimuli may have confounded the peripheral gap detection mechanism. Our simple conceptual model about this peripheral mechanism cannot explain all the data presented here, and some of the trends might be ascribed to central mechanisms. This is the interpretation in Chatterjee et al. (1998), who found short gap thresholds only when the two stimuli bounding the gap were identical; gap thresholds were long if the two stimuli were perceptually different in any way (pitch or loudness). This result suggests that even if the two stimuli marking the gap activate mostly overlapping neuron populations, the differences in neuron activation may produce a sufficiently different percept that the gap detection decision is primarily central. However, the relative importance of central and peripheral processing mechanisms is unknown.

4.4 Channel characteristics

Gap detection threshold was employed as a tool to provide more insight into the channel characteristics, i.e. the number of channels, the location of channels, the width of channels and the factors that determine channel characteristics. The present results show that the gap detection tuning curves are wider for some choices of electrodes and stimulation modes, and also vary widely across subjects. However, it is not clear at this time how to interpret the gap detection tuning curves in terms of information channels. The following two assumptions are proposed to assist in defining channel width:

(1) Each subject has a minimum gap threshold when the two stimuli are presented to the

same electrode and a large gap threshold when the two electrodes are widely separated. *It was initially assumed that the value of the gap threshold corresponds to the relative amount of neural interaction.* Thus, for each electrode for each subject, there exists a value of gap threshold relative to the minimum such that larger values of gap threshold correspond to negligible channel interaction.

(2) *It is then necessary to make an assumption as to how much neural interaction is negligible*, i.e. how much interaction can be tolerated between two neural channels for them to still be distinct channels. This value is unknown, but it might correspond to the "shoulder" of the gap threshold "tuning curves". It is speculated that this shoulder indicates the point of transition from a peripherally-limited task to a centrally limited task (figure 2.12). As a first approximation a fixed gap threshold value at 40% between the lowest and highest gap threshold can be used, which is in the general vicinity of the shoulder of the gap threshold tuning curves. Another candidate measure for deciding whether two electrode pairs correspond to two different channels, is electrode discriminability. As the electrode pairs are separated and the amount of neural overlap decreases, the stimuli become easier to discriminate. The electrode separation at a chosen level of electrode discrimination (say 75% correct) may be used to find the corresponding gap threshold (from the gap threshold tuning curves). This defines a minimum gap threshold value for electrodes to be discriminable. Gap thresholds larger than this value correspond to two electrodes constituting two different channels.

Electrode discriminability was measured in these same subjects (Hanekom and Shannon, 1996) as described in Appendix 2.A. The 40% measure does not match very well with the electrode discriminability measure. The electrode discrimination measure may be too strict. Two electrodes might be discriminable if there is any difference in the neural populations that they stimulate, but that might not be enough difference to allow them to be independent information channels. Better choices than the 40% measure may be available, but this gives an example of how channels may be defined and this measure was used in the discussion that follows.

Using these assumptions, the following deductions emerge from the results.

(1) Number of channels

The number of distinct channels may be small. Estimating the number of channels from the above assumptions, it is found that N3 (a relatively poor user) may have only a few information channels available, and the upper limit in the number of available channels may be around six or seven (for N4, the best user in the group). Clearly, the number of channels will generally be less than the number of electrodes. One key question raised by this observation is whether improvements in speech recognition can be achieved by selecting electrodes appropriately (Zwolan et al., 1997; Henry et al., 1997; Hanekom and Shannon, 1996). Specifically, can better speech recognition be achieved with a smaller number of independent electrodes or a larger number of interacting electrodes?

When speech processors are programmed with a subset of the total number of available electrodes, many combinations of electrodes are available from which to select. Because of the pattern of interactions in an individual subject, processors with the same number of electrodes can be selected that will have quite different numbers of independent channels (Hanekom and Shannon, 1996). Hanekom and Shannon (1997), using fourteen different seven-electrode processors, made a very simple estimation of the number of channels using gap detection thresholds and the assumptions above and found significant correlation between vowel recognition and the estimated number of channels. Presumably, this relation was due to a clearer formant structure when a larger number of distinct channels were available.

(2) Width and location of channels

The shapes of the gap threshold curves suggest that the tuning may be broad or that channels are relatively wide and typically span many electrodes. Channels become only slightly wider when using stimulation modes with widely spaced active and reference electrodes. Channels are in general wider for the poorest user (N3) and narrower for the best user (N4) in the group. Using the assumptions above, channel width may be between 2 electrodes for N4 (1.5 mm) and 14 electrodes for N3 (10.5 mm). The location of the best (most selective) channels may be

deduced directly from the gap detection curves.

(3) Factors determining the characteristics of channels

At least four physical factors determine channel characteristics: (1) electrode placement relative to the remaining nerve fibers; (2) electrode design, which determines the electrical field distribution (Finley et al., 1990), e.g. the Nucleus has a banded electrode design while the Clarion device (Schindler and Kessler, 1993) has a radial electrode placement; (3) nerve survival (Zimmerman et al., 1995); and (4) the current pathway that the stimulation current follows between the active and reference electrodes. Apart from these physical factors, channel characteristics may also be influenced by central auditory nervous system processing.

In existing implants we can only control the current pathway and current spread to some degree by choice of the stimulation electrode pair. Results reported here for the Nucleus device indicate very little difference between the gap threshold tuning curves for different stimulation electrode pair separations and larger variations across subjects. Electrode placement and nerve survival are fixed for an individual implant patient and so cannot be modified after surgery to achieve a larger number of channels.

Although much research has focused on the physical factors influencing cochlear implant user performance, the important question is how these effect the information actually *received*. It is proposed that more research needs to be concentrated on how the *channel capacity* depends on or is related to the physical aspects of cochlear stimulation (electrode design, electrode placement, electrical fields patterns and nerve survival patterns).

4.5 Implications for cochlear implants

4.5.1 Comparison of electrode designs

Different electrode designs exhibit different current spread characteristics (Finley et al., 1990). The results suggest that electrode designs cannot be compared by simply calculating which design produces the most localized current field. All the subjects in our study had the same electrode design, but large differences in selectivity were observed. Selectivity is not a linear function of either current spread or the spacing between the stimulation electrode pair.

4.5.2 Reduced electrode processors

In any nonsimultaneous delivery of biphasic pulses to a number of electrodes there is an inherent trade-off between the number of electrodes and the overall pulse rate. As the number of electrodes decreases a higher pulse rate can be maintained on each electrode. However, the trade-off between pulse rate and number of electrodes is not well understood in terms of their importance to speech recognition. Several recent studies (Fishman et al., 1997; Lawson et al., 1993, 1996) suggest that implant patients are not making full use of all electrodes. It is possible that better speech performance could be achieved with a smaller number of electrodes, selected to be maximally independent channels, that are stimulated at a higher pulse rate. Techniques such as gap detection, forward masking (Shannon, 1983, Lim et al., 1989; Chatterjee and Shannon, 1998), electrode discrimination (Hanekom and Shannon, 1996; Kileny et al., 1997; Zwolan et al., 1997; Henry et al.,1997), or loudness summation (Fu et al., 1996) could be used to help select electrodes for inclusion or exclusion in a processor that uses only a subset of all available electrodes.

4.5.3 Choice of electrodes for a reduced electrode processor

Gap detection thresholds may also be used to compare different choices of speech processors regarding the number of channels, using the assumptions mentioned earlier. Although the actual number of channels is unknown, this method could be used to find the speech processor that maximizes the number of calculated channels. As discussed earlier, Hanekom and Shannon (1997) found that seven electrodes in a processor can lead to a quite different number of

distinct channels depending on which electrodes are chosen. Thus, a simple relationship between number of electrodes and quality of speech recognition cannot be assumed.

4.5.4 Choice of electrode pair separation (stimulation mode).

It has been widely assumed that closely spaced bipolar electrodes are necessary for achieving good spatial selectivity in a cochlear implant. However, the present gap detection tuning curves show only minor differences in spatial selectivity as a function of the separation of the bipolar pair. Indeed, in a recent study electrode discrimination and speech recognition were each similar for monopolar and bipolar stimulation (Zwolan, Kileny et al., 1997). To the extent that stimulation mode does effect channel interaction, the optimal configuration may change from one end of the electrode array to the other. This may result in the use of multi-mode speech processors, with each information channel optimized by choosing the electrodes and stimulation modes that result in the best selectivity. Present clinical speech processor fitting software for the Nucleus device allows mixed-mode processor designs, but this feature is not generally used.

5 CONCLUSIONS

(1) Gap detection thresholds are a function of the physical separation of the electrode pairs used for the two stimuli that bound the gap. Gap thresholds increase from a minimum when the two stimuli are presented on the same electrode pair to a maximum when the two stimuli are presented on widely separated electrode pairs. This change may be due to a change-over from a peripheral, within-channel gap detection process for closely spaced electrode pairs to a central cross-channel process for widely spaced electrode pairs.

(2) When the two marker bursts are presented to the same electrode, gap detection thresholds are similar across subjects at 1-4 ms. Gap thresholds for widely separated electrodes vary considerably among subjects and may be related to speech recognition performance, with better implant users having lower gap thresholds in this condition.

(3) The area of neural activation by each electrode (as inferred from the width of the tip region of the gap detection tuning curves as a function of electrode pair separation) varies across subjects and across electrodes. For the three subjects in the present study, the better implant users exhibit sharper tuning, i.e., a smaller area of neural activation around each stimulation pair.

(4) Using stimulation modes with larger separation between active and reference electrodes has limited effect on spatial selectivity. AR stimulation mode, although presumably having larger current spread, has better neural selectivity than BP mode for some subjects. This implies that there is no fixed optimal stimulation mode, but that the optimal stimulation mode may vary across subjects and from one end of the electrode array to the other.

The research presented in this chapter was supported in part by the NIDCD (National Institute on Deafness and other Communication Disorders). The research was made possible by travel grants towards the first author by the University of Pretoria, South Africa and by the House Ear Institute, Los Angeles.

APPENDIX 2.A
SUMMARY OF ELECTRODE DISCRIMINATION STUDY

Electrode discriminability was measured in the same subjects that participated in the study described in this chapter. Details can be found in Hanekom and Shannon (1996), but as this publication may not be easily accessible, a brief description of the study is given here.

This electrode discrimination study determined the amount of electrode confusion with a pitch discrimination experiment. A place pitch ranking matrix (or electrode discrimination matrix) was compiled by using a very simple psychophysical procedure. Consecutive stimuli of 500 ms, separated by a brief quiet interval of 200 ms, were presented on two of the subject's electrodes. The subject's task was to judge which stimulus was higher pitched.

Each stimulus pair consisted of stimuli on two different electrodes stimulated in BP+1 mode. All stimuli were current-balanced biphasic pulses, positive phase first. Stimulation rate was

1000 pulses per second and the pulse phase duration was 200 microseconds. Stimuli were presented at a comfortable level of stimulation above 50%, but below 80% of the dynamic range of the subject. The stimuli were balanced for loudness to minimize confusions between loudness and pitch.

A computer program generated the appropriate stimuli and recorded the subject responses. The electrodes used for the stimulation pairs during the pitch discrimination experiment were completely randomized for each run. One run consisted of the presentation of all possible combinations of electrodes, in both orders of presentation, excluding comparisons of electrodes with themselves. Twenty runs were completed in BP+1 mode for each of the subjects, which gave a total of forty comparisons of each electrode with every other electrode.

Subject reaction, indicating which stimulus was judged to be higher-pitched, was recorded for each stimulation pair and compiled into a response matrix. The response matrix tabulated the number of times that the more basal electrode of a stimulation pair was judged to be higher pitched than the more apical (which would be the expected order based on the tonotopic organization of the cochlea). This matrix was then converted to a percentage correct matrix, under the assumption that a judgement of the more apical electrode to be higher pitched than the more basal in a specific stimulation pair, was an incorrect decision. This resulted in a lower-triangular matrix. This lower-triangular matrix was then converted to a matrix of d′ values. The d′s gave an indication of the perceptual distance between the stimuli.

Figures 4.34 and 4.36 summarize some of the results. It was found the three subjects required different distances between electrodes for two electrodes to be discriminable (d'>1). This distance (ΔE, the number of inter-electrode distances) was also found to be variable across the electrode array for all three subjects. N3 required electrodes to be far apart to be discriminable. Figure 4.36 suggests that electrodes 1 to 16 were not discriminable in N3's case. N4 generally required a ΔE of 2 (see figure 4.34), while N7 generally required a ΔE of 4 or more.

Chapter 3

MODELS OF GAP DETECTION IN ACOUSTIC HEARING

1 INTRODUCTION

Gap detection has often been used as a measure of temporal resolution in the auditory system, e.g. Penner (1977), Divenyi and Danner (1977), Fitzgibbons (1984), Formby and Forrest (1991). The gap detection task has traditionally been within-channel (i.e. gap markers were the same). Across-channel gap detection has been used in acoustic stimulation (Phillips et al., 1997) to investigate how temporal gaps between spectrally different sounds are detected, a situation that is typical in everyday speech understanding tasks. It has been hypothesized that across-channel gap detection thresholds reflect the extent of neural activation (Hanekom and Shannon, 1998; chapter 2). In both the acoustic and electrical stimulation cases, it has been found that gap thresholds increase as the gap markers are separated in frequency (Divenyi and Danner, 1977; Divenyi and Sachs, 1978; Phillips et al., 1997; Formby and Forrest, 1991; Formby, Sherlock, and Forrest, 1996) or presented on two electrode pairs with increasing spacing (Hanekom and Shannon, 1998). Examples of the resulting U-shaped gap detection "tuning curves" have been given in chapter 2.

A model of gap detection in acoustic hearing is created in this chapter, with the primary objective of investigating which underlying factors bring about the U-shaped gap detection curves found in across-channel gap detection tasks.

1.1 Models of gap detection in acoustic auditory stimulation

Previous models of auditory duration discrimination or gap detection have been presented by Creelman (1962), Formby, Sherlock, and Forrest (1996), Forrest and Formby (1996), and Heinz, Goldstein, and Formby (1996). Creelman (1962) modelled duration discrimination of

brief signals in a signal detection theoretical context. He hypothesized that an internal counter counted the number of spikes elicited by a brief signal to measure the signal duration. The spike train was assumed to be Poissonian, so that noise was present in the duration measurement, but the signal onset and offset were assumed to be known exactly. Divenyi and Danner (1977) expanded this model to include the effects of noise in the determination of the signal offset and onset in gap discrimination. These models cannot explain the U-shaped curves found in across-channel gap detection.

In a series of articles, Formby and co-workers (Formby et al., 1996; Forest and Formby, 1996; Heinz et al., 1996) described models for gap detection with sinusoidal markers that differed in frequency. They ascribed the increased gap thresholds in across-channel gap detection to peripheral filtering. They noted that listeners may enhance gap thresholds when permitted to improve signal-to-noise ratio by shifting the auditory filter centre frequency to midway between the two marker frequencies. In essence, their single channel model (Forrest and Formby, 1996) had a single auditory filter stage (including nonlinear compression to model stimulus transduction at the hair cell) that bandpass-filtered the marker frequencies. This auditory filter was centred between the two marker frequencies to optimize performance on the gap detection task. The simulation used the same 2IFC procedure as used with listeners. The filter outputs in the two intervals of the 2IFC procedure were compared and the interval with the largest ratio of maximum to minimum output (the max-min statistic) was chosen as the interval containing the gap. Simulated gap thresholds could be obtained by taking the average of several model runs. The model generated results similar to those of human listeners in a gap detection task, except that the asymptotic thresholds for large marker frequency separations could not be reproduced.

The multi-channel model of these authors (Heinz, Goldstein, and Formby, 1996) is similar to the single-channel model, but used multiple bandpass filters and calculated the max-min ratio for each channel. Each channel then decided which of the two intervals in the 2IFC procedure contained the gap, and the final decision was base on a majority vote. Only channels within a decision region centred between the two marker frequencies were included in the decision. The

model generated results similar to those of human listeners in a gap detection task.

1.2 Extension of previous models

The model described in this chapter extends the work of Formby and co-workers in two respects:

(1) Closed form equations are obtained for gap discrimination and gap detection thresholds. This makes the model easy to interpret and allows effortless investigation of the effects of model parameters.

(2) The model is developed to be more detailed and closer to the underlying neurophysiology, by using spike trains and spike train statistics to derive equations for gap thresholds. This is important, as spike train statistics are dramatically different in electrical stimulation of the auditory nerve. The models of Formby and co-workers cannot be used for prediction of gap thresholds in electric hearing.

Consequently, by using the statistics of the spike train (rather than the max-min statistic based on analogue waveforms as in Formby), modelling what happens in both electrical stimulation and acoustic stimulation is possible. The basic principle used in the model in this and the next chapter is that neural channels stimulated by gap markers that are more closely spaced (in frequency or in physical electrode position), result in an easier gap detection task so that gap thresholds decrease. Following the arguments of Formby and co-workers, it is assumed that the auditory system places a filter between the two areas stimulated by the two markers. The exact nature of this filter is unimportant. This may be an "attentional filter" as suggested by Divenyi and Danner (1977) and Phillips et al. (1997). The current model makes this notion explicit. As the markers move apart, the task of the gap detector becomes more difficult, because the difference in spike rate during the marker and within the gap becomes smaller.

Conceptually, the model presented in this chapter incorporates the following ideas. Two factors limit measurement of the duration of a temporal gap. Uncertainty exists about the times at which the edges of the gap occur, and when the edges are known (or have been estimated)

there is noise in the actual measurement of the duration of the gap. The origin of this noise is the Poissonian nature of the spike trains in the central auditory nervous system.

1.3 Objectives of this chapter

A model of gap detection in acoustic hearing is created in this chapter, with the objective of investigating which underlying factors bring about the U-shaped gap detection curves found in across-channel gap detection tasks. This model extends previous models as described above. As it is believed that the same underlying mechanisms operate in acoustic and electrical gap detection, the intention is to create a model that can also be expanded for the electrical gap detection situation (see chapter 4).

2 A MODEL FOR GAP DETECTION IN ACOUSTIC HEARING

2.1 Assumptions about the acoustically evoked spike train

At the auditory periphery, an estimate of when the edges of the gap occur may be influenced by three different situations. First, in acoustic stimulation, spike trains are random and Poisson-like for high frequencies (above 5 kHz). In this case, the positions of the edges of a gap will be the most difficult to estimate, and gap thresholds may have been expected to be the largest of the three cases. Second, for low frequencies (acoustic stimulation below 5000 Hz), spike trains are phase-locked to the cycles of the stimulus (e.g., Johnson (1980)). This may have been expected to improve the ability to judge the edges of the gap accurately. However, Shailer and Moore (1983) showed that a close correspondence exists between the auditory filter bandwidth and gap detection thresholds at low frequencies. Also, gap thresholds decrease for increasing frequencies (Fitzgibbons, 1983; Fitzgibbons, 1984). This effect has been ascribed to the decreasing auditory filter bandwidth at lower frequencies, which presumably leads to more ringing in these narrower auditory filters, and to inherent fluctuations in narrowband noise (Shailer and Moore, 1983; Shailer and Moore, 1987). The rate of fluctuation is determined

approximately by the reciprocal of the bandwidth of the filter. Smaller bandwidth results in slower fluctuations, that may be confused with the gap.

Third, for electrical stimulation, spikes are tightly phase-locked, especially to pulsatile stimulus waveforms (Javel, 1990). The only limitation in the correct judgement of the positions of gap edges is the small jitter in spike positions relative to a preferred latency. If gap detection was based on temporal information only, gap detection might have been expected to be more acute than in normal hearing, but results show that gap detection thresholds in cochlear implantees are similar to those measured in normal-hearing listeners (Shannon, 1989).

These observations suggest that temporal phase-locking does not play a major role in determining gap thresholds, but that spatial mechanisms (i.e. peripheral filtering) may play a more important role, consistent with the models of Formby and co-workers as cited above. Of course, peripheral filtering does not play a role in cochlear electrical stimulation, but the current distribution from the site of an electrode may play a role similar to an auditory filter. This idea is expanded in chapter 4.

Based on the foregoing observations, the use of Poisson processes in the gap detection model to model spike trains seems justifiable, especially for acoustic stimulation. Assuming that spike trains are Poisson processes simplifies the mathematical analysis.

2.2 Nerve fibre model

The nerve fibre model used in the acoustic gap detection models is described here. The gap detection model is based primarily on the rate response profile (called the λ-profile in this text, where λ is the average rate parameter in a Poisson process that describes neural spike train). The nerve fibre model has a significant influence on the λ-profile, which in turn influences the size of $\Delta\lambda$ when the gap marker frequencies are separated. $\Delta\lambda$ is the difference in spike rates during the gap and markers. The larger this rate difference, the more detectable the transition from marker to gap or gap to marker becomes.

2.2.1 Nature of the nerve fibre model

The fibre model described is not a biology-based model, but rather a black box model that incorporates the major spike train characteristics that are important in audition. The well-known Hodgkin-Huxley model, e.g. Kistler, Gerstner, and van Hemmen (1997), is an example of what is termed here a biology-based model. It incorporates the nerve fibre cell membrane dynamics to predict firing characteristics in response to stimuli. Neither the Hodgkin-Huxley model, nor the current model has any statistical characteristics. They always fire when threshold is reached, unlike biological models that incorporate sources of noise, e.g. Lecar and Nossal (1971).

The nerve fibre model used in this chapter is an average rate model, meaning that it models the average spike count characteristics, but not the instantaneous spike characteristics. For example, the fibre model does not incorporate phase-locking of spike trains to a preferred stimulus phase (see chapter 5). The nerve fibre model can be used for calculating λ-profiles and predicting average firing rates. Finally, the model is based on pure-tone stimulus data. The input to the fibre model is the frequency f and amplitude A of a pure tone stimulus. The output is the average firing rate at different locations in the cochlea.

2.2.2 Nerve fibre model equations

The nerve fibre model builds on a model by Colburn (1973). The model is straightforward and simple to interpret. The average rate for fibre *m* for a pure tone stimulus of frequency *f* is

$$\overline{r_m} = SR + MR \cdot g(A) \cdot H_m(f) . \tag{3.1}$$

In this equation, SR is the spontaneous rate of the fibre, MR is the maximum rate, A is the stimulus intensity in dB SPL, g(A) is a function that characterizes spike rate as a function of intensity (rate-intensity curve), and $H_m(f)$ is a tuning parameter (explained below). The function g(A) is

$$g(A) = \frac{1}{2}\left(1 + Erf\left(\frac{A - A_{thr}(m)}{\sqrt{2}\sigma}\right)\right) . \tag{3.2}$$

This is an integrated Gaussian that provides an s-shaped curve that can be fitted to typical rate-intensity data. The parameter σ controls the slope of the curve. $A_{thr}(m)$ is the threshold of fibre m in dB SPL. By definition, threshold is reached when the spike rate increases by 10% from SR. Thresholds of fibres excited by the stimulus (of frequency *f*) differ as fibres may be stimulated at a frequency different from their CF. Fibres stimulated at their CF have the lowest threshold, and the threshold increases as fibres are stimulated with frequencies increasingly different from CF. This is described by the tuning function $H_m(f)$. Thus $A_{thr}(m)$ in equation 3.2 is

$$A_{thr}(m) = A_{thr,CF} + A_{off} + 20\log H_m(f)\,. \tag{3.3}$$

$A_{thr}(m)$ is the threshold of a fibre at a specific stimulus frequency *f*, when the best frequency for that fibre is CF. $A_{thr,CF}$ is the threshold at CF in dB SPL. Typical values for $A_{thr,CF}$ are 0 to 30 dB SPL (as determined in cat by Shofner and Sachs, 1986). g(A) takes on values between 0 and 1 and is centred on A_{thr}, unless the offset parameter A_{off} is used, i.e., at A_{thr} there is already a 50% increase in spike rate. The offset parameter A_{off} is used to ensure that the spike rate has increased by 10% at A_{thr}.

For a fixed value, the parameter σ results in a fixed slope for the rate-intensity curve. However, data show that the slope changes for stimulus frequency *f* more distant from CF for frequencies above CF (Evans, 1975; also see Javel and Viemeister, 2000). A model for σ is then

$$\begin{aligned} &\text{for } f \le f_m: \quad \sigma = \sigma_0\,, \\ &\text{for } f > f_m: \quad \sigma = \sigma_0 + \left(\frac{f - f_m}{f_m}\right) \cdot k\,, \end{aligned} \tag{3.4}$$

where f_m is the CF of fibre *m*. Parameter *k* changes the slope as a function of $|f-f_m|$. To fit the data in Colburn (1973), σ_0=5 is used, while σ_0=6 and *k*=20 provides a reasonable fit to the data in Evans (1975).

The tuning function $H_m(f)$ in equations 3.1 and 3.3 characterizes the threshold of a fibre with

best frequency CF when the acoustic stimulus frequency is varied. Thus, it is (the inverse of) a normalized frequency tuning curve (maximum tuning is 1), so that 20 log H_m gives an approximation to the tuning curve, but has threshold at 0 dB. The actual threshold at CF is then specified explicitly in equation 3.3 as $A_{thr,CF}$. $H_m(f)$ is given by

$$\begin{aligned} H_m(f) &= \left(\frac{f}{f_m}\right)^{\alpha} \quad for\ f \le f_m \\ &= \left(\frac{f_m}{f}\right)^{2\alpha} \quad for\ f > f_m\ . \end{aligned} \tag{3.5}$$

The parameter α controls the rolloff of the tuning curve, and is given by

$$\begin{aligned} \alpha &= \alpha_0 \quad for\ f_m \le 800\ Hz \\ &= \alpha_0 \cdot \frac{f_m}{800} \quad for\ f_m > 800\ Hz\ . \end{aligned} \tag{3.6}$$

α_0 =4 provides a good fit to the AN data. Note that the rolloff becomes sharper at higher frequencies. These last two are the same equations as given in Colburn (1973). The tip part of the data is fitted well, but not the tail. For the current model, the tail is ignored.

The set of equations from 3.1 to 3.6 define the auditory nerve fibre model.

2.3 Cramer-Rao Lower Bound for the Poisson change-point problem

The objective of this section is to derive a simple equation for the gap threshold as a function of the difference in spike rates elicited by the pre-gap and post-gap markers. The Poissonian nature of the neural spike trains complicates the detection of the edges of a gap, as it makes a change in spike rate difficult to detect. The Cramer-Rao Lower Bound (CRLB) for the detection of the gap edges is derived for this problem. The CRLB gives the variance in the estimate of a (classical) optimal estimator.

The problem is to calculate the CRLB for the time of a step change in the rate of a Poisson

process. It is natural to suspect that the accuracy with which the change-point may be found depends on the difference in rates of the Poisson process before and after the step change, as well as the time that the detector has to observe the non-homogeneous Poisson process. The rate of change of the rate function (i.e. whether it is a step function or a ramp) should also influence the accuracy of the change-point detection.

As far as is known, the CRLB for the Poisson change-point problem has not been addressed in literature. Much literature exists on solving the change-point problem for Poisson processes (e.g.Bremaud (1981), Karr (1986), Davis (1976), Raftery and Akman (1986), West and Ogden (1994), Gal'chuk and Rozovskii (1971)), but the objective of all of these authors was to find an algorithm to detect the change-point in a given Poisson process, rather than addressing the question of how accurately this could be done.

A parallel problem is that of determining the change-point of a signal in Gaussian noise that has a step change in amplitude. The task in this case is to estimate the time of step change by observing the signal plus noise. Samples are taken at regular intervals, whereas in the Poisson problem no sampling is done, but the discrete or continuous point process is observed. This signal in Gaussian noise problem is solved in e.g. McDonough and Whalen (1995). It may be expected that the change point detection accuracy is dependent on the variance of the Gaussian noise, as well as the differences in signal amplitude before and after the step, and also the time that the detector has for observation before and after the step. The CRLB on this estimate has been calculated in Reza and Doroodchi (1996).

The CRLB for signals in Gaussian noise may then be used as an estimate of the bound for the Poisson problem. For a Poisson process, the average rate λ equals the variance, so a signal that jumps in amplitude from A_1 to A_2 at time τ, can be equated to a Poisson process that has a jump in rate from $A_1=\lambda_1$ to $A_2=\lambda_2$. The variance of the Gaussian noise is λ_1 before the jump and λ_2 after the jump. This argument redefines the Poisson problem as a signal in Gaussian noise problem.

To calculate the CRLB, a differentiable function is required to characterize the jump in rate of the Poisson process. A sigmoidal function may be used to model the change-point, instead of a step change. For a jump from average rate λ_1 to average rate λ_2 at a change-point time τ, the rate of the Poisson process is given by equation 3.7,

$$\lambda_t(\tau) = \lambda_1 + \frac{(\lambda_2 - \lambda_1)}{1 + e^{-\alpha(t-\tau)}} = \frac{\lambda_2 + \lambda_1 e^{-\alpha(t-\tau)}}{1 + e^{-\alpha(t-\tau)}}, \tag{3.7}$$

where α determines the rate of change from λ_1 to λ_2. As shown in equation 3.26, α is inversely proportional to Δt, where Δt is the duration of a transition from λ_1 to λ_2. Physical factors determine Δt, as the transition between marker and gap cannot occur instantaneously. The data of Zhang, Salvi, and Saunders (1990) show that spike rate decays exponentially after the offset of a gap marker with time constants in the order of 1 ms. Westerman and Smith (1984) measured minimum decay time constants of 1 ms in auditory nerve fibres. As shown below equation 3.6, this implies values in the order of α=4000.

The CRLB is $1/F(\theta)$ with $F(\theta)$ the Fisher information and θ a parameter. Snyder (1975) derived a general form for the Fisher information for Poisson processes,

$$F(\theta) = \int_{t_0}^{T} \left(\lambda_t(\theta)\right)^{-1} \left(\frac{\partial \lambda_t(\theta)}{\partial \theta}\right)^2 dt\,. \tag{3.8}$$

$F(\theta)$ is a function of the *intensity* λ_t of the Poisson process only. $[t_0, T]$ is the period over which the CRLB is required. For the signal model in equation 3.7, the Fisher information, parameterized by τ, may be rewritten as

$$F(\tau) = \int_{0}^{T} \left(\lambda_t(\tau)\right)^{-1} \left(\frac{\partial \lambda_t(\tau)}{\partial \tau}\right)^2 dt\,. \tag{3.9}$$

It follows from equation 3.7 that

$$\frac{\partial \lambda_t(\tau)}{\partial \tau} = \frac{\alpha e^{-\alpha(t-\tau)}(\lambda_1 - \lambda_2)}{(1 + e^{-\alpha(t-\tau)})^2}. \tag{3.10}$$

Accordingly, the Fisher information is

$$F(\tau) = \int_0^T \frac{\alpha^2 e^{-2\alpha(t-\tau)}(\lambda_1 - \lambda_2)^2}{(1 + e^{-\alpha(t-\tau)})^3(\lambda_1 e^{-\alpha(t-\tau)} + \lambda_2)} dt. \tag{3.11}$$

The integral in equation 3.11 has a solution in closed form, but this solution is a rather complicated expression that does not provide insight into the effect of the different variables (λ_1, λ_2 and α) on the accuracy with which τ can be estimated. Therefore, equation 3.11 is solved numerically. The CRLB gives the variance in the estimate of τ and is calculated as the inverse of equation 3.11.

One approach for obtaining a closed-form expression for the CRLB in terms of λ_1 and λ_2, is to approximate the Poisson process by a Gaussian process as explained above. An equivalence between the two processes exists when the Poisson process is interpreted as a discrete Poisson process. The discrete Poisson process with intensity λ_1 has an average of $\lambda_1 T$ points in a time interval T. The variance in the number of points in T is also $\lambda_1 T$. Accordingly, the discrete Poisson process is equivalent to a Gaussian process with a sampling period of T seconds, average $\lambda_1 T$ and variance $\lambda_1 T$. The problem of estimating the change-point in a Gaussian process has been solved by Reza and Doroodchi (1996). Their model is similar to equation 3.7, but they discretized the problem. In their formulation, the objective is to estimate the discrete sampling interval n_0 in which the change-point occurs. The solution is

$$CRLB = \frac{8}{(d')^2 \alpha T} \cdot \left(\sum_{n=0}^{N-1} J_k\right)^{-1} = var(\hat{n}_0), \tag{3.12}$$

where T is the sampling interval and N samples are observed. α is the rate of change parameter as before, d' is a parameter that measures the distance between the two pdfs, and J_k is defined

below. If the average value of the two Gaussian pdfs were m_1 and m_2 and the two distributions had the same variance σ,

$$d' = \frac{m_1 - m_2}{\sigma}. \tag{3.13}$$

d' may be interpreted as a signal-to-noise ratio. For $\sigma_1 \neq \sigma_2$ the average $\sigma^2 = (\sigma_1^2 + \sigma_2^2)/2$ is used. Substituting $\lambda_1 T$ for m_1 and $\lambda_2 T$ for m_2, $\lambda_1 T$ for σ_1^2 and $\lambda_2 T$ for σ_2^2, d' is found as

$$(d')^2 = \frac{2T(\lambda_1 - \lambda_2)^2}{\lambda_1 + \lambda_2}. \tag{3.14}$$

J_k in equation 3.12 is given by Reza and Doroodchi as

$$J_k = \tanh a_k - \tanh b_k - \frac{1}{3}(\tanh^3 a_k - \tanh^3 b_k), \tag{3.15}$$

with $a_k = \frac{\alpha T}{2}(k - n_0)$ and $b_k = a_k - \frac{\alpha T}{2}$.

When equation 3.14 is substituted into equation 3.12, a closed-form expression for the CRLB in terms of λ_1 and λ_2 is obtained as

$$var(\hat{n}_0) \geq \frac{4(\lambda_1 + \lambda_2)}{\alpha T^2(\lambda_2 - \lambda_1)^2} \cdot \left(\sum_{n=0}^{N-1} J_k\right)^{-1}. \tag{3.16}$$

The solution in equation 3.16 is an approximation to the actual CRLB that can be calculated from equation 3.11, but as figure 3.1 shows, the approximation is quite acceptable. The variance in equation 3.16 is in number of sampling intervals, so that $var(\hat{\tau}) = T^2 \cdot var(\hat{n}_0)$.

If multiple observations of the same signal is available, it is generally true that the signal-to-noise ratio (SNR) for estimating the signal improves. Thus if the same change in rate of the Poisson process is observed in M simultaneous but independent Poisson processes, the signal-to-noise ratio (SNR) improves as indicated in equation 3.17,

$$SNR_M = \sum_{i=1}^{M} SNR_i. \tag{3.17}$$

Generally, SNR_i will not be the same in different fibres that convey information about the same signal, but in this model the M fibres are within the same critical band (as explained in section 2.4) so that the SNR_i are of comparable magnitudes in these channels. To simplify calculations, it is assumed that the SNR_i are identical in the M channels, so that $SNR_M = M \cdot SNR$. For the CRLB calculation this implies that the minimum variance in the estimate of τ is obtained when the SNR is the signal-to-noise ratio of the channel with the largest difference between λ_1 and λ_2 (if it is assumed that λ_1 and λ_2 are not equal on all M channels).

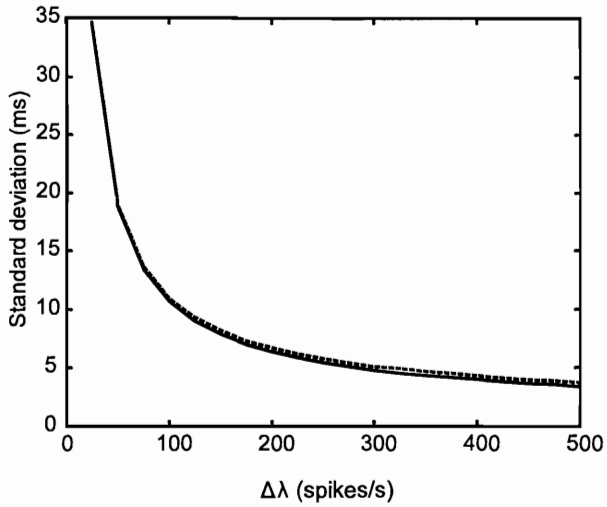

Figure 3.1

This figure shows the standard deviation in the estimate of the position of a gap edge, calculated from equation 3.11 (solid line) and equation 3.16 (dashes). The values calculated from equation 3.16 were multiplied by T to obtain the standard deviation in ms.

For the detection of a gap, λ_1 may be the rate during the marker, while λ_2 may be the rate during the gap. For within-channel gap detection, the task is to detect the change in rate, and not to estimate the time of change. Thus, the probability P of detection as a function of $\Delta\lambda$,

where $\Delta\lambda = |\lambda_2 - \lambda_1|$, is required. The gap threshold is assumed to be at P = 0.76. For across-channel gap detection, the task is to discriminate between gap lengths in the two intervals.

The final equation for the variance in the estimate of τ is therefore

$$var(\hat{\tau}) = \frac{T^2 \cdot 4(\lambda_1 + \lambda_2)}{\alpha T^2(\lambda_2 - \lambda_1)^2} \cdot \frac{1}{\sum_{n=0}^{N-1} J_k} \cdot \frac{1}{M}$$
$$= \frac{4(\lambda_1 + \lambda_2)}{\alpha(\lambda_2 - \lambda_1)^2} \cdot \frac{1}{\sum_{n=0}^{N-1} J_k} \cdot \frac{1}{M}. \qquad (3.18)$$

The summation in equation 3.18 does not lend itself to physical interpretation. Arriving at an even simpler form of equation 3.18 by the following derivation is possible. Suppose the task is to estimate the time of change τ of the average value of a Gaussian process (figure 3.2), from a value A_1 to a value A_2. The Gaussian process has a standard deviation σ_y. Let the change take place within Δt seconds. The task is to estimate the time of change τ at the centre of the transition. This task is the same as estimating the time at which the Gaussian process has an average value exactly halfway between A_1 and A_2. Because of the Gaussian noise in measurement of σ_y, there is noise in the estimate of τ. Thus, the Gaussian variable on the y-axis is transformed to another Gaussian variable on the t-axis. The average value of this Gaussian is τ, while the standard deviation is σ_t. It can easily be shown that

$$\sigma_t = \frac{\Delta t}{A_2 - A_1}\sigma_y. \qquad (3.19)$$

This is the standard deviation in the estimate of τ when one sample is taken at the time when the average value of the Gaussian is $(A_1+A_2)/2$. It follows for M samples that

$$var(\hat{\tau}) = \Delta t^2 \cdot \frac{\sigma_y^2}{(A_2 - A_1)^2} \cdot \frac{1}{M} = \frac{\Delta t^2}{SNR} \cdot \frac{1}{M}. \tag{3.20}$$

If, as before, the Poisson process is discretized, SNR may be substituted by equation 3.14 to obtain

$$var(\hat{\tau}) = \frac{\Delta t^2 \cdot (\lambda_1 + \lambda_2)}{2T(\lambda_2 - \lambda_1)^2} \cdot \frac{1}{M}. \tag{3.21}$$

Therefore,

$$std\ dev(\tau) = \sqrt{\frac{\Delta t^2(\lambda_1 + \lambda_2)}{2T(\lambda_2 - \lambda_1)^2} \cdot \frac{1}{M}}. \tag{3.22}$$

The equivalence between equations 3.18 and 3.22 is clear. Note that while equation 3.18 gives the minimum standard deviation predicted by the CRLB, equation 3.22 is the actual standard deviation of an estimate formed as described above. For $\alpha >> 1$, (for rapid transitions) the summation term in equation 3.18 saturates at a value of 4/3. It will be shown below that α will always be large in these calculations. Equation 3.18 simplifies to

$$var(\hat{\tau}) = \frac{3(\lambda_1 + \lambda_2)}{\alpha(\lambda_2 - \lambda_1)^2} \cdot \frac{1}{M}. \tag{3.23}$$

Comparison of equation 3.21 and equation 3.23 suggests that an equivalent α may be found for a transition duration of Δt seconds. If the slope of a sigmoidal function at $t=\tau$ in equation 3.7 is equated to the slope of the transition in figure 3.2, α can be solved for in terms of Δt. The slope is

$$\frac{\partial \lambda_t(\tau)}{\partial t} = \frac{\alpha (A_2 - A_1) e^{-\alpha (t-\tau)}}{(1 + e^{-\alpha (t-\tau)})^2} . \tag{3.24}$$

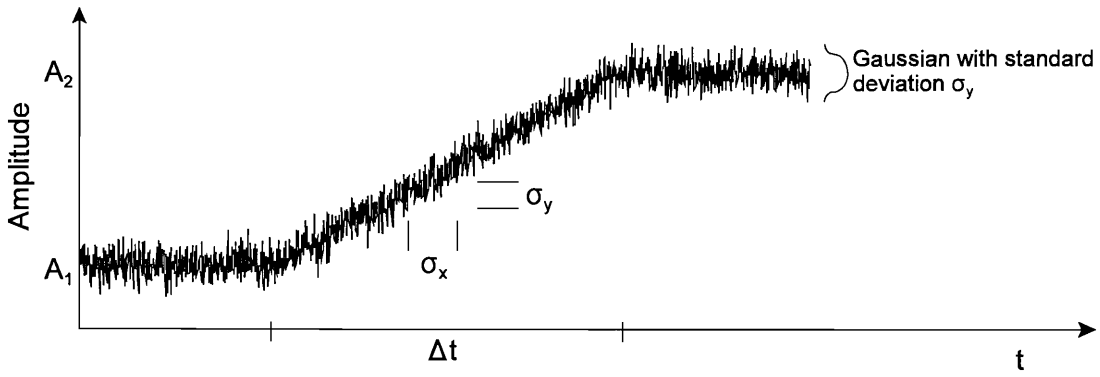

Figure 3.2

The Poisson change-point problem is re-interpreted as a Gaussian change-point problem, which is shown in this figure. The transition from average value A_1 to average value A_2 takes Δt seconds.

Replacing t with τ and equating this with the slope of the transition in figure 3.2,

$$\frac{A_2 - A_1}{\Delta t} = \frac{\alpha (A_2 - A_1)}{4} , \tag{3.25}$$

so that

$$\alpha = \frac{4}{\Delta t} . \tag{3.26}$$

For example, if calculation resolution is 1 ms, the fastest transition takes place in 1 ms. Thus, Δt=1 ms and α=4000. This shows that large α is always used, which in turn shows that the approximations in equation 3.23 are correct. The choice of Δt=1 ms is also supported by the auditory nerve fibre decay time constant data discussed after equation 3.7.

Using equation 3.26 in equation 3.23, it is found that

$$var(\hat{\tau}) = \frac{3\Delta t}{4} \cdot \frac{\lambda_1 + \lambda_2}{(\lambda_2 - \lambda_1)^2} \cdot \frac{1}{M}. \tag{3.27}$$

If, in equation 3.21, the transition takes place in one sampling period, then T=Δt, so that the variance in the estimate of τ predicted by equation 3.27 is 1.5 times the variance predicted by equation 3.11. However, it is merely the nature of the transition model (ramp in equation 3.19 or sigmoidal in equation 3.7) which results in this difference. The sigmoidal function with slope as calculated in equation 3.26 does not quite complete a transition within one sampling period. When α= 6/Δt, the transition is completed within one sampling period. With this α in equation 3.23, it is seen that equation 3.21 equals equation 3.23. Thus the estimate is efficient and the variance in estimate is given by the simplified version of equation 3.21,

$$var(\hat{\tau}) = \frac{\Delta t \cdot (\lambda_1 + \lambda_2)}{2(\lambda_2 - \lambda_1)^2} \cdot \frac{1}{M}. \tag{3.28}$$

Equation 3.28 is a simple closed form expression for the variance in the estimate of the change point. This equation can be used to predict the standard deviation in the estimate of the time of transition from an average rate λ_1 to a new average rate λ_2 in a Poisson process, as a function of $\Delta\lambda$, the difference between the two rates.

2.4 Bounds on the gap detection and discrimination thresholds as a function of ΔF

For the gap detection or gap discrimination problem, the variance in the estimate of the time of transition is required as a function of ΔF, the difference in frequency between the two markers that mark the gap. As it is known from the derivations above how $\Delta\lambda$ influences the accuracy with which a change-point can be located, a relation between ΔF and $\Delta\lambda$ is required.

It is assumed that the auditory system places a filter halfway between the two marker frequencies F_1 and F_2. This "filter" is just an "attention window" or observation window and

does not apply any weighting. The width of the window is assumed to be one critical band. When F_1 and F_2 are closely spaced (closer than one critical band), the largest part of the activation pattern of each of the marker frequencies falls within the attention window. A central integration centre that observes inputs from fibres within this observation window will see large differences in firing rate when the fibres are excited by the presence of a marker tone, as opposed to the low firing rate during the gap. Not all the fibres fire at the same rate, of course. The central integration centre will see a firing rate profile. The data of Zhang, Salvi, and Saunders (1990) show that discharge rate may be suppressed below SR during the gap. The rate during the gap is assumed to be SR or lower. The rate profiles for the pre-gap and post-gap markers will be similar, but not equal (figure 3.3). As the marker frequencies are separated (ΔF increases) the observation window will eventually contain just the tails of the

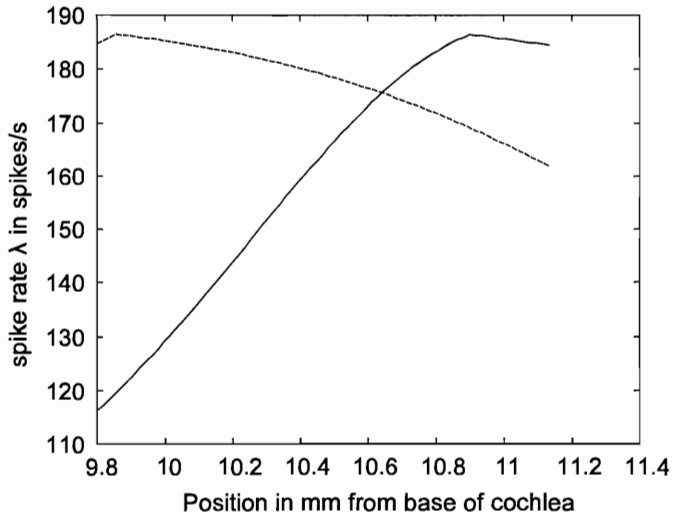

Figure 3.3

Spatial excitation profiles of two marker frequencies are shown when observed in a window one critical band wide (midway between the two sites of maximum excitation). The solid line is for excitation by the first marker of 500 Hz, and the dashed line is for excitation by the second marker of 600 Hz.

activation patterns. Little or no difference in firing rate between the gap and marker conditions exists for this situation, so that the transition between the gap and marker will be difficult to

detect. Using the model in paragraph 2.1 to obtain the firing profile, different values $\lambda_1^{(i)}$ for each fibre in the observation window during the marker is obtained, while it is assumed that a fixed λ_2 exists on all fibres during the gap. λ_1 is the rate during a marker, while the superscript *(i)* indexes the fibre number.

To combine information from the spike trains in the observation window, one option is to calculate the SNR on each fibre and then to add the SNRs (see equation 3.17; Green and Swets, 1966, equation 9.1) to obtain a combined SNR for N channels. This can be substituted into equation 3.20 to obtain a new equation 3.21. Or, to simplify calculations, the maximum λ_1 could be used. This will give the best possible SNR and thus the best possible performance. The effect of N fibres can then be added by multiplying the SNR by N. This method does not take the poorer signal-to-noise ratio on other fibres into account. The first option was used in calculations. This was used in equation 3.21 to calculate the variance in the estimate of τ, the time of the transition between the marker and the gap.

It was further assumed in calculations that the two markers had the same average value of λ_1, and that both transitions (from marker to gap and from gap to marker) were of equal difficulty to estimate. If the noise in measuring the exact time of transitions were the only source of noise in estimating gap length, then the variance in the estimate of the gap length is the sum of the variances of the estimates of the two transition times. Following Siebert (1970), the standard deviation in the estimate of gap duration can be equated to the gap duration discrimination threshold jnd_{gap},

$$jnd_{gap} = \sqrt{2\,var(\hat{\tau})}\,. \qquad (3.29)$$

$Var(\hat{\tau})$ is calculated from equation 3.21. Equation 3.29 is for a single fibre and is valid under the assumption that the two random variables indicating the two transitions on either side of the gap are independent.

Figure 3.4 shows data sets for gap detection (Formby et al., 1996) and gap discrimination

(Divenyi and Danner, 1977) along with predictions obtained from equation 3.29 and other predictions to be described in the sequel.

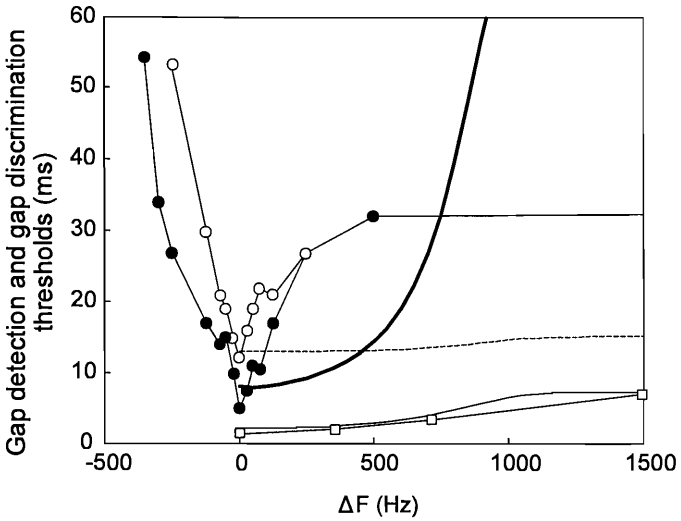

Figure 3.4

Gap detection and discrimination data and model predictions for gap discrimination are shown. Circles are gap detection data from Formby et al. (1996). The standard (first marker) was at 500 Hz in these data. Open circles are for a signal-to-noise ratio of 40 dB with noise presented at 10 dB SL, while closed circles are for the same signal-to-noise ratio, but presented at a louder level (noise at 30 dB SL). Squares are gap discrimination data from Divenyi and Danner (1977). Model predictions have been obtained only for F_2 (post-gap marker frequency) higher than F_1 (pre-gap marker frequency). The thin solid line and dashed line are model predictions for gap discrimination, while the bold solid line is a prediction for gap detection. The dashed line for gap discrimination was obtained when the Poissonian timer was implemented, and the solid line was obtained without the Poissonian timer. Model parameters were: fibre threshold = 45 dB, k=20, σ_0=5, SR=35 spikes/s, maximum spike rate = 200 spikes/s.

2.5 Measurement of gap duration with a Poissonian timer

It is possible that the internal time measurement is also noisy, i.e. when the only source of noise is not only the noisy measurement of gap edges, but also imperfect measurement of the time between the two estimates of gap edges. Creelman (1962) and Divenyi and Danner (1977) addressed the problem of measuring the gap duration with a timing mechanism that measured the gap duration by counting the number of spikes of a Poisson spike train when the gap edge positions are known. In this document, this mechanism of measuring gap duration is referred to as a "Poissonian timer".

With the gap edge positions known, but measurement of gap duration is done with a Poissonian timer, d' is (Divenyi and Danner, 1977)

$$d' = \sqrt{SNR} = \frac{\sqrt{2}(\mu_{t+\Delta t} - \mu_t)}{\sqrt{\sigma^2_{t+\Delta t} + \sigma^2_t}} . \tag{3.30}$$

μ_t is the average number of spikes during a gap with duration t. For a Poisson process, $\mu_t=\lambda.t$, with λ the rate during the gap, $\mu_{t+\Delta t}=\lambda.(t+\Delta t)$, and the spike count variances are $\sigma_t^2=\lambda.t$ and $\sigma_{t+\Delta t}^2=\lambda.(t+\Delta t)$. The jnd for gap duration is Δt.

The average number of spikes during a gap of duration t is still the same when the gap edges cannot be determined accurately, but the variance in the number of spikes increases. If σ_τ^2 is the variance in the estimate of the gap edge time, the additional number of spikes for one uncertain edge is $\sigma_\tau.\lambda$. Thus, the total variance in the spike count when estimating gap duration is

$$var(spike\ count) = 2 \cdot (\sigma_\tau \lambda)^2 + \sigma_t^2 . \tag{3.31}$$

Then, substituting equation 3.31 in equation 3.30, d' for gap duration is

$$d' = \frac{\sqrt{2}(\mu_{t+\Delta t} - \mu_t)}{\sqrt{\sigma^2_{t+\Delta t} + 4\sigma^2_\tau \lambda^2 + \sigma^2_t}}, \tag{3.32}$$

which is an equation also given in Divenyi and Danner (1977).

The gap duration jnd can be solved for from equation 3.32 by setting d'=1 (this is by definition the discrimination threshold; at d'=1 the percentage of correct decisions is 76%). μ_t and σ_t^2 are substituted by $\lambda.t$, and $\mu_{t+\Delta t}$ and $\sigma^2_{t+\Delta t}$ by $\lambda.(t+\Delta t)$ and then equation 3.32 is solved for Δt, where Δt is the jnd in gap duration,

$$\Delta t = \frac{1}{4\lambda} + \frac{1}{4\lambda}\sqrt{1 + 16\lambda t + 32\sigma^2_\tau \lambda^2}. \tag{3.33}$$

The variable t is the base duration for gap discrimination in this equation, and λ is the average spike rate that the Poissonian timer uses to measure the gap duration. By assumption, this rate is different from the spike rate during the gap, but is the spike rate used by a Poissonian timer somewhere in the central auditory system. To use this equation, σ_τ^2 is required. This is the variance in the estimate of the transition from gap to marker as given in equation 3.29. Figure 3.4 shows the effect on gap thresholds of using a Poissonian timer.

2.6 Gap detection

The derivation so far is valid for a gap discrimination task. For gap detection, the task is quite different, and in this section equations are derived for gap detection thresholds.

In the gap detection task, the listener has to detect the presence of a gap, but does not have to estimate the gap duration. If both markers have the same frequency (the within-channel condition), the task is to detect a transition. If the gap markers differ in frequency (the across-channel condition), there will be a transition whether a gap is present or not. To detect the gap in a 2IFC paradigm, the listener will have to decide which interval has two transitions. Accordingly, two transitions must be detected. The detection of a transition depends on both

the rate difference between two Poisson processes (during the gap and markers), and the time that the detector has to observe the Poisson process before and after the transition. If the gap is too brief, the change in spike rate after the first gap edge will not be detected, and neither will the change in spike rate at the end of the gap. Only one transition will be detected if the markers differ in frequency.

In other words, if the gap is too brief, the listener does not discriminate between the Poisson spike rate during the marker and during the gap. If the difference in spike rate is not discriminated, the gap cannot be detected. Therefore, the central detection mechanism has to obtain an estimate of the spike rate.

Spike rate estimates can be obtained by an optimal nonlinear estimator as described in chapter 5, or sub-optimal estimates may be obtained from spike counts. The nonlinear estimator described in chapter 5 uses the spike train as input and estimates the spike rate of that specific spike train. It bases estimates on a prior model of the expected signal. To obtain statistical information (e.g. variance in estimate), Monte Carlo analysis is required.

On the other hand, statistical methods can be used to evaluate the performance of a classical estimator. In this chapter, the central gap detection mechanism is hypothesized to discriminate spike trains based on the sub-optimal statistic of spike counts. If two Poisson processes that differ in rate have to be discriminated based on the number of spikes in an interval T, the signal-to-noise ratio for the discrimination task is (Rieke et al., 1997; see also equation 3.14).

$$SNR = \frac{T \cdot (\Delta\lambda)^2}{\lambda_{avg}} = \frac{2T(\lambda_2 - \lambda_1)^2}{\lambda_1 + \lambda_2}. \tag{3.34}$$

In this equation, λ_{avg} is the average λ and T is the total observation interval of the Poisson process with rate of either λ_1 (during a marker) or λ_2 (during a gap). As the critical interval for discrimination between the two spike rates is the gap (because it is brief), T can be equated to

the gap duration. In other words, the assumption is that if an observation period of T seconds is too brief to discriminate between two rates λ_1 and λ_2, the gap will not be detected. If T is long enough for discrimination of the spike rate during the first marker from the gap spike rate, and also for discrimination of the gap spike rate from the second marker spike rate, the gap will be detected. If it is assumed that both marker stimuli elicit the same spike rate, and that the (lower) spike rate during the gap is stationary, detection of both transitions are equally likely.

The threshold value for the detection of the gap is where P_d=P(gap detected)=0.76 (i.e., when d'=1 in a 2IFC paradigm). For the detection of a gap,

$$P(\textit{gap detect}) = P(\textit{detect first transition}) \cdot P(\textit{detect second transition}) . \qquad (3.35)$$

For achieving a detection probability of P_d=0.76, the probability of detection for either transition has to be 0.87 ($0.87^2 = 0.76$), if it is assumed that the detection of both markers is equally likely.

The detectability d' required for P=0.87 is then calculated from

$$P(\textit{detection of transition}) = 1 - Erfc\left(\frac{\sqrt{2}\ d'}{2}\right) = 0.87 , \qquad (3.36)$$

which is valid for a 2IFC experimental paradigm.

For P=0.87, it is found that d'≈1.6. This value for d' is substituted into equation 3.34, keeping in mind that d' = $\sqrt{}$SNR. Solving for T,

$$T = \frac{1.6^2\ (\lambda_1 + \lambda_2)}{2(\lambda_2 - \lambda_1)^2} . \qquad (3.37)$$

This is the final equation for the gap detection threshold, where T is the gap duration, and λ_1 and λ_2 are the spike rates during the marker and gap respectively. For comparison with the gap discrimination threshold, equation 3.22 is repeated here,

$$std\ dev(\tau) = \sqrt{\frac{\Delta t^2(\lambda_1 + \lambda_2)}{2T(\lambda_2 - \lambda_1)^2} \cdot \frac{1}{M}}. \qquad (3.22)$$

Note that in equation 3.37 the quantity $(\lambda_1 + \lambda_2)/2(\lambda_2 - \lambda_1)^2$ appears, while the square root of this same quantity appears in equation 3.22. This characterizes the difference between gap detection and discrimination. This is equivalent to the situation of intensity discrimination, where it is well known that for detection, data can be fit by a normal integral of A^2 (where A is the stimulus intensity), while for discrimination the data is fit by a normal integral of ΔA (e.g., Laming (1986)).

As before, values of $\Delta\lambda$ correspond to a particular ΔF, so that curves for the gap detection threshold as a function of ΔF can be plotted. Examples of gap detection threshold predictions are shown in figures 3.4-3.7 (discussed below).

3 RESULTS

Figures 3.4-3.7 show gap detection data (Formby, Sherlock, and Forrest, 1996) and gap discrimination data (Divenyi and Danner, 1977) for acoustic stimulation and listeners with normal hearing, along with model predictions. The Formby data were measured for F_1=500 Hz and a variable F_2. Two data sets are shown, both measured at a SNR of 40 dB, but the data set with the lower gap threshold was measured at higher intensity (the noise level was 30 dB SPL instead of 10 dB SPL). The data show that discrimination of gap duration is an easier task than gap detection under the conditions of these experiments (where noise complicated the gap detection task). Possibly, the presence of the base duration makes this task simpler than gap detection.

Model predictions for various parameter choices are shown in figures 3.4 to 3.7. Model parameters are described in the figure captions. Three model curves are shown in each figure, one for gap detection thresholds (equation 3.37), one for gap duration discrimination when the

only source of noise is the uncertainty in gap edge positions (equation 3.22), and one for gap duration discrimination when the gap duration is measured with a Poissonian timer (equation 3.33 together with equation 3.21).

Overall, the trends shown by the gap detection model and the gap duration discrimination model parallel those observed in the data. The gap detection curves have steeper slopes than the gap duration discrimination curves. The latter curves are wide, bowl-shaped curves. However, the predicted gap detection thresholds are lower than the gap discrimination threshold for within-channel gap detection ($\Delta F=0$) when the Poissonian timer is used.

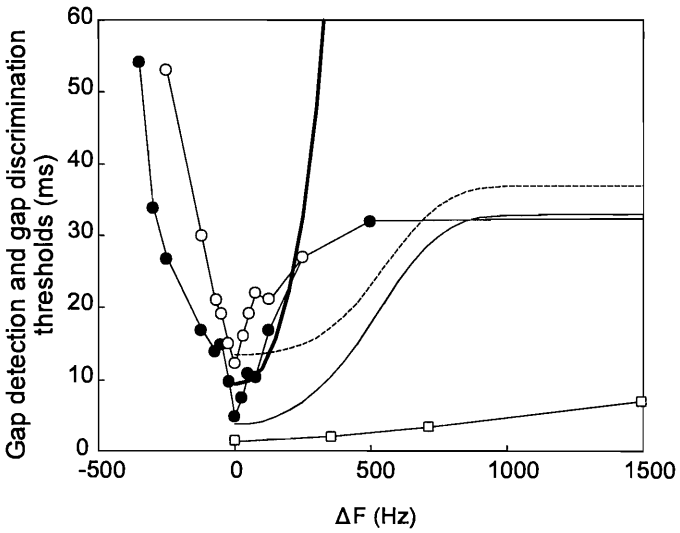

Figure 3.5

Same as figure 3.4, but with different model parameters: fibre threshold = 55 dB, k=20, σ_0=5, SR=25 spikes/s, maximum spike rate = 200 spikes/s. The thin solid line and dashed line are model predictions for gap discrimination, while the bold solid line is a prediction for gap detection.

For the parameter choices in figure 3.5, the tip region of the gap detection data is predicted by the gap detection model, while the tails of the gap detection data are predicted by the gap discrimination model. However, neither model predicts the gap discrimination data. Note that for both the gap detection and gap discrimination models, the same fibre parameters were used. It is, however, conceivable that gap detection and gap discrimination are performed by fibres

with different parameters.

The model predictions in figure 3.6 are interesting. For these parameter values, the gap detection model predicts the gap detection data, while the gap discrimination model with the same parameters predicts gap duration discrimination thresholds in the correct range, although the wide-bowl shaped curve is not predicted. Considering the gap discrimination model where the Poissonian timer is used, it is observed that the sharp peak in the electrical stimulation data (see figures 2.2, 2.3 and 2.12) is predicted by the gap detection model, while the wide bowl-shaped part of the electrical stimulation data is predicted by the gap discrimination model. There is a change-over point at a gap duration of 13 ms, so it seems possible that at this duration the task changes from detection to discrimination. This is at a value of ΔF of 100 to 150 Hz, which is approximately one critical band at a marker frequency of 500 Hz. Interestingly, this is supported by pitch discrimination data for electrical stimulation (Hanekom and Shannon ,1996), where electrode sounds generally become distinct when electrode pairs are 1.5 mm apart (approximately one critical band).

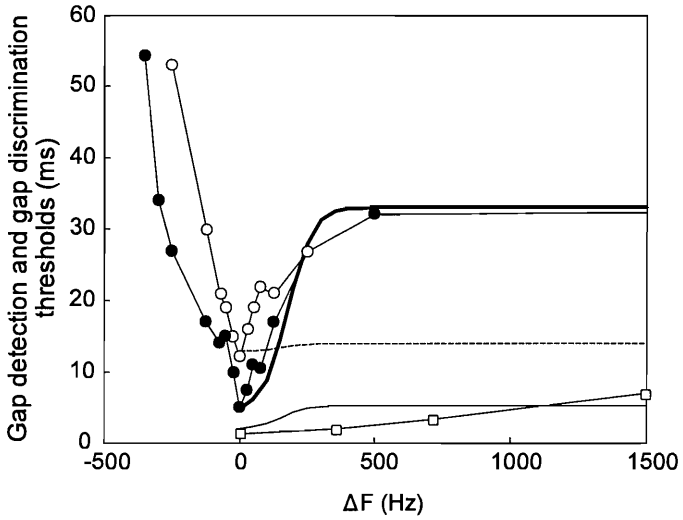

Figure 3.6

Same as figure 3.4, but with different model parameters: fibre threshold = 55 dB, k=20, σ_0=2, SR=35 spikes/s, maximum spike rate = 200 spikes/s.

If the resolution is 1 ms, and if α is chosen as $4/\Delta t$, then α=4000. With this choice and with fibre threshold set at 45 dB in the gap discrimination model, it is seen in figure 3.4 that a

reasonable fit is obtained to the gap discrimination data. This suggests the possibility that the gap detection thresholds and gap discrimination thresholds are not determined by the same population of nerve fibres. For these parameter choices, it is also observed that only one fibre is required to achieve the gap discrimination data if gap duration measurement is perfect, but if the measurement is done with a Poissonian timer, more fibres are required. Other parameter choices also need to be changed to predict the data in this case. Specifically, the base duration for discrimination has to be 0 to achieve low enough gap discrimination values with the Poissonian timer. Equation 3.33 shows that gap thresholds may be decreased by decreasing the base duration, or by increasing λ, the rate of the Poissonian timer.

In some of these simulations, the neural threshold was set at 55 dB. This is an unrealistic choice. This value is required to achieve a sharper cutoff in λ_2 (i.e. a narrower region of activation). This means that after the auditory nerve, an additional system must exist which sharpens neural tuning, or, equivalently, inhibits neural firing when the frequency difference between CF and the stimulus frequency is large enough.

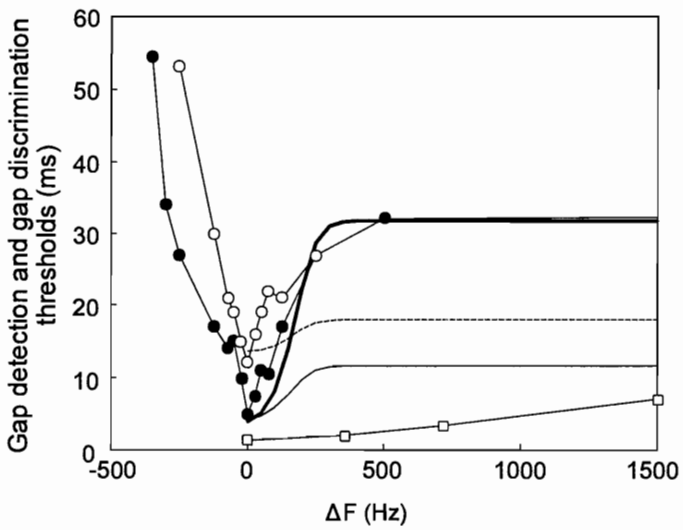

Figure 3.7

Same as figure 3.4, but with different model parameters: α_0=12, fibre threshold = 35 dB, k=20, σ_0=7, SR=40 spikes/s, maximum spike rate = 200 spikes/s.

Apart from the unrealistically high value of threshold, another parameter controls the roll-off

rate of λ_2. This is the tuning parameter α in equation 3.5, which controls the sharpness of tuning. With α_0=4 in equation 3.6, the model tuning curves provide good approximations to the data of Kiang et al. (1965). This is at the auditory nerve level. With α_0=12, tuning is much sharper, but compares with the tuning found at higher levels of the central auditory nervous system.

Cells in the DCN (dorsal cochlear nucleus) are known to exhibit sharp cutoff and inhibitory areas. Cells have also been found in the DCN that have extremely sharp roll-off (Rhode and Greenberg, 1992). Also, frequency tuning in the central auditory system improves deeper into the auditory system. Tuning is sharper at the auditory cortex than at the CN level, which in turn has sharper tuning than the auditory nerve (Sutter, 2000). Sutter gives examples of cells in A1 that have much sharper tuning than what is required to fit the gap detection data. It is quite possible that gap detection takes place at the CN level but the goal here is not to prove or disprove this.

Using this sharper tuning (α_0=12 in equation 3.6), and other more realistic parameter choices, it is shown in figure 3.7 that the gap detection data are approximated quite well by the model. The parameters are shown in the figure caption. Note that the sharp curve is still obtained for high threshold fibres only, but now the model fibre threshold is more realistic. For low threshold fibres, the model predicts better performance. Thus, it seems the auditory system does not perform optimally in this task.

4 DISCUSSION

4.1 Strengths and shortcomings of the model

The expressions that were derived for gap detection and discrimination thresholds as a function of the spike rates during the gap and markers, seem intuitively correct. Building on these, the model provides closed form expressions for gap detection and discrimination thresholds. It is

seen that the model can predict both the magnitude and the trends in the gap detection and gap discrimination data for certain choices of model parameters. The sharper curves for gap detection and the shallower curves for gap discrimination are predicted.

The model assumes that the gap detection mechanism uses sub-optimal statistics (spike counts) in a classical detector to detect the presence of a gap. No prior knowledge of the signal dynamics is used (as opposed to the model of chapter 5). However, as argued in chapter 5, it is likely that the auditory system uses an internal model of the spike generation mechanism.

The models for gap detection and discrimination do not provide an explicit mechanism that can detect a gap from a spike train input. Therefore, it is perhaps not immediately evident that the model for gap discrimination implicitly assumes that λ_1, λ_2 and λ_3 are known, i.e. the calculated CRLB for position of the gap edge assumes that the spike rate during the gap and markers are known. This is not an unreasonable restriction, as it is possible that the auditory system estimates spike rates. Regardless, the CRLB gives a lower bound on the accuracy of the estimation of gap edge times, and possibly the auditory system does not achieve this lower bound. This is supported by gap detection and discrimination data.

To use classical signal detection theory to predict gap thresholds, meant that a simple signal model (spike train model) was required. Closed form expressions could be obtained because spike trains were assumed to be Poissonian. The Poissonian assumption is usually only valid at higher frequencies or deeper in the central auditory nervous system where no phase-locking occurs. However, as argued in the introduction, it is reasonable to use Poissonian spike trains in a model for gap detection, as it seems that temporal information does not play a significant role in determining across-channel gap detection thresholds. Chapter 4 discusses the gap detection problem when phase-locking occurs.

4.2 Parameter sensitivity and the origin of the shape of the gap detection tuning curves

The predicted magnitude of gap detection thresholds by equation 3.37 is consistent with data when realistic values for λ_1 and λ_2 are used. The shapes of the gap detection tuning curves are

determined by several model parameters.

The sharp tips of the gap tuning curves are controlled primarily by the roll-off of the λ versus ΔF function. The sharper the slope of this function, the sharper the tip of the gap detection tuning curve. The slope of this function is, in turn, controlled by α_0 in equation 3.6, which determines the sharpness of tuning.

The slopes of the tuning curve flanks are controlled by σ_0, which determines the slope of the rate-intensity curve. Larger σ_0 result in shallower rate-intensity curves and shallower tuning curve slopes.

The flattening of the gap detection tuning curve beyond the shoulder of the curve is obtained when the spike rate during a gap is less than the SR. The spike rate during the gap may be less than SR when suppression of spike rate occurs after the offset of the first marker (Zhang et al., 1990). The flattening of the tuning curve occurs because, for widely spaced marker electrodes (large ΔF) the spike rate in the observation window during a marker reaches a steady state value (the spontaneous rate SR). As the spike rate during the gap is also fixed in the model, it is seen from equation 3.37 that gap threshold T will reach a steady state value.

4.3 Temporal and spatial models for gap detection

The model hinges on the idea that the auditory system uses an attention filter or observation window that is placed midway between the regions excited by the two markers. As explained in more detail in chapter 4, evidence does exist which challenges this idea. Nonetheless, the model has made concrete the notion that "the task becomes more difficult" as marker frequencies are separated. Increased difficulty of the task is reflected in spike rates during the markers and gap that become more similar as marker frequencies are separated. This is a spatial model for gap detection, as opposed to a temporal model for gap detection. As discussed in the introduction, evidence exists that supports the notion that at least across-channel gap thresholds are based on spatial mechanisms.

Zhang, Salvi, and Saunders (1990) have shown that spike rate decays exponentially after the offset of the first marker, so that the gap is filled with continuing spike activity and therefore brief gaps are difficult to detect. Filling of the gap provides a feasible explanation for within-channel gap detection thresholds, but cannot explain increases in gap thresholds in the across-channel condition.

Whether temporal mechanisms or spatial mechanisms are used to explain gap thresholds, it is clear that separating the marker frequencies results in deteriorated signal-to-noise ratios in the spike train or set of spike trains that the detection mechanism observes. The "signal" that has to be detected in either case is the transition from marker to gap (or gap to marker) that is masked by some source of noise.

4.4 Interpretation of modelling results

To predict gap detection data accurately, especially the sharp tip region, the model requires unrealistically high fibre thresholds when neural tuning curves are modelled after auditory nerve tuning. When sharper tuning is assumed, similar to that found in some DCN fibres or deeper in the central auditory nervous system, the sharp tip region is predicted more easily, but still using high threshold fibres. There is no apparent reason why only high threshold fibres would take part in gap detection.

Assuming more realistic fibre parameters, predicted gap thresholds are smaller than measured gap thresholds in the across-channel condition. The auditory system seems to operate sub-optimally in this task, an observation that has also been made regarding frequency discrimination ability (Siebert, 1970; Hanekom and Krüger, 2001). Spike count statistics are already suboptimal for rate discrimination or transition detection, but even so deterioration of gap detection thresholds in the across-channel condition is faster than the model predicts.

The fact that sharp tuning is required to predict gap detection data with this model suggests that theoretically the auditory system should be able to perform much better on the gap detection task, but is somehow limited by internal noise or other confounding factors. It may

be that subjects find it difficult to attend to the gap because of the different-sounding markers (van Wieringen and Wouters, 1999), but this is probably more important in cochlear implantees than in normal-hearing listeners. The task of detecting a gap between spectrally different sounds occurs frequently in normal hearing, e.g. voice onset time is one feature that distinguishes different consonants.

It is outside the scope of this work to determine where in the auditory system the gap detection mechanism is situated, but it can be speculated on. With many free parameters controlling the shape of the model-predicted gap detection tuning curves, it is important that parameter choices should be biologically plausable. It is therefore adequate to show that the parameter values required to fit the data are in line with neural parameters published in literature. The fact that sharp tuning is required to predict gap detection data points to the possibility that the gap detection mechanism is implemented at the CN level or deeper in the central auditory nervous system.

Finally, comparing model predictions with the data, it seems possible that the task changes from gap detection to gap discrimination at a certain frequency separation (possibly at one critical band, as shown before).

5 CONCLUSIONS

The CRLB for the Poisson change-point problem, which appears not to be documented in literature, has been derived here. The result is intuitively satisfying.

It is shown that the quantity $(\lambda_1 + \lambda_2)/2(\lambda_2 - \lambda_1)^2$ appears in the model for the gap detection task, while the square root of this same quantity appears in the gap discrimination model. A similar observation regarding intensity discrimination has been made before.

A model that can predict the U-shaped curves found in across-channel gap detection has been created. The model is simply based on signal detection theory considerations.

The model shows that a spatial mechanism, as opposed to temporal mechanisms, may contribute to gap detection thresholds in the across-channel condition. This is important in cochlear electrical stimulation, where spike trains are strongly phase-locked to the stimulus and where temporal mechanisms do not seem to determine gap detection thresholds. Accordingly, techniques similar to those used in this chapter are applied in chapter 3 to create a model that can predict gap detection thresholds in auditory electrical stimulation.

Chapter 4

MODELS OF GAP DETECTION IN ELECTRIC HEARING

1 INTRODUCTION

Gap detection is often employed as a way to probe the temporal resolution ability of the auditory system (Shannon, 1989). Gap detection thresholds in electrical stimulation were measured by Shannon (1983a), Preece and Tyler (1989), Shannon (1989), Hanekom and Shannon (1998), Chatterjee, Fu, and Shannon (1998), Busby and Clark (1999) and Van Wieringen and Wouters (1999) under various conditions.

Shannon (1989) measured gap detection as a function of intensity of stimulation, using sinusoids and pulse trains, with both gap markers presented on the same electrode pair (a within-channel condition). He found that gap thresholds are a strong function of stimulus level, with longest gap thresholds of more than 50 ms (at near audible threshold stimulation levels), and shortest gap thresholds of less than 1 ms (at high stimulation levels).

Preece and Tyler (1989), who measured gap detection thresholds in cochlear implants using sinusoidal electrical stimulation, ascribed the longer gap thresholds at lower stimulation levels to a decay of sensation theory of gap detection (Penner, 1977). According to this theory, the gap is detectable when the sensation has decayed by a just-noticeable amount, or when the signal level has changed by more than the intensity difference limen, which is larger near auditory threshold.

Although neural synchrony is much higher in biphasic electrical stimulation than in acoustic stimulation (Javel, 1990) and differs in the pattern and extent of activated nerve fibres (Kral

et al., 1998), Shannon (1989) showed that electrical and acoustic stimulation produced similar gap detection performance. This suggests that gap thresholds are not primarily determined by temporal processing of individual fibre spike trains.

Hanekom and Shannon (1998) employed gap detection, not as a measure of temporal resolution, but as a way to measure electrode interaction in cochlear implants (see chapter 2). The hypothesis was that if the two gap markers were presented on different electrode pairs, gap thresholds would be small when the two electrode pairs stimulated the same neural population (a within-channel condition), and would increase as the neural populations became more disjunct (an across-channel condition; electrode interaction becomes less). If this hypothesis is true, the pattern of gap thresholds as a function of distance between the two marker electrode pairs will also reflect the current distribution in the cochlea. This suggests that one should be able to estimate the current distribution pattern in the cochlea from gap detection tuning curves.

Chatterjee et al. (1998) argued that the U-shaped curves achieved by Hanekom and Shannon (1998) (see chapter 2) could be attributed to a perceptual discontinuity, rather than reflecting the amount by which stimulated neural populations overlap (an across-channel mechanism). They showed that U-shaped curves can also be obtained with a within-channel gap detection mechanism, by presenting both gap markers on the same electrode, but using different stimulation pulse rate or intensity of stimulation for the two markers. Busby and Clark (1999) showed that rate of stimulation (pulse rates of 200 pps, 500 pps and 1000 pps were used) does not influence gap thresholds when the rate of stimulation is the same on both markers.

Van Wieringen and Wouters (1999) measured gap detection thresholds in users of the LAURA cochlear implant. They measured gap detection thresholds in various within-channel and across-channel conditions. Not all of their data showed an effect of distance between electrode pairs when pre- and post-gap markers were presented on different electrode pairs (as was found in Hanekom and Shannon (1998)) and, furthermore, a strong effect of training was found. It appeared that once subjects had been trained to attend to the gap in the presence of

confounding factors, the effect of distance between pre- and post-gap electrode pairs disappeared in some subjects. A primary interpretation of their results was that gap detection thresholds depend more on the subject's inability to attend to the gap when confounding factors are present than on neural interaction. Finally, it is noted that, as in most cochlear implant psychoacoustics studies, the subjects in the Van Wieringen and Wouters study showed much inter-subject variability.

The poorer gap detection performance in acoustic across-channel experiments has been attributed to an auditory attentional process (Phillips et al., 2000). These authors proposed that two different mechanisms operate in within-channel and across-channel gap detection tasks. For within-channel gap detection (in cochlear implants, this is when both markers are presented on the same electrode pair), the task of the auditory system is to detect a discontinuity. In across-channel gap detection (in cochlear implants, this is when markers are presented on different electrode pairs that stimulate disjunct or overlapping but non-identical neural populations), the discontinuity always exists. The task now becomes one of gap duration discrimination. The authors speculated that (in across-channel gap detection) attention resources are allocated to the first marker, resulting in deteriorated ability to measure the time interval between the two markers. A related idea, expanded on later, is employed in the current model. It should be noted that Oxenham (2000) provides reasonable arguments against the idea that attention resources are loaded to the extent that gap detection ability deteriorates.

1.1 Information available to the central detector in acoustic and electric gap detection

1.1.1 Temporal response properties

Two major differences exist in the information available to the central detector in acoustic and electric gap detection. First, in acoustic gap detection, spike trains are much more Poissonian in nature than the highly entrained spike trains found in electric gap detection. This is especially true for acoustic gap detection employing bandlimited noise markers (as used in e.g. Phillips et al. (2000)). Spikes may also occur during the gap, which is seldom the case in electric

hearing (Shepherd and Javel, 1997), as spontaneous activity is usually not present in the deafened auditory system. In the acoustic gap detection situation, the central gap detector probably has to base decisions on the envelope of the marker-gap-marker auditory event as reflected in spatially distributed neural spike train patterns, rather than on individual inter-spike intervals. As explained below, it is likely that the gap detector bases its decisions on inter-spike intervals in electric hearing.

Phase locking of spike trains to the stimuli, similar to what happens in electric hearing, occurs (below 5 kHz, Johnson, 1980) in acoustic gap detection experiments using sinusoidal markers (e.g. Formby and Forrest, 1991). However, 100% entrainment (spikes occur in response to each stimulus pulse) to the preferred stimulus phase as is found in electrical stimulation (Javel, 1990) seldom occurs in acoustic stimulation (Rose et al., 1968). 100% entrainment in pulsatile electrical stimulation can occur at pulse repetition frequencies of up to 800 Hz for healthy or short-term deaf ears, but the ability to entrain to high frequencies is reduced in long-term deafness. Shepherd and Javel (1997) observed entrainment only up to 400 pps in long-term deaf fibres. This is probably a result of prolonged refractory periods that occur in nerve fibres that are demyelinating (Shepherd and Javel, 1997).

Spike position jitter relative to a preferred latency (following a stimulus pulse) has significantly larger standard deviation in acoustic stimulation (Javel and Mott, 1988) than in electrical stimulation (Javel and Shepherd, 2000). Spike position jitter increases with higher frequency in electrical stimulation (Javel, 1990; Javel and Shepherd, 2000), but decreases in acoustic stimulation (Javel and Mott, 1988).

Different response types may be evoked in response to an electric stimulus pulse. As shown in Javel and Shepherd (2000) and Javel (1990), spikes that occur in response to electrical pulse stimuli, cluster into one of four discrete latency windows. These have been labelled *A*, *B*, *C* and *D* responses, in order of increasing latency. Different spike latencies probably result from spike initiation at multiple sites. *A* responses have short latency (in response to the stimulus pulse) and have small temporal jitter. *B* responses have longer latency and more spike position jitter.

Both *A* and *B* responses are observed in deafened ears, but the *A* response is predominant. Just *A* responses are often observed in the majority of fibres in long-term deaf ears (Shepherd and Javel, 1997).

D responses occur with long latency and are not observed in deaf animals and are therefore not relevant to cochlear implants. The *C* response is primarily found at near-threshold intensities, and is possibly just a near-threshold extension of the *B* response. This response probably arises from spikes initiated at the peripheral processes of the spiral ganglion cells. The short latency *A* response probably results from spikes initiated at the central processes of spiral ganglion cells. Latency transitions from *B* to *A* responses are sometimes observed as stimulus intensity is increased, but this transition is not found in long-term deaf ears (Shepherd and Javel, 1997). Van den Honert and Stypulkowski (1987a) observed that both phases of a biphasic pulse stimulus can elicit spikes and that these spikes may be initiated at different sites, and therefore with different latencies. *A* and *B* responses may result from different phases of a biphasic stimulus pulse.

1.1.2 Spatial excitation patterns

The second difference between acoustic and electric gap detection is that a wider extent of neural activation is brought about with electrical stimulation. Maximum slopes of spatial tuning curves of 8 dB/mm have been measured for bipolar electrical stimulation with the Nucleus electrode array (Kral et al., 1998), while tuning curves in the 8 kHz region can have slopes of up to 100 dB/mm in acoustic stimulation. Furthermore, current spread may increase in long-term deafness, due to demyelination of nerve fibres (and therefore changing resistive pathways) (Shepherd and Javel, 1997).

In summary, in electric gap detection a wide spatial extent of spike trains, weakly to highly entrained to the stimulus pulse train, contains the gap. The area of activation is narrower in acoustic gap detection and the gap is contained in much noisier spike trains.

1.2 Differences in the gap detection task between acoustic and electric hearing

Although the neural circuitry of the central detector will be the same in both cases, the task of the central detector in electric and acoustic gap detection is probably different. As discussed before, the within-channel gap detection task in acoustic stimulation is one of detection of a transition. The across-channel gap detection task is a time duration discrimination task, with gap detectability limited by (i) the masking of edges of the gap as a result of Poisson encoding, and (ii) the measurement accuracy of the "Poissonian timer" (see chapter 3).

When entrainment is 100% in electrical stimulation, the task is to detect one slightly longer inter-spike interval within a series of fixed duration inter-spike intervals observed with high signal to noise ratio (the noise being the spike position jitter). This is a classical detection task.

When entrainment is less than 100%, the real gap is hidden by spike trains containing "gaps" that are multiples of the stimulation pulse repetition period. Detecting the actual gap when it is small (a fraction of one pulse period for 100% entrainment) will be quite difficult for a detector by observing a spike train from just a single fibre. In this case, the gap will have to be longer than the typical inter-spike intervals obtained when entrainment is 100%. This notion will be explained in more precise terms later.

In both cases (100% entrainment or less) in electric hearing, the detector will need to compare each inter-spike interval against a threshold. Accurate measurement of the inter-spike interval is not required. So it seems the task in gap detection in electric hearing does not reduce to a duration measurement task in the across-channel condition.

1.3 Models for gap detection

Several models for gap detection in acoustic hearing were discussed in the previous chapter. As far as is known, no previous models of gap detection in electric hearing exist. An appropriate model for gap detection should be able to predict (i) the sensitivity of gap detection

thresholds to intensity of stimulation as found by Shannon (1989), (ii) the U-shaped curves found by Hanekom and Shannon (1998) and Chatterjee, Fu, and Shannon (1998) (under different conditions) and (iii) the correct magnitude of gap detection thresholds as found by several authors. Such a model should employ a biologically plausible mechanism and plausible parameter values.

Two possible mechanisms, or a combination thereof, may be adopted to explain gap detection thresholds. The first explanation is primarily based on temporal mechanisms. It has been hypothesized that ringing in the mechanical response of the basilar membrane following stimulus offset prolongs the neural response, so that these spikes will tend to fill the gap (Zhang et al. 1990).

Cochlear implants stimulate nerve fibres in the cochlea directly, so that the cochlear filtering step is bypassed. Neural synchronization is high, with 100% entrainment of neural spike trains to the stimulation pulses often occurring in fibres close to the electrodes (Javel, 1990), and generally no neural activity occurring when the stimulus is absent. Improved gap detection thresholds might reasonably have been expected for electrical stimulation of the auditory system. However, evidence suggests that gap detection thresholds do not depend on neural synchronization alone (Shannon, 1989), and that other sources of noise affect the ability of the central gap detection mechanism to detect the presence of the gap accurately.

A second possible explanation of gap detection thresholds in cochlear implants lies in the mechanism of "auditory attention". This is a spatial, rather than temporal, mechanism. The hypothesis is that the auditory system can only attend to a limited spatial extent of neural activity. Accordingly, an idea related to the auditory attention mechanism suggested by Phillips et al. (2000) is employed in the current model. The model proposes that the *detectability* of the second marker onset deteriorates when the marker electrode pairs are moved apart. In contrast to the explanation of Phillips et al. (2000), no time measurement is required in the model.

It is implicitly assumed that the listener attends to the gap in the gap detection task. Taking a neurophysiological view, it is assumed that attention in this case means that the central detector observes a spatially limited extent of fibres. To detect the gap, it is necessary that the detector observes a set of fibres with spike trains entrained (not necessarily 100%) to *both* the gap markers. If this is not so, the central detector would either observe the offset of the pre-gap marker, or the onset of the post-gap marker, and the gap will not be detected. The limited extent of fibres observed will consequently probably be situated somewhere midway between the two electrode pairs defining the two markers. When the marker electrodes are far apart, the observation window will contain fibres with low entrainment.

This attentional mechanism is explored in this chapter. It is emphasized that this is a viewpoint that is based on unproved assumptions.

1.4 Objectives of this chapter

In summary, this chapter intends to build understanding of the processes underlying the results achieved in the psychoacoustics of gap detection in electric hearing, by creating models that can predict the data reported in chapter 2. It is shown that gap detection data can be predicted by a model that uses cochlear spatial information, using plausible model parameters. Accordingly, gap detection ability does not necessarily reflect temporal processing mechanisms. The model hinges on the idea that a spatially limited neural observation window is available to a central detector. A final and important result in this chapter is the estimation of current distribution profiles in the cochlea from psychoacoustic gap detection data.

2 A MODEL FOR GAP DETECTION IN ELECTRIC HEARING

2.1 Assumptions about the electrically evoked spike train

The model assumes a long-term deaf auditory system. Consequently, it is assumed that spikes

are highly entrained to the stimulus pulse train and that only *A* responses occur. This last assumption is motivated by the fact that responses other than the short latency, small jitter *A* responses are seldom observed in electrically stimulated long-term deaf ears. The assumption is made to simplify the model so that obtaining analytical predictions for gap thresholds without reverting to time-consuming (and sometimes less enlightening) Monte Carlo simulation techniques is possible. It would, however, be possible to incorporate "noisier" spike train models (that include *B* and *C* latency responses as well) in a numerical model.

2.2 General formulation of a gap detection model for electric hearing

In a general formulation, the gap detection problem in electric hearing is similar to the acoustic gap detection problem, where (for sinusoidal markers of low frequency) spike trains are phase-locked to the stimulus. 100% entrainment is assumed in this treatment. So it is assumed that spikes occur in response to each marker stimulus pulse with some temporal dispersion (or jitter) and that spontaneous discharges occur during the gap. The average spike rates λ_1 and λ_3 during the markers differ in this general formulation, and they also differ from the average spike rate λ_2 in the gap. To simplify the formulation, it is assumed that spontaneous discharges during the gap do not occur randomly, but rather that these discharges are periodic. Spike position standard deviation is σ_{s1} during the markers, and σ_{s2} during the gap. Figure 4.1 visualizes the spike train model for this formulation.

The optimal detector for an a priori known signal is a matched filter. In the gap detection problem, the "known signal" is the expected transition in spike rate that will occur when the gap occurs. Ideally, the matched filter will know the spike rate during the gap and during each marker. The matched filter for the first transition is shown in figure 4.1(e). If more samples are available, the matched filter detection probability will improve. It is assumed that the signal samples that the matched filter operates on are the inter-spike intervals. The matched filter in figure 4.1 observes the spike train for a time 2T. Assume that N inter-spike intervals are observed in T.

If T is too short, the gap will not be detected. The shortest T for which the percentage of correct transition detections is 76% (d'=1), is the "transition threshold". This threshold will depend on how much λ_1 and λ_2 differ. Both transitions (pre-gap marker to gap, and gap to post-gap marker) need to be detected to detect the gap. The gap will be detected if two consecutive transitions are detected with d'=1.6 as explained in chapter 3.

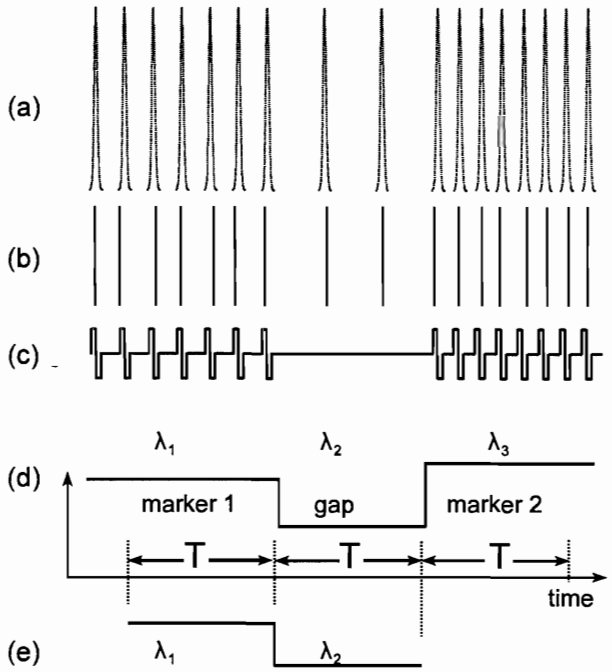

Figure 4.1

This figure gives a general model for the gap detection problem in electric hearing. The progression from the first marker with spike rate λ_1, through the gap with spike rate λ_2, to the second marker with spike rate λ_3 is depicted in (d). The stimulus (gap-marker-gap) is shown in (c). The matched filter shape is shown in (e). (a) shows the distribution of spikes around the preferred stimulus latency. The resulting spike train is shown in (b).

If just one sample (one inter-spike interval) is observed, the detectability is

$$d' = \sqrt{\frac{2(\mu_2 - \mu_1)^2}{\sigma_1^2 + \sigma_2^2}} = \sqrt{\frac{2(\mu_2 - \mu_1)^2}{2\sigma_{s_1}^2 + 2\sigma_{s_2}^2}} = \sqrt{\frac{(\mu_2 - \mu_1)^2}{\sigma_{s_1}^2 + \sigma_{s_2}^2}}. \qquad (4.1)$$

with $\mu_1=1/\lambda_1$ and $\mu_2=1/\lambda_2$, where λ_1 and λ_2 are the spike rate during the first marker and gap respectively. σ_1 and σ_2 are the standard deviations of samples during the marker and gap respectively. The standard deviation σ of a sample is the standard deviation in an inter-spike interval, which is $2\sigma_s$, where σ_s is the spike position standard deviation. σ_{s1} is the standard deviation during a marker, and σ_{s2} during the gap.

If T is longer, so that N samples are observed in T, detectability improves with the square root of N,

$$d'_N = \sqrt{N}d', \tag{4.2}$$

(Green and Swets, 1966). Hence,

$$d'_N = \sqrt{N}\sqrt{\frac{(\mu_2 - \mu_1)^2}{\sigma_{s_1}^2 + \sigma_{s_2}^2}}. \tag{4.3}$$

The number of samples in T is

$$\lambda T - 1, \tag{4.4}$$

where λ is λ_1 or λ_2. The detectability is then

$$d'_N = \sqrt{\lambda_2 T - 1}\sqrt{\frac{(\frac{1}{\lambda_2} - \frac{1}{\lambda_1})^2}{\sigma_{s_1}^2 + \sigma_{s_2}^2}}, \tag{4.5}$$

where λ_2 was used in the expression for N, because it is assumed that $\lambda_2 < \lambda_1$, which implies that spike rate λ_2 during the gap is the critical factor that will determine smallest detectability.

For detection of the gap (detection of two transitions), $d'_N=1.6$ and equation 4.5 is solved for T,

$$T = \frac{1}{\lambda_2} + \frac{\sigma_{s_1}^2 + \sigma_{s_2}^2}{\left(\frac{1}{\lambda_2} - \frac{1}{\lambda_1}\right)^2} \cdot \frac{1.6^2}{\lambda_2}. \tag{4.6}$$

This T is the shortest T that the matched filter will require to detect the gap with at least 76% probability. When the spike standard deviations σ_{s1} and σ_{s1} are small,

$$T \approx \frac{1}{\lambda_2}. \tag{4.7}$$

The minimum value of T is $1/\lambda_2$. Therefore, for small spike jitter (as obtained in pulsatile electrical stimulation), this shows that the matched filter does not require the observation of several stimulation pulses, but will be able to detect the gap by comparing just two inter-pulse intervals (as reflected in the inter-spike intervals), i.e. one interval that does not contain the gap, and one longer interval that does contain the gap.

The gap detection problem can then be reformulated as in figure 4.2. If the duration of the interval that contains the gap is $1/\lambda_2$,

$$\begin{aligned} \textit{gap threshold} &= T - \frac{1}{\lambda_2} \\ &= \frac{\sigma_{s_1}^2 + \sigma_{s_2}^2}{(\frac{1}{\lambda_2} - \frac{1}{\lambda_1})^2} \cdot \frac{1.6^2}{\lambda_2}, \end{aligned} \tag{4.8}$$

which is the gap threshold if the detector searches for two transitions.

2.3 Gap detection model structure for electric hearing

The gap detection problem for electric hearing is depicted in figure 4.2. Having proven that the gap detector only needs to compare each observed sample (inter-spike interval) with the standard inter-spike interval, the task of the detector simplifies to the detection of one longer inter-pulse interval embedded in a pulse train with fixed period (1 ms in this case). The detector does not need to search for two transitions, but just for the longer interval, so that the gap threshold is not given by equation 4.8, but by equation 4.11 below. The gap threshold is taken as the additional time added to the standard stimulus period, when this longer interval is just detectable (76% correct decision in a 2IFC experiment).

2.3.1 Gap detection based on a single fibre when entrainment is 100%

In the experiments modelled here, the stimulation pulse repetition frequency is always the same on both markers. In the simplest scenario, the pre- and post-gap markers are presented to the same electrode pair (the within-channel condition) and the input to the central detector is a spike train on a single nerve fibre. In this scenario, the input to the central gap detector is a spike train that is ideally 100% entrained to the stimulation pulse train with only a small amount of spike position jitter.

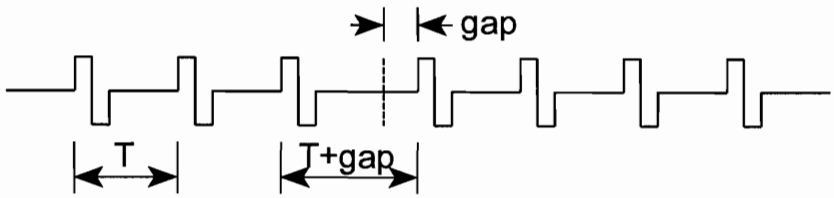

Figure 4.2

A part of a stimulation pulse train with pulse repetition frequency 1/T, showing the definition of the gap.

In this situation, the gap detection problem reduces to a simple signal detection problem: the central detector has to decide which of two inter-pulse intervals occurred, the standard, or the probe (standard plus gap). For this problem, d' is given as

$$d' = \sqrt{\frac{(\mu_1 - \mu_2)^2}{\sigma_{s_1}^2 + \sigma_{s_2}^2}}, \tag{4.9}$$

with $\mu_1=1/\lambda_1$ and $\mu_2=1/\lambda_2$, where λ_1 and λ_2 are the spike rate during a marker and during the gap respectively. For 100% entrainment, the spike rate λ_1 equals the stimulus rate.

Note that in general spikes may occur during the gap, with the spike rate during the gap being λ_2. It was, however, shown in the previous paragraph that under ideal circumstances, when spikes occur only on stimulus pulses (with 100% entrainment), the window T required to detect a gap is equal to $1/\lambda_2$ plus a small duration that is determined by the spike jitter (spike position standard deviation, σ_s). Hence, for 100% entrainment, the temporal window T

required to detect a gap is just slightly longer than the pulse repetition period, so that the detection problem in figure 4.2 is obtained. The detector does not have a large number of samples (inter-spike intervals) in the gap to compare with inter-spike intervals during the marker. For each input sample, the detector has to decide whether the gap was present or not. This is a standard signal detection problem.

With standard deviation in spike jitter σ_s=0.1 ms (assumed to be identical for spikes marking the gap, and spikes marking other inter-pulse intervals), and using d'=1 for 76% correct decisions, μ_2 is calculated as 1.14 ms. The gap threshold is then μ_2 - μ_1. Hence, the shortest detectable gap is 0.14 ms for this example. More generally,

$$d' = \frac{gap\ threshold}{\sqrt{\sigma_{s_1}^2 + \sigma_{s_2}^2}} = \frac{gap\ threshold}{\sqrt{2}\sigma_s} \tag{4.10}$$

if σ_{s1}=σ_{s2}, which is assumed to be true. Hence for d'=1 the gap threshold is given by

$$gap\ threshold = \sqrt{2}\,\sigma_s \tag{4.11}$$

for the case of 100% entrainment.

2.3.2 Gap detection based on multiple fibres when entrainment is less than 100%

More generally, the detector will observe M parallel neural channels (M adjacent or closely spaced nerve fibres), of which not all spike trains will be 100% entrained. A stimulus on a particular electrode pair will produce a current distribution with current decaying away from the electrode, so that it may be expected that some fibres close to the electrode will have 100% entrainment, but distant fibres may not fire at all, and some fibres will have intermediate entrainment. Fibres stimulated considerably above their thresholds will have 100% entrainment, but fibres stimulated near threshold will not have a spike probability of one on each stimulus pulse.

2.3.2.1 Model assumptions

Several assumptions are made for modelling purposes. These assumptions are intended to fulfil one of two purposes: either they reflect a certain belief about the nature of the gap detection mechanism, or they are meant to disregard "noise" so as to focus on the underlying signal processing hypothesized to take place.

(1) It is assumed that the central detector observes an attentional window of nerve fibres situated midway between the two markers when the markers are not presented on the same electrode pair. This observation window is chosen to provide optimal detection probability. If the observation window is closer to one electrode than to the other, one electrode might elicit high spike entrainment, while the other electrode does not elicit any neural response in the observation window. The gap will then not be detected. An observation window midway between the two marker electrode positions will provide optimal detection probability.

(2) It is assumed that this observation window cannot be of any given spatial extent, but is limited. In the current model, the observation window is assumed to be limited to one critical band or auditory filter. The observation window is essentially just an off-frequency auditory filter. The psychophysical concepts of critical bands, auditory filters and off-frequency listening are discussed in (Patterson and Moore, 1986).

(3) Finally, it is assumed that the observation window can shift along the axis of the cochlea with a resolution of one nerve fibre, i.e. the M nerve fibres observed can be shifted by one nerve fibre towards the apex or base. This seems to be a reasonable assumption, as critical bands shift dynamically to centre on a given stimulus frequency (Patterson and Moore, 1986).

2.3.2.2 Inter-spike interval probability density function for less than 100% entrainment

The further the marker electrode pairs are apart, the more distant the observation window will be from the electrodes, and the smaller the probability of 100% entrainment becomes. The inter-spike interval histogram (ISI histogram) measured on a single fibre has more than one

mode if entrainment is less than 100%. These modes occur at multiples of the stimulation pulse repetition period. The standard deviation of each mode is σ_s. If the ISI histogram is normalized to an area of 1, it is the pdf (probability density function) of the inter-spike interval of a single fibre (referred to as the ISI pdf in the text that follows). A typical ISI histogram is shown in figure 4.3.

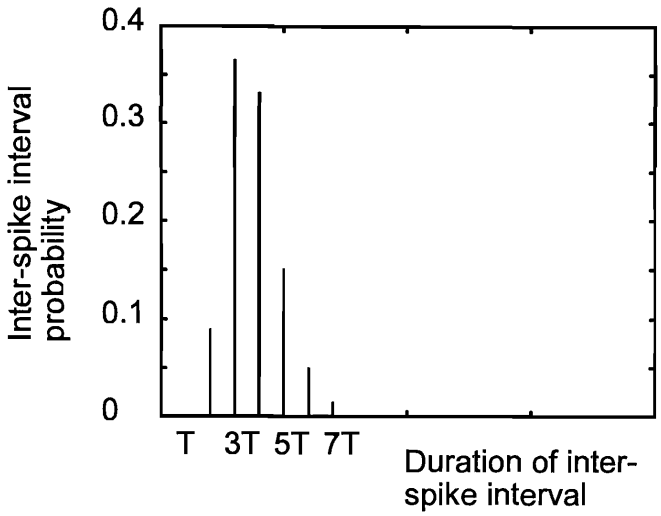

Figure 4.3

A typical ISI pdf for entrainment less than 100%. T is the stimulation pulse repetition period. This figure was calculated using the model of (Bruce et al., 2000). Note that although the figure shows a mode standard deviation of σ_s=0, mode standard deviation is typically up to 0.1 ms. The model used here does not take mode standard deviation (spike jitter) into account.

Assume now that the task is to detect the gap when there is not 100% entrainment. The signal detection task of the central detector is now complicated by the multi-mode pdfs that characterise the two conditions that have to be discriminated, i.e. the pdf of inter-spike intervals when a marker is observed, and the pdf of inter-spike intervals when the gap occurs. The first two modes in these two pdfs will be at T and T+gap respectively, as shown in figure 4.4. It is clear that, to achieve a large fraction of correct decisions, T+gap will be more

distant from T than in the single-mode pdf case. Hence, in this case the gap threshold will be determined by the standard deviation of the multi-mode pdf and not by the spike position jitter.

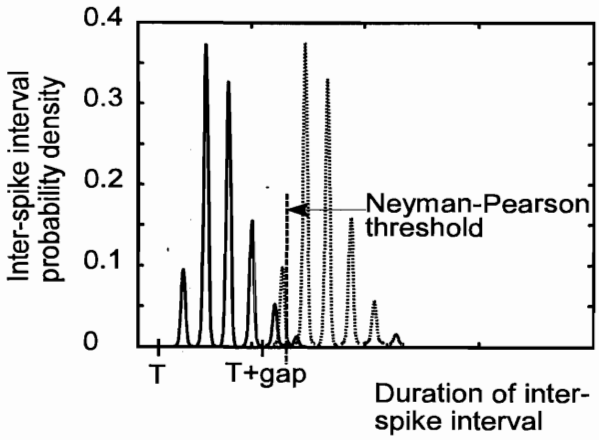

Figure 4.4

Two ISI pdfs, with first mode (of zero amplitude) at T and T+gap respectively, with T+gap chosen for high gap detection probability in this figure.

The model of Bruce, as described in Bruce et al. (2000), Bruce et al. (1999a), Bruce et al. (1999b) and Bruce (1997) calculates the probability of a spike occurring on each stimulation pulse in a pulse train for a given current distribution and stimulation pulse repetition frequency. The complete mathematical model is derived in Bruce et al. (2000) and the derivation is not repeated here. Their model implementation in Matlab is employed for some of the calculations in the current model (with permission from the author of the code).

The Bruce model calculates the normalized ISI histogram internally. It does not take spike position jitter into account. Typical ISI histograms obtained with the Bruce model are shown in figure 4.5.

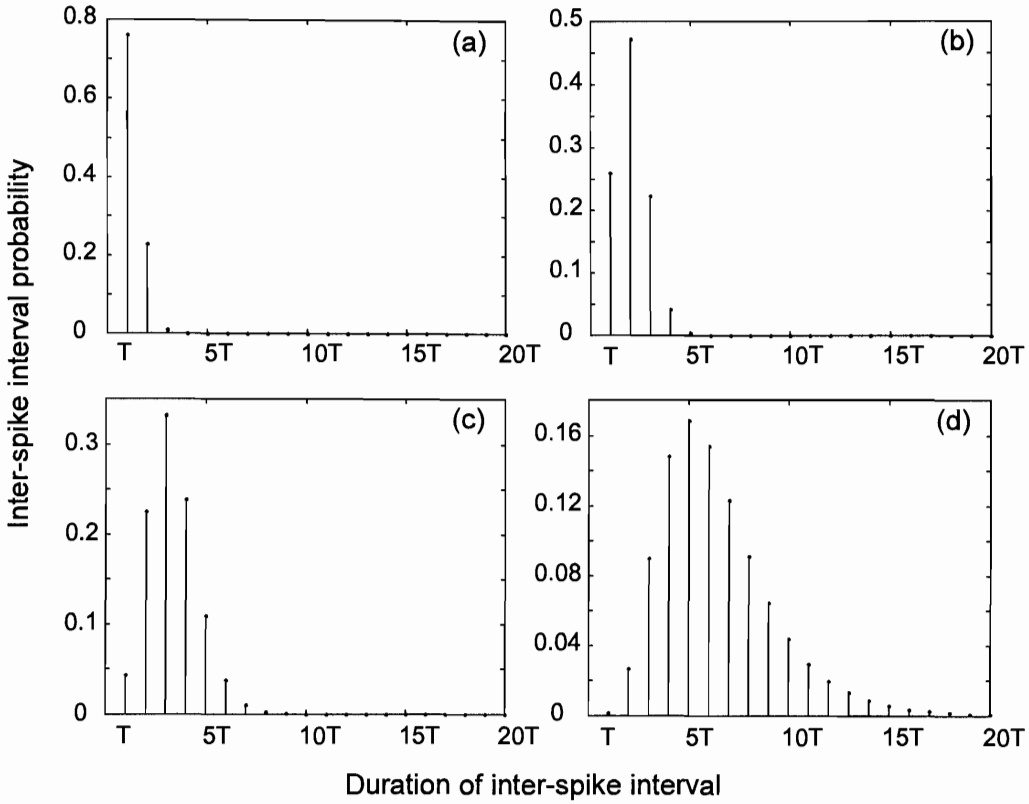

Figure 4.5

ISI pdfs (with no jitter) for several situations in which entrainment is not 100%. The probability of a spike occurring on each stimulus pulse is as follows: (a) P(spike)=0.8, (b) P(spike)=0.5, (c) P(spike)=0.3 and (d) P(spike)=0.16.

Bruce uses the concept of renewal time in his model. The renewal time is the waiting time between the occurrence of one spike and the next. The average renewal time and standard deviation in renewal time are calculated by the Bruce model. These are respectively the mean and standard deviation of the ISI pdf. For example, the average renewal time in figure 4.5a is 1.25T and the standard deviation in renewal time is 0.46T, where T=1 ms is the period of the stimulus pulse train. The average renewal time in figure 4.5c is 3.3T and the standard deviation in renewal time is 1.25T.

2.3.2.3 Approximation to the ISI pdf

To simplify calculations, the ISI pdfs are approximated by their envelopes. For small standard deviation in renewal time, the envelope of the ISI pdf is well approximated by a gamma pdf,

$$pdf(x) = \frac{a^{b}\, x^{b-1}\, e^{-ax}}{\Gamma(b)}, \tag{4.12}$$

where x is the time axis, and a and b are parameters used to fit the gamma pdf to the ISI pdf envelope. Parameters a and b together determine the mean and standard deviation of the gamma pdf. Parameter b determines the order of the gamma function. To obtain a good fit to the ISI pdf data, it was found that a can be taken as

$$a = \frac{2}{\sigma_r}, \tag{4.13}$$

where σ_r is the standard deviation in renewal time calculated by the Bruce model. For small σ_r (below 2 ms), the fit is quite good for a gamma function of order b=5. For larger σ_r, smaller values of b are required to obtain a good fit. A polynomial was fit to the required b values for values of σ_r between 2 ms and 100 ms

$$b(\sigma_r) = p_1\sigma_r^7 + p_2\sigma_r^6 + p_3\sigma_r^5 + p_4\sigma_r^4 + p_5\sigma_r^3 + p_6\sigma_r^2 + p_7\sigma_r + p_8\,. \tag{4.14}$$

The polynomial coefficients are given in Appendix 4.A.

2.3.2.4 Combining information from M channels

Assume now that the observation window is M neural channels wide (i.e. M nerve fibres terminate on the central detector). The information from the M channels is assumed to be combined optimally. The ISI pdfs for the M channels in the observation window are very similar if the observation window is not wider than one critical band, and may be approximated by an observation of M identical neural channels. For M identical channels, d'_M improves with the square root of M,

$$d_m = \sqrt{M}\, d', \tag{4.15}$$

where d' is for each individual channel (Green and Swets, 1966).

2.3.2.5 Calculation of the gap threshold

The gap threshold is now obtained as follows. First, the d' required in each individual channel is calculated from equation 4.15, assuming that the gap threshold is reached when $d'_M=1$. This is an equivalent d', as d' is only defined for Gaussian pdfs. From this d', the required probability of detection P_d in each channel is calculated, for a 2IFC experimental context.

$$P_d = 1 - Erfc\left(\frac{\sqrt{2}\ d'}{2}\right). \tag{4.16}$$

P_d is

$$P_d = \int_{x_{th}}^{\infty} pdf(x|T + gap)\,dx\,. \tag{4.17}$$

The pdf in this equation is the inter-spike interval pdf, conditioned on the occurrence of the gap. The integral is from x_{th}, the detection threshold that the detector uses. This detection threshold is obtained by the Neyman-Pearson criterion (Kay, 1998), which employs a constant false alarm probability P_{fa}. A false alarm probability of $P_{fa}=0.05$ is used, and x_{th} is calculated from

$$P_{fa} = \int_{x_{th}}^{\infty} pdf(x|T)\ dx\,. \tag{4.18}$$

Equation 4.17 must then be solved for T+gap. The integrals in equations 4.17 and 4.18 can be found in closed form. The solution to the integral

$$\int_{x_0}^{\infty} \frac{a^b x^{b-1} e^{-ax}}{\Gamma(b)}\ dx \tag{4.19}$$

is

$$\frac{\Gamma(b, x_0 \cdot a)}{\Gamma(b)} . \tag{4.20}$$

The definitions of the gamma function $\Gamma(b)$ and incomplete gamma function $\Gamma(b,a)$ are given in Appendix 4.A.

Equations 4.17 and 4.18 are nonlinear equations that can be solved with optimization procedures. The gap threshold model was implemented in Matlab. In the Matlab implementation, equation 4.18 is first solved for x_{th}, using the *fsolve* function in the Matlab Optimization Toolbox. This routine employs a Newton method nonlinear least squares algorithm to solve (a system of) nonlinear equations.

With x_{th} known, and the required P_d known from equation 4.16, the gap threshold is obtained from equation 4.17. In implementation it is equivalent, but easier, to solve for the value of x_1 (using the fsolve function) that satisfies

$$\int_{x_1}^{\infty} pdf(x|T)\, dx = P_d . \tag{4.21}$$

The gap threshold is then

$$x_{th} - x_1 . \tag{4.22}$$

2.3.2.6 Software implementation

The procedure defined by equations 4.12 to 4.22 was used to calculate the gap threshold for electrical stimulation when entrainment of some nerve fibres in the observation window was less than 100%. The gap threshold was calculated from equation 4.11 for 100% entrainment. All calculations were done in Matlab 5.3 on a Pentium III based computer running Windows 2000.

2.4 Model parameters

The following model parameters can be varied: neural parameters of fibre threshold and RS (relative spread, a parameter that characterizes the slope of the rate-intensity function, Bruce et al., 2000), current distribution, distance between electrode and fibres, extent of the observation window, and stimulation parameters (intensity, pulse width and pulse repetition frequency). It is shown that with realistic choices of model parameters, predicting measured gap thresholds is possible.

2.4.1 Current distribution

A simple exponential model of current decay, which assumes a homogeneous resistive medium, was one of the two models of current spread investigated in the current study. Bruce et al. (1999b) approximated the current distribution with a point source of current at the active electrode, with a current decay of 4 dB/mm for bipolar stimulation. The model and measurements of Kral et al. (1998) found a current decay of between 4 and 8 dB/mm for bipolar electrical stimulation with the Nucleus-22 electrode array. Electrodes were spaced 0.75 mm apart. Black et al. (1983) measured current decay of 6 dB/mm close to the bipolar electrode (within 1 mm of the electrode), but slower current decay further away from the electrode. Various rates of decay were investigated as appropriate models for the data.

The three-dimensional spiralling finite element model of Hanekom (2001) does not assume a homogeneous medium, but attempts to model the impedance characteristics of the different types of tissue in the cochlea more accurately. Her model uses various electrode positions to predict the voltage distribution in the cochlea. The voltage distribution on the cochlear nerve is then used as input to a nerve fibre model that predicts which fibres will fire. The model clearly shows that, for bipolar stimulation, two peaks of neural activity exist close to the active and return electrodes. The peaks do not occur simultaneously, but on each phase of the biphasic stimulation pulse. The model does not take refractory effects into account.

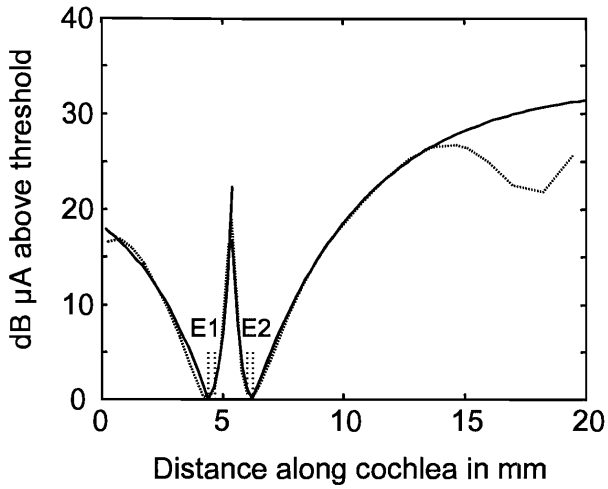

Figure 4.6

The figure shows predicted neural thresholds obtained by the finite element model of Hanekom (2001) (dotted line) along with a curve fit to these data (solid line). The dashed lines E1 and E2 on the figure mark the positions of the two electrodes in a bipolar pair (BP+1 stimulation mode).

Hanekom's results are given as threshold predictions. The threshold of the nerve fibre that fires at the lowest stimulation current (typically a fibre closest to an electrode) is used as reference, and the thresholds of fibres further away from the electrodes are calculated relative to this reference threshold. An example is shown in figure 4.6. For use in calculations, curves have been fit (figure 4.6) to the model-predicted data of Hanekom (2001). Hence, the threshold current (stimulation current at the electrode) of each nerve in the observation window can be calculated.

Assuming that fibres have fixed thresholds, it can be calculated how far above threshold a fibre is stimulated,

$$I_a = I_e - I_{e,th}\,, \tag{4.23}$$

where I_a is the current relative to threshold (in dB µA), I_e is the electrode stimulation current

and $I_{e,th}$ is the electrode current at which the fibre reaches threshold. Then the current at position x is given by

$$I_x = I_{th} + I_a \,, \tag{4.24}$$

where I_{th} is the fixed fibre threshold. The I_x values for all positions x in the observation window define the current spread profile in this window. Position x is defined as the linear distance along the axis of the cochlea, as shown in figure 4.7. Figure 4.7 shows how the physical setup is modelled in the current model. The distance y in this figure is the shortest distance between the electrode array and the neural plane, while the distance z is the direct distance between a specific electrode and a specific nerve fibre.

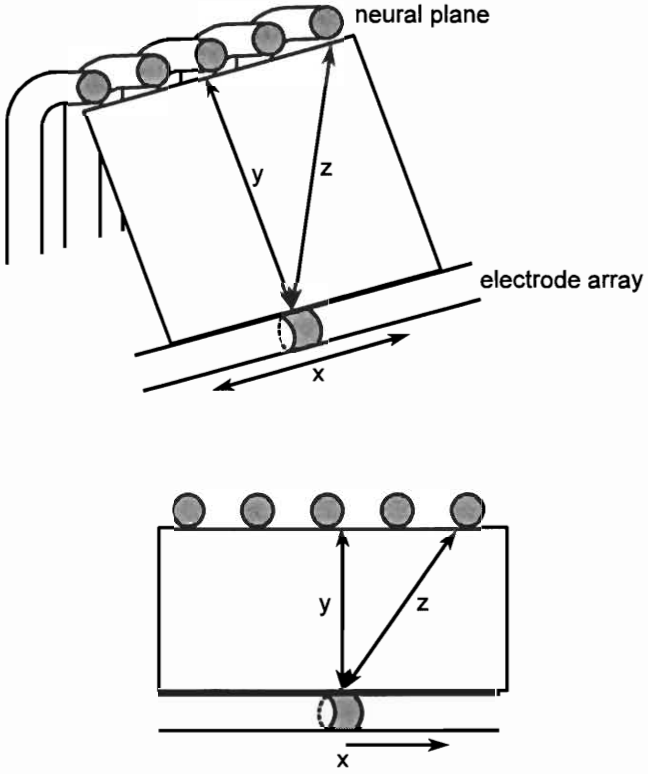

Figure 4.7

This figure defines the spatial relationship between the electrode array and the closest point of stimulation on the neural plane.

The current distribution model in equation 4.24 was the second of the two models of current spread investigated in the current study.

2.4.2 Distance between electrodes and nerve fibres

Distance y in figure 4.7 is the shortest distance between the electrode array and the neural plane. In the Bruce model (Bruce et al., 2000), y=0. More realistically, the electrode array is further away from the neural plane in the base region of the cochlea and nearer to the neural plane in the apex region as a result of the tapering of the cochlea (figure 4.8). Typical values are y=0.8 mm 25 mm from the base, and y=1 mm near to the base of the cochlea (Shepherd et al., 1993). In the Hanekom (2001) model, the electrode is 0.62 mm from the neural plane for an electrode that lies against the outer wall of the cochlea.

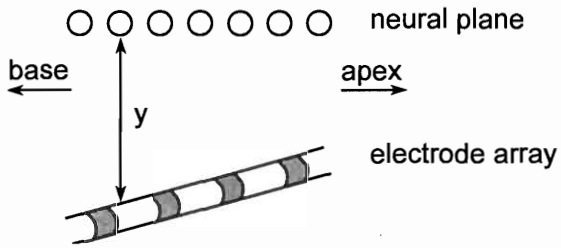

Figure 4.8

The electrode array is closer to the neural plane towards the apex. The distance y varies linearly along the length of the cochlea.

In the current model, fixed distance between the electrode array and neural plane, and also varying distance, were investigated. The fixed distance was taken as 0.5 mm and the electrode array that varied in distance from the neural plane was 0.1 mm away at the apex and 2 mm at the base. Although these values were not realistic, they were chosen to amplify the effect of electrode distance. It turns out that electrode distance has little influence in the model.

2.4.3 Rate-intensity function of nerve fibres

Electrically stimulated nerve fibres have very limited dynamic range, and this is also a function of the state of the nerve fibres. The dynamic range of input intensity (between threshold and saturation) can be as little as 2 dB μA for fibres in a non-deafened cat (Javel and Shepherd, 2000). For the non-deafened cat as well as for a short-term deaf animal, the rate-intensity function is logarithmic (Javel and Shepherd, 2000). However, for the long-term deaf animal, wider dynamic ranges are found, and the rate-intensity function is linear. For the current model, which is intended to model psychoacoustic data of subjects that have been deaf for a long time, it is assumed that the fibre rate-intensity function follows the pattern seen in long-term deaf animals.

The saturation discharge rate, and consequently the dynamic range, is also determined by the stimulation pulse rate for electrical stimulation. Spike rates can be entrained to pulse rate up to very high pulse rates (800-1000 Hz), although some fibres reach saturation rates at lower pulse rates (Javel, 1990; Javel and Shepherd, 2000), especially in long-term deaf ears.

Bruce et al. (1999b) determined mean values and standard deviations for fibre thresholds and RS for the population of (non-deaf) nerve fibres measured by Dynes (1996) (see also Dynes and Delgutte, 1992) and modelled by Bruce. The average fibre threshold was 46 dB μA and the average RS was 0.14. However, as is shown in figure 4.9, larger values of RS replicate the rate-intensity data much better. In the current model, fixed neural parameters (for threshold and RS) are assumed, and specifically, the data for neuron 3-21 as documented in Bruce et al. (1999b) is used. Neuron 3-21 had RS=0.151. In the current model, the effect of RS=0.151, RS=0.3 and RS=0.5 are investigated.

2.4.4 Other stimulation parameters

The following stimulation parameters were fixed in the current model.

The observation window was never wider than one critical band. The number of hair cells per critical band is around 150 (Zwicker and Fastl, 1990), and it was assumed that the central

detector used information from anything between M=1 and M=150 fibres.

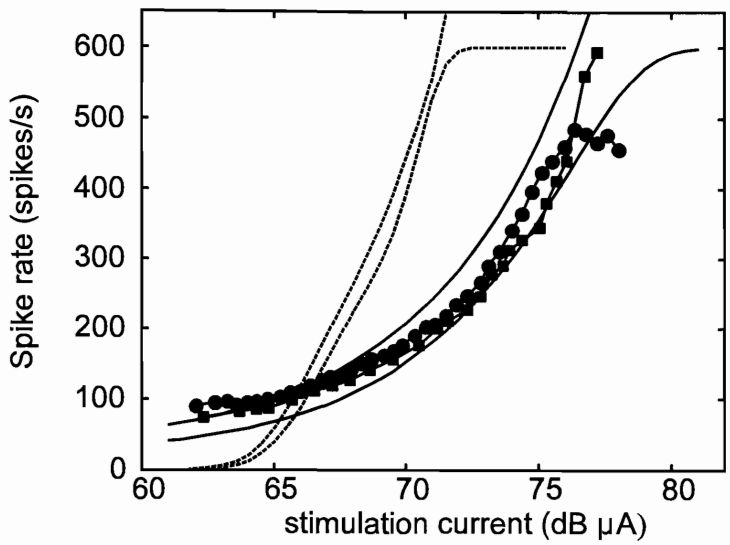

Figure 4.9

The data of Javel and Shepherd (2000) for a long term deaf cat stimulated at 600 pps (●) and at 800 pps (■) are shown together with rate-intensity functions obtained with the Bruce model, with RS=0.5 (solid lines) and RS=0.151 (dotted lines). In both cases, curves for 600 pps and 1000 pps stimulation rates are shown. The 600 pps curves saturate at 600 spikes per second.

Stimulation pulse rate of 1000 Hz and pulse width of 200 μs/phase were used in the model, as these were the parameters used in the psychoacoustic experiments that this model is intended to model. For similar reasons, BP+1 mode was assumed, and it was assumed that the electrode array was the Nucleus-22 design (Clark et al., 1990). Electrodes in a bipolar pair were 1.5 mm apart.

3 RESULTS

3.1 Gap detection tuning curves predicted by model

Figures 4.10 to 4.22 show gap detection tuning curves for various parameter choices in the model. Parameters investigated for their influence on gap detection thresholds are described below.

(1) The effect of the number of fibres in the observation window (M=1 to M=150) was investigated. The width of the observation window was at most one critical band.

(2) The effect of current distribution was investigated. The rate of current decay away from the active electrode site was varied for simple exponential current decay. The twin-peaked current distributions predicted by the model of Hanekom (2001) were also investigated as an alternative current distribution model.

(3) Neural parameters (fibre threshold and RS) were either random (according to the model of Bruce et al., 1999b) or fixed (parameters of neuron 3-21 in Bruce et al., 1999b).

(4) A random component was added to the current distribution to simulate the effects of current decay that varied as a result of a non-homogeneous current path impedance.

(5) The distance between the neural plane and the electrodes was either fixed at 0.5 mm, or varied linearly between 2 mm (near the base) and 0.1 mm (near the apex). These values are not realistic, but were chosen to magnify the effect of the relative distance between the neural plane and the electrodes.

(6) In one embodiment of the model, gap threshold was always calculated from equation 4.17 (which assumes less than 100% entrainment). In a second embodiment, equation 4.11 was used when entrainment was 100%, and equation 4.17 when entrainment was less than 100%. These two conditions are identified as a single process model and dual process model respectively.

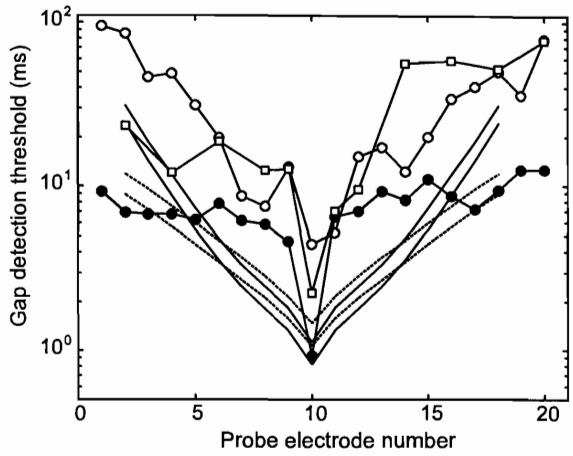

Figure 4.10

Model predictions for gap detection thresholds for a pulsatile stimulus of 55 dB μA. The neural parameter RS was either 0.3 (solid lines) or 0.5 (dashes). In these two sets of curves, either M=1 (upper curve) or M=150 (lower curve). A single process model was employed, and the distance between the electrode plane and neural plane was 0 mm. The exponential current decay was 4 dB/mm. The data with electrode 10 as standard for subjects N3 (○), N4 (●) and N7 (□) are also shown.

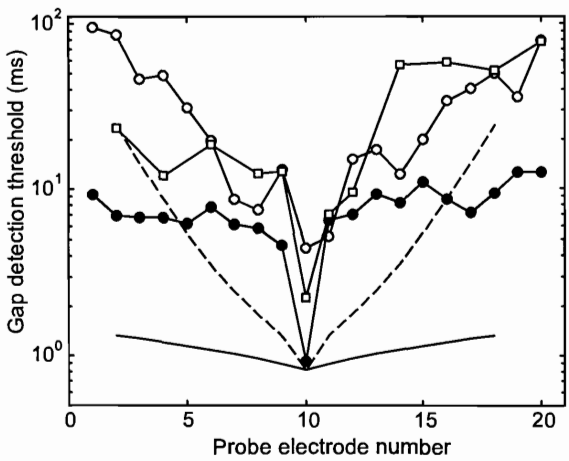

Figure 4.11

This figure shows the effect of a different rate of exponential current decay. The solid curve is for a current decay of 0.5 dB/mm, while the dashed curve is for 4 dB/mm. M=150 in both model curves, RS=0.3, and the subject data are also shown.

Figure 4.10 shows model predictions with M=1 and M=150, for a stimulus of 55 dB μA. A single process model was employed, and the distance between the electrode plane and neural plane was 0 mm. The exponential current decay was 4 dB/mm, and the neural parameter RS was (for a long-term deaf ear) either 0.3 (solid lines) or 0.5 (dashes). The Neyman-Pearson detection threshold was calculated to obtain a false alarm probability of 0.05. The gap detection tuning curves for N3, N4 and N7, with electrode 10 as standard (from Hanekom and

Shannon, 1998, or chapter 2), are also shown. The model-predicted gap tuning curves do not show the sharp tip seen in the data of N4 and N7, but has the shallow bowl shape of the N3 data.

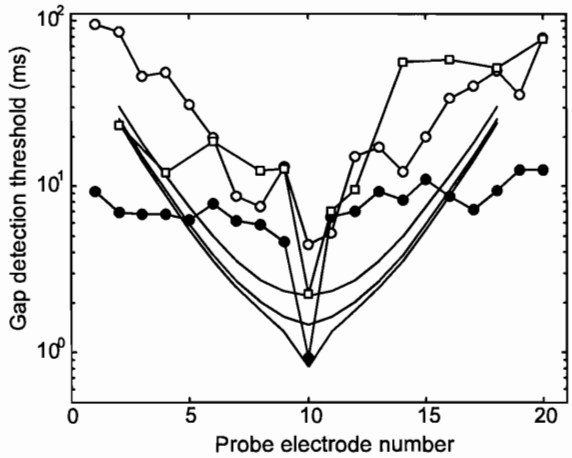

Figure 4.12

This figure shows the effect of the distance between the neural plane and the electrode plane, for a non-varying distance between the electrode array and the neural plane. From top to bottom, the model-predicted curves are for distance y of 1mm, 0.5 mm and 0 mm. Exponential current decay was 4 dB/mm, the stimulus intensity was 55 dB μA, and M=150.

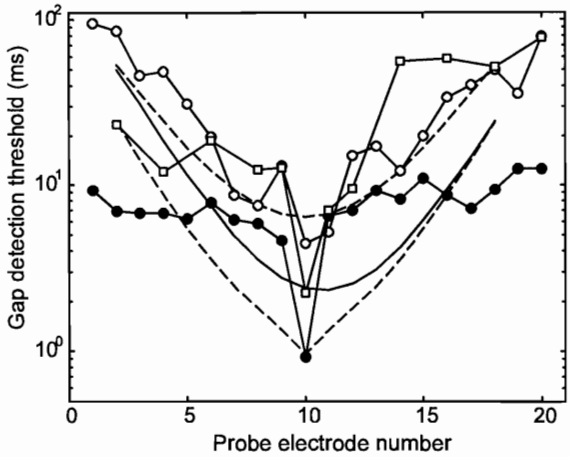

Figure 4.13

The distance between the electrode array and the neural plane varied along the length of the array in this figure (solid curve). The distance varied linearly between 2 mm for the most basal electrode to 0.1 mm for the most apical electrode. This is compared to fixed electrode array distances of 0.1 mm (lower dashed curve) and 1 mm (upper dashed curve).

Figure 4.11 shows the effect of a different rate of exponential current decay. M=150 in both model curves, and the subject data are also shown. The slope of the 0.5 dB/mm model curve is similar to the slope of N4's data, but does not show the sharp tip. A current decay of 4 dB/mm is typical of bipolar stimulation, while 0.5 dB/mm is more typical of monopolar stimulation. The N4 gap detection data exhibit trends that can be predicted by fast current decay (> 8 dB/mm) near the electrode, but to predict trends in the far field, a current decay

more typical of monopolar stimulation is required.

Figure 4.12 shows the effect of the distance between the neural plane and the electrode plane, assuming a non-varying distance between the electrode array and the neural plane. More distant positions for the electrode array result in larger gap detection thresholds, and the sharp tip is not observed for more distant positions.

The distance between the electrode array and the neural plane varied along the length of the array in Figure 4.13. The distance varied linearly between 2 mm for the most basal electrode to 0.1 mm for the most apical electrode. This is compared with fixed electrode array distances of 0.1 mm and 1 mm. The electrode array with linearly varied position is shown to deform the gap tuning curve towards the apex.

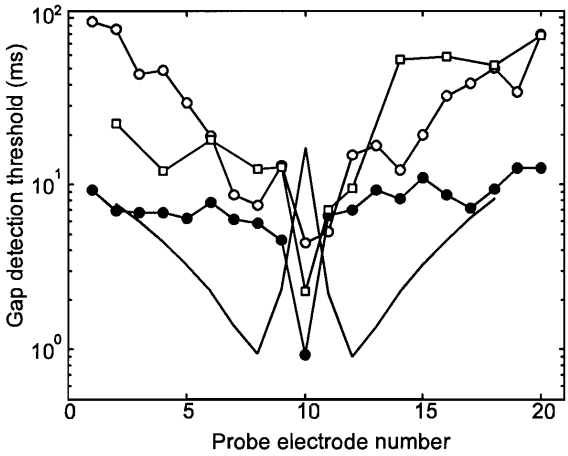

Figure 4.14

Model-predicted gap detection tuning curves when the twin-peaked current distributions predicted by the model of Hanekom (2001) are used. Model parameters are M=150, and Neyman-Pearson threshold=0.1.

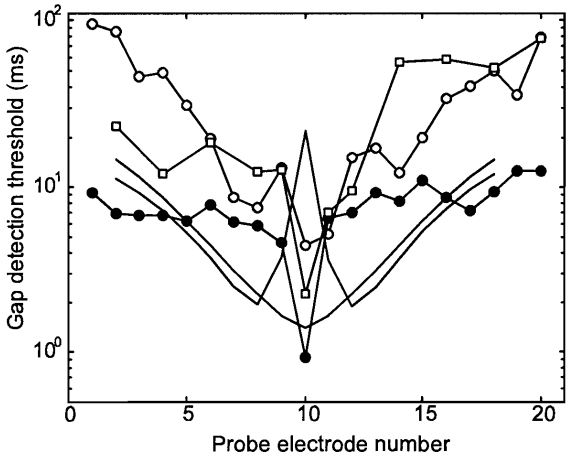

Figure 4.15

The gap tuning curve predicted by the Hanekom current distribution model is compared to a decay of 6 dB/mm.

The twin-peaked current distributions predicted by the model of Hanekom (2001) result in the gap detection tuning curves in figure 4.14. Model parameters are M=150, and Neyman-Pearson threshold=0.1. The twin-peaked current distribution results in a twin-tipped gap detection tuning curve with large gap thresholds when the standard and probe are on the same electrodes. This is not consistent with the trends observed in the data. The flanks of the gap tuning curves predicted by the Hanekom model exhibit a 6 dB/mm decay, as shown in figure 4.15 where it is compared to a single-peaked exponential decay of 6 dB/mm.

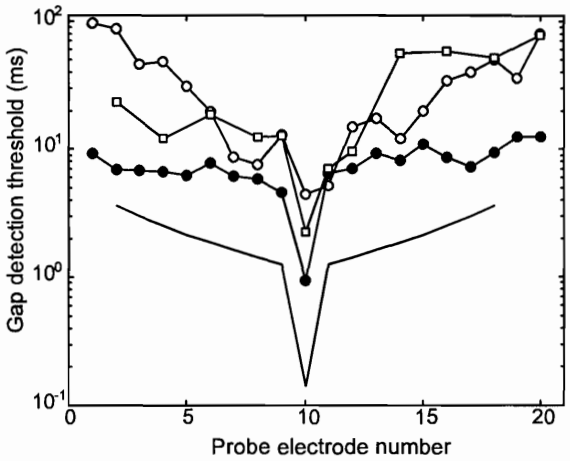

Figure 4.16

The dual process model predicts the sharp tip at high intensity (57 dB μA).

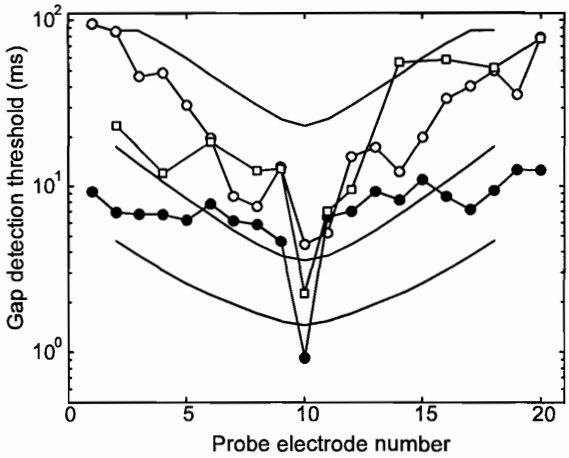

Figure 4.17

The effect of neural threshold is shown at fixed stimulation intensity of 50 dB μA, with M=1 and RS=0.3. From top to bottom, the neural threshold was 55 dB μA, 49 dB μA and 45 dB μA.

The dual process model predicts the sharp tip at high intensity (figure 4.16). A 2 dB/mm decay is used to obtain the simulation results in figures 4.16 and 4.17. This current distribution produced tuning curve flanks consistent with the data of N7. The dual process model is employed in figure 4.17 and M=1. The effect of neural threshold is shown at fixed stimulation intensity of 50 dB μA. As neural threshold increases, the gap threshold also increases, as expected.

With a 2 dB/mm current decay, different values of stimulation intensity result in gap tuning curves that show some of the trends observed in the data (figure 4.18). At lower intensities, the model predicts a bowl-shaped curve similar to the data of N3. At higher intensities, the model predicts the sharp tip seen in the data of N4 and N7.

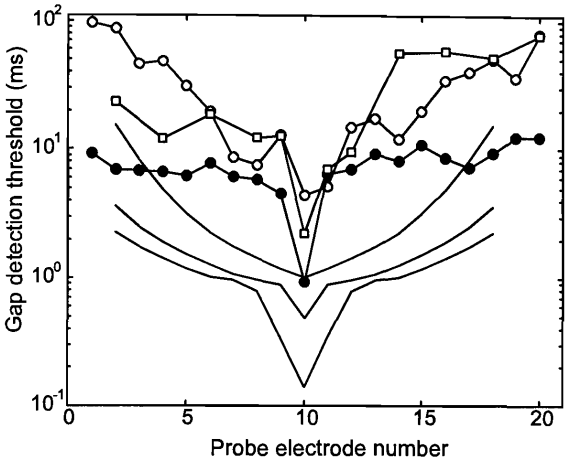

Figure 4.18

The effect of intensity of stimulation on predicted gap thresholds. From top to bottom, stimulus current was 53 dB μA, 55 dB μA and 57 dB μA. Other parameters are y=0, RS=0.151, current decay is 2 dB/mm, M=1.

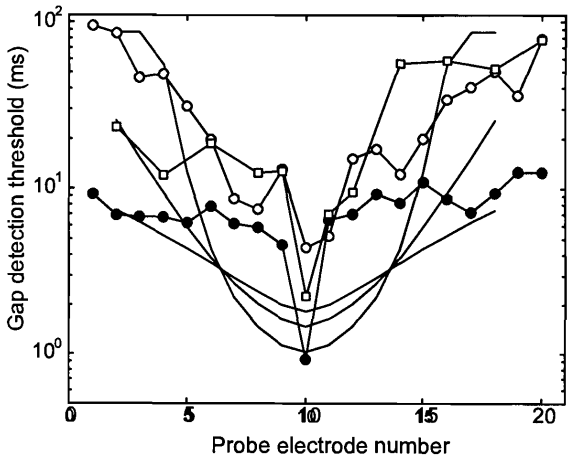

Figure 4.19

The influence of neural parameter RS (relative spread) is shown in this figure. The shallow curve is for RS=0.6, the steepest curve is for RS=0.15, and the third curve is for RS=0.3. Other parameters are M=150, stimulation intensity is 55 dB μA, and current decay is 4 dB/mm.

A likely explanation is that the electrode array for N3 is more distant from the neural plane than that of N4. The argument is that the gap tuning curves for N3 and N4 where obtained at similar loudness levels. Possibly then, the N3 gap tuning curve may also have shown the sharp tip at higher intensities. This in turn suggests that the current distribution for N3 is wider, although figure 4.11 suggests a wider (far field) current distribution for N4 (or slower far field current decay). However, it is also seen that the slopes on the flanks of the gap tuning curves are shallower at higher stimulation intensities, similar to the trend in N4's data, which supports the idea that N4's array is closer to the nerve fibres.

The neural parameter RS (relative spread) influences the fibre rate-intensity curve and therefore affects the shape of the gap tuning curve (figure 4.19). When RS is larger, the tip of the gap tuning curve is sharper, and the slopes on the flanks of the curves are steeper.

When the neural parameters are allowed to be random (in accordance with Bruce et al., 1999b), the gap tuning curves show the jagged trends seen in the data. When the neural threshold is allowed to vary over a 10 dB range, large variance is seen in the gap tuning curves (figure 4.20), while a range of 1 dB or 5 dB simulates the trends in the data more closely (figure 4.21).

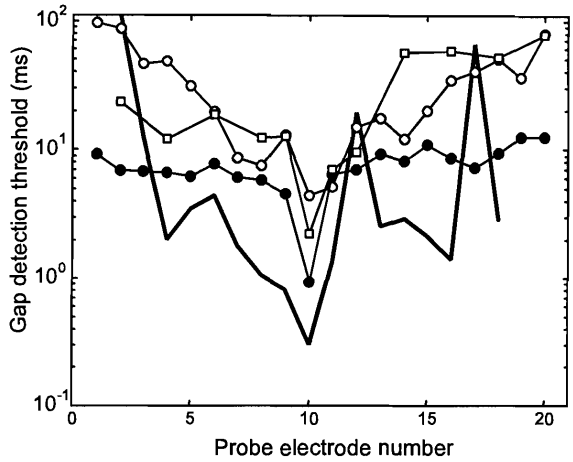

Figure 4.20

The neural threshold varies over a 10 dB range. Other parameters are M=150, stimulation intensity is 50 dB μA, and current decay is 4 dB/mm.

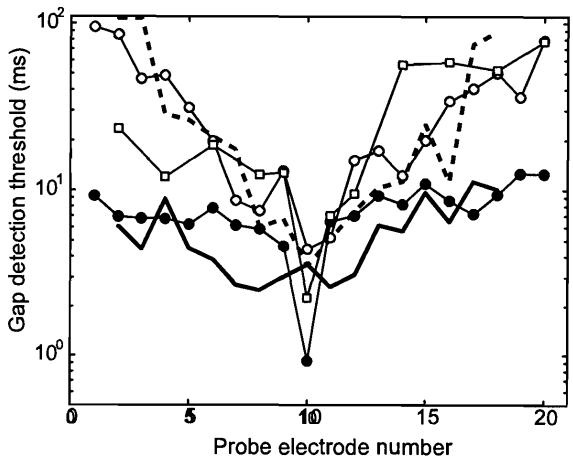

Figure 4.21

The neural threshold varies over a 5 dB range in this figure. Other parameters are M=150, stimulation intensity is 50 dB μA, and current decay is 1 dB/mm (solid line) or 4 dB/mm (dashed line).

Using fixed neural parameters, but adding a random component of standard deviation of just 1 dB to the current distribution, it is seen in figure 4.22 that the non-monotonicity of the data can also be predicted by variation in current decay. The figure shows that the trends of data of N3 are predicted (solid line) by a general trend of 4 dB/mm current decay, with a 1 dB standard deviation in the current that reaches the (fixed threshold) fibres. Non-monotonic

current decay can result from non-homogeneous impedance paths. The general trends of N4's data are predicted (dashed line) by current decay of 1 dB/mm (with a 1 dB standard deviation in the current).

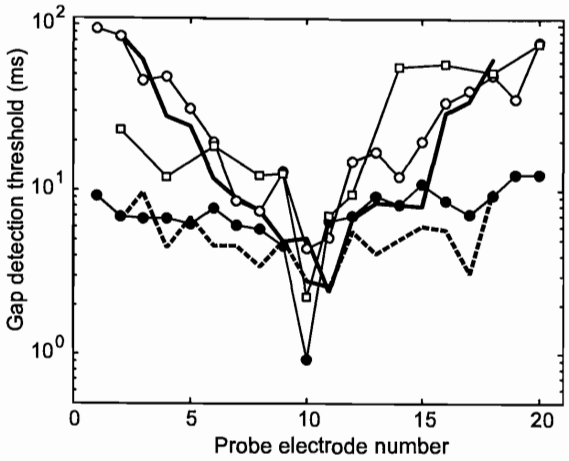

Figure 4.22

A random component of standard deviation of 1 dB is added to the exponential decay current distribution. The solid line has a 4 dB/mm exponential current decay, and the dashed has a current decay of 1 dB/mm.

3.2 Current distributions predicted by model

Figures 4.23 to 4.25 show the electrode current required by the model to obtain the measured gap tuning curves for the three subjects. These curves assumed an electrode-neural distance that varied across the length of the electrode array. A general trend in these curves is that larger currents are required to obtain the measured gap detection thresholds when the electrode is more distant from the nerve fibres. However, the curves are non-monotonic and all three curves have a peak at the standard electrode, so that larger currents are required to obtain the small values of gap threshold that are generally observed when both markers are on the same electrode.

From these curves, the current distribution is calculated (figures 4.26 to 4.28), i.e. for a fixed

electrode stimulation current, these curves show the calculated current at a number of positions on the neural plane. More specifically, the model can only calculate the currents at the set of positions of observation windows used by the model.

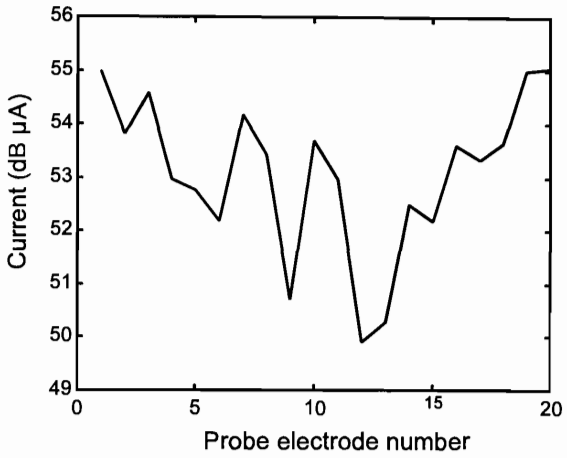

Figure 4.23

This figure shows the electrode current that is required in the model to obtain the measured (Hanekom and Shannon, 1998) gap tuning curves for subject N3. Model parameters are as follows. M=1 and RS=0.3, while the model assumed a 4 dB/mm current decay.

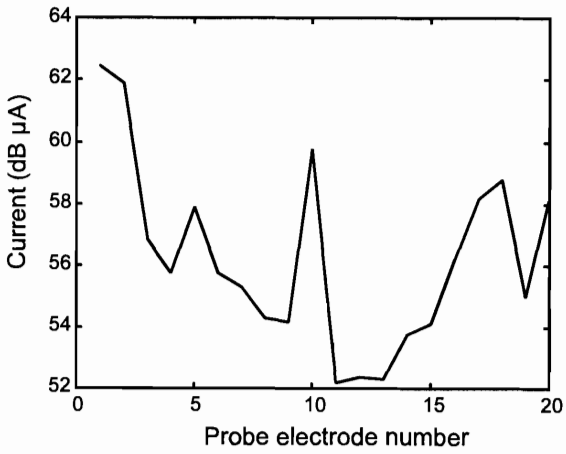

Figure 4.24

Same as figure 4.23, but for subject N4.

The current distribution curves obtained depend on model assumptions. In figures 4.26 to 4.28, an electrode array with varying distance y to the neural plane was assumed. RS=0.3 was used in these simulations. A numerical value for the current decay model parameter (e.g. 4 dB/mm) is required by the model to calculate the gap thresholds in the optimization routine that solves for the required current. Different choices of this parameter lead to slightly different predictions for the current distribution. In calculation of the current spread, the current decay parameter will specify only the general trends in the data. As the results show, the actual (model-calculated) rate of decay is non-monotonic and not necessarily close to 4 dB/mm. This is as a result of variations in current pathways and non-homogeneous impedance.

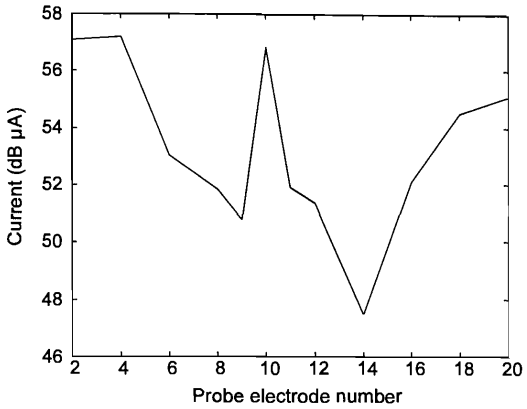

Figure 4.25

Same as figure 4.23, but for subject N7.

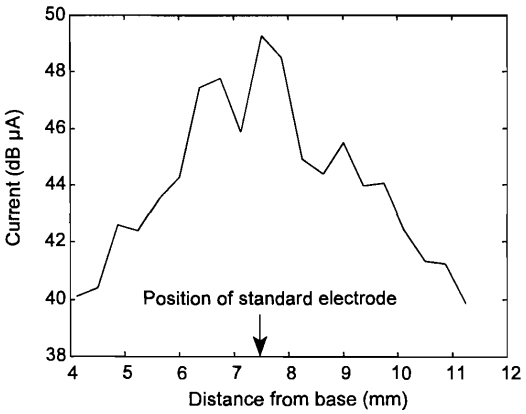

Figure 4.26

The current distribution curve for N3 as calculated from figure 4.23, assuming an electrode array with varying distance y to the neural plane.

Figure 4.26 shows a wide, unfocussed current distribution for N3, with a rate of current decay of 2.7 dB/mm (towards base) and 2.5 dB/mm (towards apex). N4 has a sharp current peak at fibres closest to the standard electrode position and then a gradual decay of current towards the more distant nerve fibres (< 1 dB/mm). Hence, N4 has sharply focussed stimulation close to the electrode, but also a wide activation pattern with higher levels of current than N3 and N7 still present far from the electrode (figure 4.29).

N7 also has a sharply focussed current distribution close to the electrode (8 dB/mm towards the apex), and a somewhat more gradual decay in current of 3.2 dB/mm (towards the apex) and 2.67 dB/mm (towards the base) when measured from the tip to the tail on either side of the tip.

The current intensity does not decay more than 14 dB for any of the subjects even at the most distant nerve fibres modelled. Predictions are not available throughout for the current at the –10 dB point, as the current has not always decayed that much. As a measure of the focussing of current, the –6 dB focussing width and the –10 dB focussing width are calculated. Curves

were extrapolated if the current did not decay by 10 dB or more to obtain the values for focussing width. The –6 dB focussing widths for the subjects are 4.4 mm (N3), 0.7 mm (N4) and 1.1 mm (N7). The –10 dB focussing widths for the subjects are 8 mm (N3), 5.3 mm (N4) and 5.3 mm (N7).

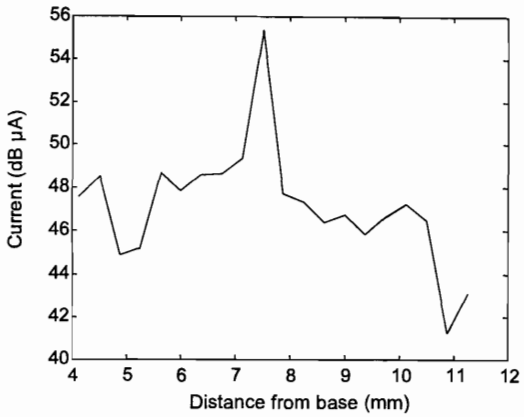

Figure 4.27

Same as figure 4.26, but for subject N4 and calculated from figure 4.24.

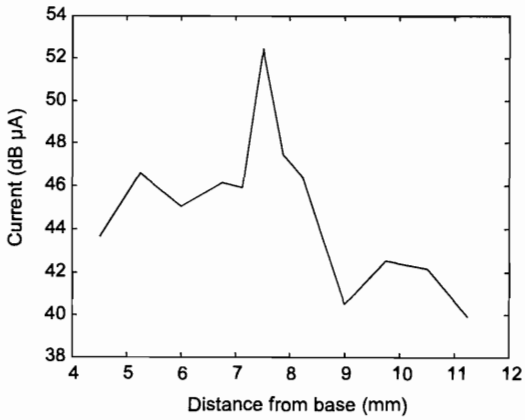

Figure 4.28

Same as figure 4.26, but for subject N7 and calculated from figure 4.25.

With different model assumptions than those for figures 4.26 to 4.28, different model-calculated current distribution curves are obtained. The predicted current distribution does not change by more than 5% when the electrode array is at fixed distance of 0.5 mm from the neural plane, while the trends remain the same. However, when the model assumption for the current decay rate is different, the predicted current distributions are notably different to those obtained in figures 4.26 to 4.28. A current decay of 4 dB/mm appears to model the gap data for N3 quite well (figure 4.22), but N4 has a current decay of closer to 1 dB/mm on the flanks (figure 4.22), while the N7 data is probably modelled better by a 1 dB/mm decay on the base flank and a 4 dB/mm decay on the apex flank of the data. Using 1 dB/mm for both the N4 and N7 data, the current distribution curves in figure 4.30 and 4.31 are obtained.

Particularly notable in these model-calculated current distribution curves are the widely varying patterns of predicted current distribution in these three subjects. Many psychophysical studies

have reported considerable inter-subject variability and this is regarded as a major obstacle in the way of progress in development of more advanced cochlear implants.

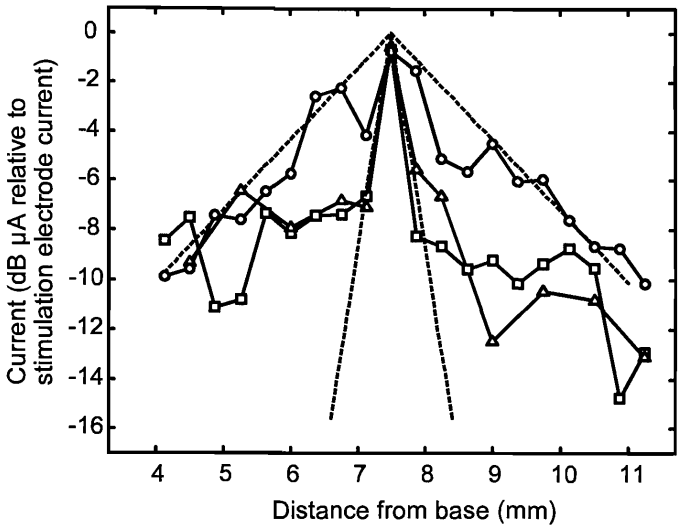

Figure 4.29

Current distributions for all three subjects plotted on the same graph (circles: N3; squares: N4; triangles: N7). Current decay is plotted relative to the current that reached the fibres closest to the standard electrode. Also shown (dashed lines) are two exponential current decay curves with length constants of 0.5 mm and 3 mm respectively (decay of 8.7 dB/mm and 1.4 dB/mm respectively).

Published data for *current* distributions in cochlear electrical stimulation are scarce. More often, *voltage* distributions are measured (Ifukube and White, 1987; Spelman et al., 1995) or modelled (Hanekom, 2001). However, as was shown by Black et al. (1983) and Kral et al. (1998), voltage distributions do not necessarily reflect the current distributions at the neural excitation sites. This is because of non-homogenous impedance in the cochlea. Voltage distributions cause currents and these currents excite nerve fibres. Accordingly, using current distributions to predict which fibres will be excited is more appropriate.

It is noted, however, that some more complex voltage distribution models (Frijns et al., 1995;

Hanekom, 2001) calculate nerve excitation directly from the voltage distributions via finite difference or finite element models that take the non-homogeneous cochlear impedances into account and via Hodgkin-Huxley (or equivalent) nerve fibre models that predict when fibres will fire.

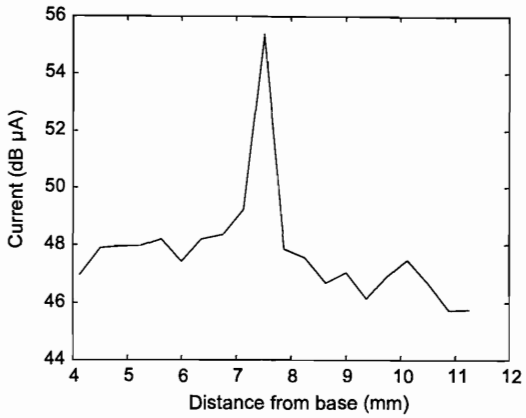

Figure 4.30

With different model assumptions than in figure 4.24, the predicted current distribution for subject N4 is as shown in this figure. The model now assumes a 1 dB/mm current decay.

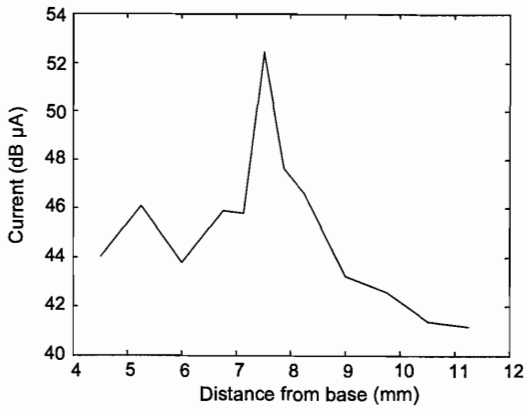

Figure 4.31

Same as figure 4.30, but for subject N7.

As a further note, spatial excitation data are also usually given as threshold data rather than as spatial excitation profiles. Several studies measured the thresholds of fibres at various cochlear positions in response to electrical stimulation at a fixed position in the cochlea (e.g. Van den Honert and Stypulkowski, 1987b; Shepherd and Javel, 1997), but no data on supra-threshold spatial excitation patterns appear to be available.

Figure 4.32 compares the model-calculated current distribution curves to current distribution data from Black et al. (1983). These data were not obtained in vivo, but were measured in a tank model (a 5 mm saline-filled tube). The data show a current decay length constant of 0.7 mm at the tip, which is the same as the length constant obtained by Kral et al. (1998) in their measurements. Close to the peaks the model-calculated current distributions for N4 and N7

follow the data quite well. The tails of the calculated current distributions are elevated by 5-10 dB relative to the tank data.

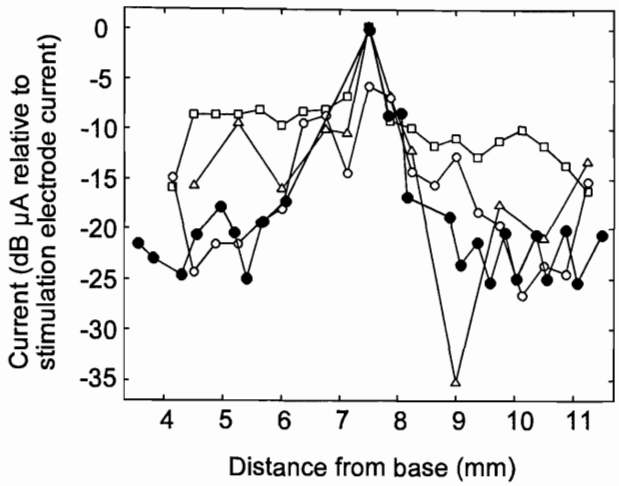

Figure 4.32

This figure compares the model-calculated current distribution curves to measured current distribution data. Filled circles are data from Black et al. (1983). Open symbols are predicted current distributions for the three subjects (open circles: N3; open squares: N4; open triangles: N7).

3.3 Predictions for electrode discrimination from the predicted current distributions

Predictions for current distributions were obtained for all the electrodes for which gap detection tuning curves were available for all three subjects. Predictions were then obtained for the discriminability of electrodes. This was done as follows. The average and standard deviation of each electrode's current distribution was required. As the rate of current decay was different towards the base than towards the apex, the standard deviation was calculated on either side of the peak current. Electrode discriminability was then calculated from

$$d' = \sqrt{\frac{2(m_2 - m_1)^2}{\sigma_{E1}^2 + \sigma_{E2}^2}}, \qquad (4.25)$$

where m_1 and m_2 are the active electrode positions (in mm from the base), and σ_{E1} and σ_{E2} are the standard deviations of the current spread on the appropriate flanks of the current distribution curves of the two electrodes. The predicted electrode discriminability is shown as two-dimensional contour plots (figure 4.33 and 4.35) for N4 and N3 respectively. These figures show how discriminability varies with distance between the electrode pairs. In both cases, a distance of at least two electrodes (in BP+1 mode in the Nucleus cochlear implant, see chapter 2) is required to achieve d' > 1 (defined as the discrimination threshold).

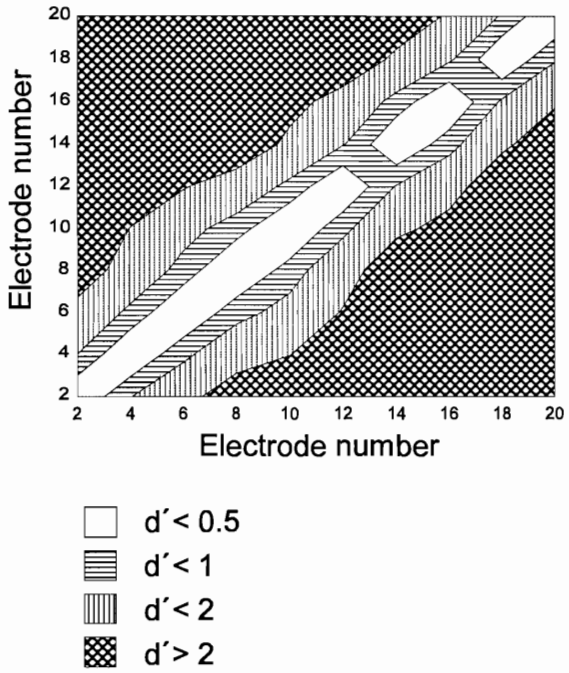

Figure 4.33

Predicted electrode discriminability for N4 in BP+1 mode.

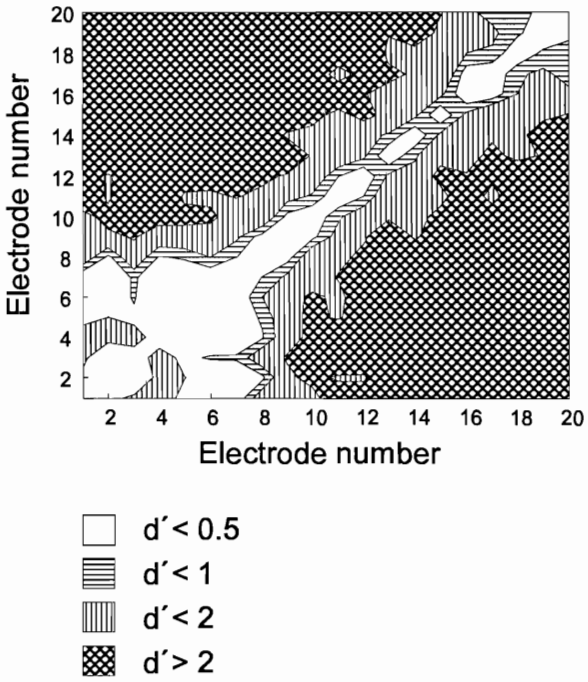

Figure 4.34

Electrode discriminability for N4 (BP+1 mode) measured in a pitch discrimination experiment (Hanekom and Shannon, 1996).

This means that electrodes need to be spaced at least 1.5 mm apart (in the Nucleus cochlear implant, see chapter 2) to be discriminable. As the two electrodes in an electrode pair is 1.5 mm

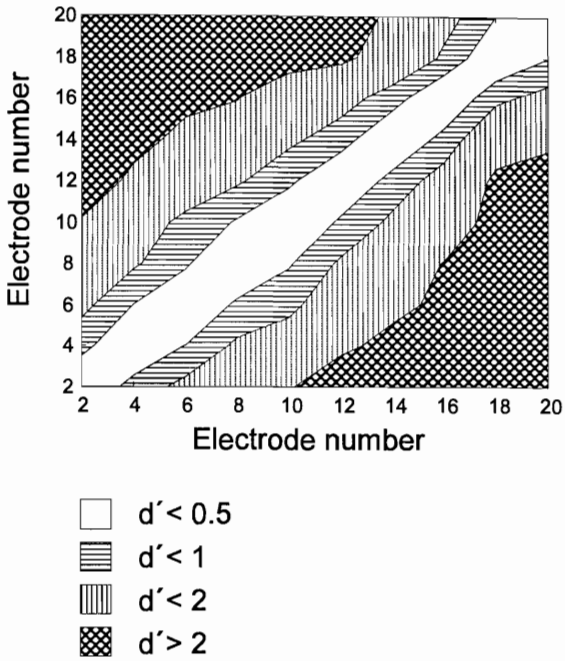

Figure 4.35

Predicted electrode discriminability for N3 in BP+1 mode.

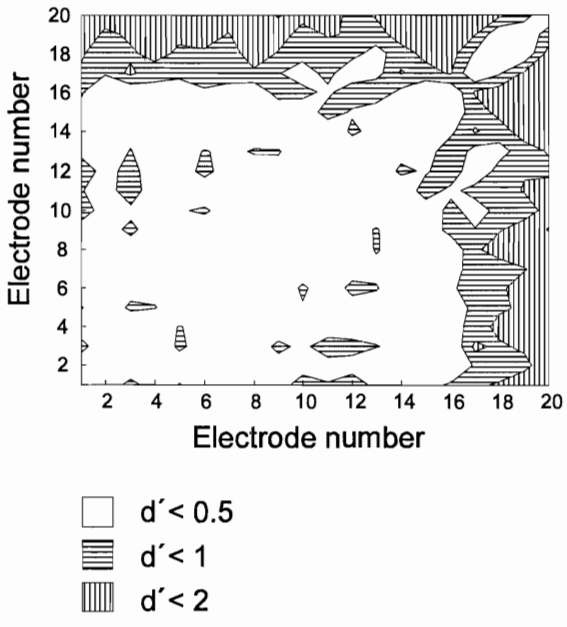

Figure 4.36

Electrode discriminability for N3 (BP+1 mode) measured in a pitch discrimination experiment (Hanekom and Shannon, 1996).

apart, this means that stimuli on overlapping electrode pairs will be poorly discriminated. This is probably because these electrode pairs stimulate the same neural population to a large extent. The predictions in figures 4.33 and 4.35 also suggest that electrodes with closer spacing will be more discriminable for N4 than for N3.

Predicted discriminability of electrodes is compared with electrode discrimination data (figures 3.34 and 4.36) for the same subjects (Hanekom and Shannon, 1996). The predicted discriminability correlates well to the measured electrode discrimination data for N4, but not for N3 or N7 (not shown). The data for N4 suggests that a separation of at least two electrodes

is required to achieve d' > 1, as is also predicted by the model. This suggests that the current distributions predicted by the model is a fair reflection of the actual current distributions for N4.

N3 (and N7, not shown) should have performed better on the electrode discrimination task than what is reflected in the data in figure 4.36 if the current distribution predictions are accurate. Other factors may have played a role in N3's case (and in N7's) in the electrode discrimination task. The actual task was a pitch discrimination task, and it is possible that electrodes were discriminable, but that these subjects were not able to judge pitch. Anecdotaly, N7 reported that he "has always been tone deaf". Thus, the electrode discrimination data do not provide conclusive evidence that the predicted current distributions are correct.

4 DISCUSSION

4.1 Modelling of gap detection data

The model can predict both the magnitude of gap thresholds and the U-shaped curves with the correct choice of parameters. The trends in the slopes on the flanks of the predicted gap thresholds curves follow those seen in the data, and depend primarily on the current distribution and rate of current decay. Realistic choices for rate of decay can predict the slopes.

If an exponential decay is assumed, the model cannot predict the sharp tip seen in the data of N4 and N7, but the shallow bowl is predicted. It is possible that two separate mechanisms exist that determine gap thresholds, so that gap thresholds are very small in the within-channel condition and larger in the across-channel condition. It is possible that a perceptual similarity process plays a role in the across-channel condition as suggested in chapter 2.

However, the model shows that hypothesizing different mechanisms for across-channel and within-channel conditions is not necessary. When entrainment is close to 100%, gap thresholds are determined by the spike position jitter (spike position standard deviation) alone. Gap

thresholds are then expected to be small for pulsatile electrical stimulation, as temporal jitter of spikes is small. This situation is most likely to occur when the two markers are presented on the same electrode.

When entrainment in the observation window is lower that 100%, gap thresholds are primarily determined by the standard deviation of the inter-spike interval pdf and the temporal jitter of spikes does not play an important role. The standard deviation of the inter-spike interval pdf for less than 100% entrainment is typically much larger than temporal jitter, so that gap thresholds are expected to be much larger. Figures 2.5 to 2.7 provides further support for this notion. The sharp tips in the gap detection tuning curves disappear at lower stimulation levels where entrainment should be below 100%. Thus it appears that this model provides a plausible explanation for the origin of the gap detection tuning curves.

Finally, it is not clear why the twin-peaked voltage distribution predicted by Hanekom (2001) fails to be a good predictor for gap thresholds. A less than satisfying explanation is that the nonhomogeneous cochlear impedance introduces enough variability in the current arriving at different fibres to mask the twin-peaked pattern.

4.2 Modelling of current distributions

Predicted current distribution profiles are consistent with tank measurements (Black et al., 1983) of current distribution (figure 4.32) for the first millimeter on either side of the current peaks and the trends are similar throughout the modelled region. Exponential current decay cannot explain the sharp tip, but a more sharply focussed current distribution can explain the sharp tip (see figures 4.26 to 4.28). Differences between model predictions and tank data in the tail sections of the current distributions may be due to differences between the nonhomogeneous cochlear impedances and homogeneous tank impedances. Shepherd and Javel (1997) remarked that current spread may increase in long-term deafness due to demyelination of nerve fibres. The length constant of the predicted current distributions of N4 and N7 is 0.5 mm near the tip, which is comparable to the length constants measured by Kral

et al. (1998). The length constant of the current distribution of N3 is 3 mm, which is also the length constant measured by Black et al. (1983).

Interesting results are presented in figures 4.29 and 4.33. Close to the electrodes, predicted current distributions for N4 and N7 show current decays that are typical of bipolar stimulation, while the current decay of N3 is more reminiscent of monopolar stimulation. The data of N4 and N7 show sharp (predicted) current peaks close to the electrodes, with current decay very similar to what has been obtained in a tank model. The predicted current distribution tails more distant from the electrode follows the trend seen in the tank model data.

The wide current distribution pattern of N3 is reflected in the electrode discriminability data shown in figure 4.36, while the more sharply focussed current distribution (close to the stimulation electrodes) of N4 is reflected in the electrode discriminability data shown in figure 4.34. Even though the tail in the predicted current distribution of N4 is equally far below the current peak than N3's predicted current distribution, N4 has much better electrode discriminability. Thus it appears the tail of the current distribution does not contribute to the listener's ability to discriminate electrodes. This emphasizes the importance of current focussing techniques in stimulus pattern design (Townshend et al., 1987) and electrode designs to obtain better current focussing (Cords et al., 2000).

It is clear from figure 4.36 that N3 had difficulty in discriminating closely spaced electrodes, which may be ascribed to the wide current distribution. Hanekom and Shannon (1996) endeavoured to relate place pitch discrimination data to current spread and the definition of information channels in electric hearing. While the measurement of electrode discrimination relied on the pitch discrimination ability of subjects in the experiments of Hanekom and Shannon, gap detection may provide a more objective approach to arrive at predictions for current distributions. The similarity between the trends seen in the predicted current distributions and the electrode discrimination data is encouraging.

There is, however, still a large discrepancy between the predicted electrode discriminability

pattern and the place pitch discrimination data (figures 4.35 and 4.36). It is possible that the pitch discrimination experiment of Hanekom and Shannon (1996) underestimated the ability of subjects who had poor pitch perception to discriminate electrodes. Experiments that are designed for electrode discrimination (e.g. Collins et al., 1997) rather than for place pitch discrimination would have been more appropriate for comparisons between predictions and data. It is possible that electrode discrimination is better than reflected in the place pitch discrimination data.

As a final remark, predicted current distributions vary widely between subjects, which may be a major factor determining speech recognition ability in cochlear implant users.

4.3 Applicability of the model

The model described in this chapter is only applicable for auditory electrical stimulation, and specifically the situation where the gap is defined as in figure 4.1. If stimulation pulses can occur in the gap, i.e. when the gap is not silent, the model will not necessarily model the task of the central detection mechanism very well. Also, when spikes can occur during the gap (in a model with spontaneous activity), the task changes. These situations were investigated in paragraph 4.2.2. The model is applicable in a gap discrimination situation, but as no gap discrimination data for cochlear electrical stimulation is available, model predictions for the discrimination task were not explored further.

4.4 Free parameters and parameter sensitivity

The model has a small number of free parameters that control different aspects of the gap threshold tuning curves. The U-shaped gap tuning curves and correct magnitude of gap thresholds are obtained by the model primarily because spike probability decreases as gap marker electrodes are separated. This results in larger standard deviation in the pdfs in the signal detection task, which in turn results in larger gap thresholds.

The number of fibres combined in a central detector and the electrode-nerve distance can be varied, but the model is not very sensitive to these parameters. The model suggests that variations in electrode-nerve distance may slightly skew the gap tuning curves, but is not as important as current distribution. This statement is based on the assumption that fibres are excited at a fixed threshold. However, the model of Hanekom (2001) shows that apparent fibre thresholds change considerably when electrodes are placed close to the modiolus.

The model suggests that the major factor determining the gap threshold tuning curve shapes is the current distribution, which depends on electrode design and cochlear impedance characteristics. Exponential current decay at different rates can control the slopes of the gap tuning curve flanks. Other irregular current distributions can predict the jagged shape of the gap tuning curves.

Neural parameters (threshold and RS) can be varied and, as has been shown, the variability of these can also predict the non-monotonic shape of the gap tuning curves. RS controls the slope of the gap tuning curve flanks. Finally, stimulation parameters (intensity of stimulation, pulse width, and frequency of stimulation) can be varied. The magnitude as well as the shape of predicted gap thresholds is sensitive to the intensity of stimulation.

4.5 Strengths and weaknesses of the current model

4.5.1 Strengths

As far as is known, this is the first model for gap detection in electric hearing. The model provides a simple explanation for gap detection thresholds as measured in cochlear implantees, using realistic model parameters. Gap detection threshold magnitudes are correct, and the U-shaped curves can be predicted. It is based on simple signal detection theory considerations. The model implementation performs analytical calculations and does not require lengthy Monte Carlo simulations.

The model also provides estimates of the current distributions based on psychophysical gap

detection data. Hence, current distributions may be estimated non-invasively. Evaluating whether the predicted current distributions are a fair reflection of the actual current distributions is impossible with currently available techniques. Tank measurements (Black et al., 1983) show similar trends than observed in the model predictions. If it can be proven that these predictions are good estimates of the actual current distributions, the model will be a particularly useful tool to individualise settings in cochlear implant programming.

4.5.2 Weaknesses

Three important criticisms of the current model are discussed here. First, a primary model assumption is that a limited extent observation (or attention) window is employed by the central gap detection mechanism. Not enough neurophysiological evidence is available to support this assumption. In fact, gap detection studies in normal acoustic hearing have shown that gap detection improves when the gap is present in more neural channels (Hall et al., 1996), although a similar result was not obtained in cochlear implantees in the study by Van Wieringen and Wouters, 1999.

Second, model assumptions about the neural spike train may be an oversimplification. It is assumed that only the *A* response occurs, while the *B* and *C* responses may also occur and complicate the central gap detector's task. However, it is believed that for modelling purposes it is reasonable to assume that only the *A* response occurs, as this response is predominant in long-term deaf ears. Occurrence of multiple spikes with different latencies in response to an electrical stimulus pulse will complicate the task of the central detector and makes it difficult to obtain model predictions for gap thresholds without reverting to Monte Carlo modelling. Creating such a model is possible. However, conceptually these multiple spikes just decrease the signal to noise ratio (i.e. they "fill the gap"), so that gaps become more difficult to detect and gap thresholds increase.

Finally, the signal detection calculations were based on gamma functions that approximated the multi-mode inter-spike interval histograms. This was to simplify calculations, but will influence the predicted gap thresholds.

4.5.3 Other arguments against the model

Chatterjee et al. (1998) and Van Wieringen and Wouters (1999) have argued against the notion that U-shaped gap detection curves originate from an across-channel gap detection mechanism where smaller neural overlap leads to increased gap thresholds.

One argument was that the same U-shaped curves can also be obtained with within-channel gap detection by either using markers with different stimulation pulse rate or intensity of stimulation (Chatterjee et al., 1998). This result will also be obtained with the current model. Although no calculations have been made, it is conceptually easy to see that the signal detection task is complicated by markers that differ in frequency, as gap detection is based on inter-spike interval duration. In the model, the gap detection mechanism searches for a change in inter-spike interval duration and detects a gap when this occurs. Inter-spike intervals that change without a gap occurring, confounds the gap detection mechanism, which will result in increased gap thresholds. So, although Chatterjee et al. (1998) are correct to suggest that within-channel processes can also result in U-shaped gap tuning curves, this does not negate the current model.

The role of perceptual dissimilarity processes cannot be precluded. U-shaped gap tuning curves obtained with differences in marker loudness (Chatterjee et al., 1998) are explained more easily by perceptual dissimilarity than by the current model. Van Wieringen and Wouters (1999) also argued that perceptual processes probably determine gap thresholds. A strong argument by these authors was that training eliminated the increase in gap thresholds in across-channel conditions in some subjects.

4.6 Future improvements of the model

One model limitation is the way that information is combined across a number of fibres that carry information about the gap. It has been assumed that spike train information from one critical band is combined to obtain predictions of gap thresholds. Under the (valid) assumption that these spike trains are very similar, the model simply calculates what the improvement in

signal-to-noise ratio is, and predicts gap thresholds under the assumption that spike trains are observed by a classical detector.

The model does not take into account that spike trains from widely-spaced positions in the cochlea may be combined to obtain predictions for gap thresholds. One future improvement will be that the model will incorporate mechanisms to combine non-similar spike trains from more widely-spaced positions.

Also, the model provides no *implementation* for the detection mechanism, but employs a *statistical* approach based on classical detection theory to provide gap threshold predictions. The model currently does not allow one to have a set of (simulated) spike trains containing the gap as input to a gap detection mechanism, which then has to detect the gap. Bayesian (as opposed to classical) estimation techniques, which employ an internal model of the signal to be detected, are explored in chapter 5 (for frequency discrimination). Chapters 5 to 7 explore the implementation-driven modelling approach, rather than the statistical approach used in this chapter and in chapter 3, by which an implementation for an optimal detector is devised. Such a detector can then use simulated spike trains as input and predictions can be derived with Monte Carlo simulation. The advantage of this approach is that complexities in spike trains (e.g. multiple spikes with multiple latencies) can be investigated.

Finally, the model takes only spatial, and not temporal, information into account. Decay of spike rate does not occur in cochlear electrical stimulation as in acoustic gap detection, but spike latencies are also not taken into account. Furthermore, adaptation and bursting (Shepherd and Javel, 1997) are not taken into account. All these complexities can be incorporated into an implementation-driven model.

5 CONCLUSION

A number of conclusions may be drawn from the modelling exercise in this chapter.

(1) Obtaining realistic gap detection thresholds in electric hearing by a spatial model that

does not allow for temporal models of gap detection is possible.

(2) Predictions for the current distribution in the cochlea of a specific implant user can be obtained from the gap detection tuning curves. It is not known how accurate these predictions are, but they do show trends similar to those found in current distribution measurements.

(3) The sharp tip seen in some gap tuning curves is possibly obtained when entrainment is close to 100%, when the primary factor determining gap thresholds is probably the temporal dispersion of spike placement in response to a stimulus pulse.

(4) The shallow bowl portion of the gap tuning curve is probably obtained when entrainment is not close to 100%, when gap thresholds are probably determined by standard deviation of the inter-spike interval pdf.

(5) Electrode array placement relative to the neural plane plays a secondary role in determining gap thresholds.

(6) The primary factor determining gap thresholds is probably the shape of the current distribution. Modelling results suggest that exponential current decay is not a good model of current distribution in the cochlea. Possibly, for bipolar stimulation, sharper current peaks are obtained close to the electrode.

(7) The model rests on the important assumptions that an attention window or observation window exists, that this window is placed for optimal gap detection probability, and that the gap detection mechanism uses spike train information from this window exclusively. This assumption cannot be proven or disproven by the current model.

APPENDIX 4.A

Equation 4.14 is a polynomial that was fit to obtain the required b values for values of σ_r between 2 ms and 100 ms. The coefficients in this equation are $p_1 = -5.2745 \times 10^{-12}$, $p_2 = 2.1284 \times 10^{-9}$, $p_3 = -3.5133 \times 10^{-9}$, $p_4 = 3.0544 \times 10^{-5}$, $p_5 = -0.0015$, $p_6 = 0.0416$, $p_7 = -0.6214$ and $p_8 = 5.8933$.

The definitions of the gamma function $\Gamma(b)$ and incomplete gamma function $\Gamma(b,a)$ as used in chapter 4 are given in equations 4.A.1 and 4.A.2.

$$\Gamma(b) = \int_0^\infty t^{b-1} e^{-t} dt \tag{4.A.1}$$

$$\Gamma(b,a) = \int_a^\infty t^{b-a} e^{-t} dt \tag{4.A.2}$$

Chapter 5

A SPATIAL MODEL OF FREQUENCY DISCRIMINATION IN ACOUSTIC HEARING

The results in this chapter have previously been published: Hanekom, J.J. 1999, "A model of frequency coding in the central auditory nervous system", *South African Journal of Communication Disorders*, vol. 46, pp. 81-89.

1 INTRODUCTION

A sound stimulus received by the peripheral auditory system is transformed to neural spike train activity in a population of auditory nerve fibres (called a spike train pattern). The auditory system varies a number of parameters of the neural spike train pattern to accurately represent a sound, e.g. average spike rate, spread of excitation over a specific subset of the neural population and synchronization of spikes to the stimulus waveform (Delgutte, 1996). This process of transformation of the original sound stimulus to an internal spike train pattern representation is called coding (Bialek, 1991) and the internal representation of a stimulus is referred to as the neural code for this stimulus (figure 5.1). Two mechanisms known to be involved in frequency coding in the auditory system are rate-place coding and phase lock coding (Delgutte, 1997; Moore and Sek, 1996). In rate-place coding, the auditory system may use the excitation pattern across the entire auditory nerve population to determine the stimulus frequency. Rate-place coding operates over the entire stimulus frequency range, but is dominant for the coding of high frequencies (above about 5000 Hz) (Kim and Parham, 1991; Moore, 1973). Phase lock coding, i.e. synchronization of neural firing rate to individual cycles of a periodic stimulus, is the primary cue used for determination of the frequency of a pure tone at low frequencies. Phase locking is progressively lost as stimulus frequency increases above about 2500 Hz (Delgutte, 1996), although there is evidence that temporal information may be useful up to at least 10000 Hz (Heinz et al., 2001). Both coding mechanisms probably

operate in parallel over a large range of frequencies, but it is not clear yet to which extent the central auditory system uses either mechanism alone or both mechanisms simultaneously in the determination of the frequency of a pure tone (Johnson, 1980). It is however known that at increasingly higher auditory nerve centres more phase-locking is lost and the auditory system relies increasingly on rate-place codes alone (Langner, 1992).

Shofner and Sachs (1986), Kim et al. (1990) and Kim et al. (1991) studied spatial response profiles of the discharges of populations of auditory nerve fibres. A "spatial response profile" or "rate response profile" is simply the spatial distribution of neural responses (average neural firing rates) to single tone stimuli along the length of the cochlea. An earlier study by Kim and Molnar (1979) indicated that the rate response profiles become very broad and exhibit very little tuning at all but very low stimulation intensities (20 dB SPL). However, they did not clearly distinguish between low spontaneous rate and high spontaneous rate fibres in their study. Low spontaneous rate (SR) fibres have wider dynamic range (Shofner and Sachs, 1986) and are more likely candidates for rate coding. Shofner and Sachs studied specifically the rate response profiles at low frequency for very low SR fibres (fibres with SR < one spike per second). These fibres account for about 15% of the afferent auditory nerve population. Shofner and Sachs found that the rate response profiles for these fibres exhibit clear peaks at the stimulation frequency (1500 Hz in their experiments) over a wide range of sound pressure levels (in their experiments, from 34 dB SPL to 87 dB SPL). Kim et al. (1990) observed the same effect with a 1000 Hz stimulus. These studies suggested that the rate-place code operates not only at high frequencies, but also over a wide range of sound pressure levels at low frequencies.

High SR fibres saturate at relatively low stimulation intensities (Kim et al., 1990) and the peak in the excitation pattern at the stimulus frequency is quickly flattened as excitation begins to spread along the length of the cochlea at higher stimulation intensities. Spread of excitation is a cue for loudness (Smith, 1988) and it is therefore fair to assume (for the purposes of this chapter) that high SR fibres are primarily involved in intensity coding. Of course, low and high SR fibres are involved in coding of frequency via phase locking as well. However, this chapter focusses on rate coding only and so only low SR fibres are considered.

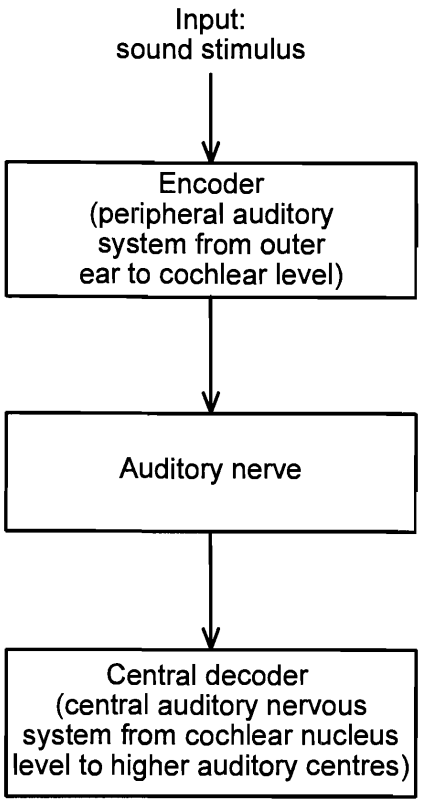

Figure 5.1.

A simple model of the encoding of sound stimuli by the auditory system. The model presented in this chapter describes the encoding process (which primarily takes place in the cochlea) and the decoding process (which takes place in the central auditory nervous system at the cochlear nucleus level and higher). The encoder output is the sound stimulus encoded as a neural spike train pattern.

The articles by Shofner and Sachs (1986), and Kim et al. (1990) established that a pure tone stimulus is represented in the rate code, but it is not actually known whether the rate information is utilized by the central auditory system for the extraction of frequency, and if so, how this is achieved. In other words, how does the central auditory system go about extracting a single frequency tone from about 28000 (Kim, 1984) nerve fibre spike trains? Of course, the auditory system does not have to extract the tone explicitly, i.e. there need not be an explicit representation of the tone somewhere in the central auditory system. By this it is meant that it is not necessary that the auditory system has a representation where (for

example) only a single neuron fires somewhere in the central auditory system when a 1000 Hz tone at 60 dB SPL is heard. It is however known that pure tones are extracted and represented centrally in some form, as is demonstrated by (for example) the ability of subjects to discriminate between two frequencies and vocalize which was higher in pitch.

This chapter presents a modelling study and proposes a mechanism for the extraction of frequency information from a population of nerve fibres. This model is phenomenological and does not reveal the complexity of the underlying biophysical processes. The intention of the model is to explain how a tone might be extracted from the activities of a population of nerve fibres (i.e. the tone is encoded by a population code), and this is achieved by demonstrating how we can account for psychophysically observed frequency discrimination difference limens or just noticeable frequency differences (jndfs or Δf's).

Mathematical detail of the numerical model is not included in the description in this chapter. The philosophy was to elucidate the principles behind the model rather than to obscure them with mathematical details of the implementation.

2 POPULATION CODING

In neural systems (including the auditory system), sensory information is represented in the activities of a population of nerve fibres. Population coding models have been studied before in a more general context of parameter coding by neurophysiological systems (Baldi and Heiligenberg, 1988; Shadlen and Newsome, 1994). In population coding systems, input sensory information (in this chapter, auditory information in the form of a single tone) is sampled by a limited number of receptors (the inner hair cells in this case). The receptors have rather wide tuning profiles which overlap considerably. For the tuning profiles of the receptors of the auditory system, see, for example, Kiang (1965) or Ruggero (1992). Because of the broadness of tuning, a large number of receptors will be activated by the stimulus. For example, if a listener hears a 1000 Hz pure tone, not only are the fibres that have a characteristic frequency (CF) of 1000 Hz activated, but so are many other fibres with CF in

the vicinity of 1000 Hz. However, the fibres with CF of 1000 Hz are maximally activated and fibres not tuned to 1000 Hz are activated less strongly. Because the tuning profiles are so wide (see, for example, Johnson (1980) or Kiang (1965)), even fibres an octave away from the stimulus may be activated, although weakly.

By comparing the relative activity of all the different receptors, an internal picture may be formed by the central auditory system of the acoustic environment (i.e. the actual physical signal, the pure tone, in this chapter). Activities of fibres in the neural population are in the form of trains of action potentials, or spike trains.

In the model described in this chapter, the viewpoint of an ideal central observer (Bialek, 1991) is adopted. In other words, the central auditory system is imagined as being a central observer with no knowledge of the "outside world", except that which is reflected by the activity of the population of auditory nerve fibres. Each nerve fibre forms an information channel.

The central observer is a conceptual model of all the signal processing that take place in the entire central auditory system. It has available an entire population of nerve fibres responding with different activities to the same stimulus. The reconstruction of the physical signal by the central observer can be much more precise than the spacing between adjoining receptors (Snippe and Koenderink, 1992). Just noticeable differences are typically smaller than the tuning widths of the individual receptors.

In the human auditory system, the central observer receives its only image of the acoustic environment by observing spike trains from about 28000 afferent auditory nerve fibres (Kim, 1984; Spoendlin and Schrott, 1989). From these 28000 spike trains, it has to somehow extract the single tone (or tone in noise) which is presented to the listener in a frequency discrimination experiment.

3 POINT PROCESS DESCRIPTION OF SPIKE TRAINS

The spikes (or action potentials) of a neural spike train are all very similar in shape and size, but the information-bearing aspect of a spike train is the times of occurrence of the spikes. Furthermore, spikes are random in the sense that two identical presentations of the same stimulus do not lead to two identical spike trains. Spike trains may differ in the number of spikes in a given time period and in different times of occurrence of spikes.

Spike trains can be described mathematically as point processes (Johnson, 1996). The theory of point processes describes the occurrence of isolated events (in this case, individual neural spikes) with the mathematical tools of probability theory and statistics. The point process description can provide a basis for mathematical analysis of coding of information in spike trains. The point process having the simplest structure is the Poisson process. Here a Poisson process is a train of spikes, such that the spikes have a Poisson distribution. The Poisson distribution is a mathematical function describing the probability of having exactly k spikes, placed at entirely random moments, in a time interval T. The Poisson process is characterized by the intensity parameter λ. The number of spikes (k) in an interval T is random, so that the exact number of spikes in the interval T is not known when λ is known. However, for larger values of λ, a larger number of spikes are expected in the time interval T. The expected number of spikes in an interval T is λT.

Nerve fibres respond to the stimulus by modulating their firing rates (Shadlen and Newsome, 1994). Every nerve fibre shows an increase in discharge rate over a specific range of pure tone frequency, but the spikes are randomly spaced over the duration of the stimulus and repetitions of the same stimulus do not produce the same number of spikes. This suggests that the Poisson distribution provides an adequate description of the statistics of neural spike trains on the auditory nerve (Keidel et al., 1983; Shadlen and Newsome, 1994). After the occurrence of a spike, there is a short period (the refractory period) during which the nerve fibre is unable to produce another spike. The Poisson process disregards refractory effects (Johnson and Swami, 1983) and is therefore not an entirely accurate model for neural spike trains. However, it was chosen for use in the model described in this chapter, as it is easy to

manipulate mathematically as it has only one parameter (the intensity or rate parameter, λ). The rate parameter λ gives an indication of the instantaneous rate of spikes, and even though spikes occur at random times, λ can under certain conditions be estimated in a straightforward way by counting the number of spikes N in a time period T and by then dividing N by T, i.e. $\lambda^*=N/T$, where λ^* is an estimation or guess of the value of λ. This equation is only a good estimate for the value of λ in the case where it is known that λ does not change during the time period T.

The rate of the spike train, λ, is the only parameter that the central observer needs to extract (Johnson, 1996) in order to have complete knowledge of the stimulus. However, any estimation of λ will never be entirely accurate because of the random nature of the Poisson process. For example, if a time period of T seconds is observed, different numbers of spikes will be observed at each repetition of the same stimulus, although on average the number of spikes will equal λT. Thus, there will be variance in the spike count (as a result of the mathematical definition of a Poisson process, both the average spike rate and the variance equal λ).

The central observer has to expect any normal speech or sound pattern as input and has no way of "knowing" that it is presented with a single pure tone only in a frequency discrimination experiment. So the central observer cannot assume that the λ-parameter remains constant for a specific neural channel; it has to assume that the spike rate is driven by the normal acoustic environment and thus that the λ-parameter is constantly changing. The λ-parameter is a function of the external acoustic signal s(t). This signal is in general random (consider for example a speech signal) and so the random spike train described by a Poisson process has a rate λ driven by a random input signal. Such a process is called a doubly stochastic (i.e. doubly random) Poisson process. The task of the central observer in this general context is to estimate s(t) from the observed set of spike trains. In general, the estimation task of the central observer is extremely difficult. Although the central observer may assume a constant rate λ and use $\lambda^*=N/T$, this will be an extremely poor estimation of the actual λ. If, however, the statistics of the signal s(t) are known (for example, the average and the variance of s(t)), it is possible to obtain better estimators for s(t) and λ. It is assumed

that the auditory system has prior knowledge of the mechanisms that generate signals that it receives. This prior knowledge may be summarized mathematically by a Markov process. A Markov process evolves with known transition probabilities between states. The states in this case are the possible frequencies of the pure tone input. The task of the estimator is to judge which state best represents the frequency of the pure tone input. When a large number of different input frequencies are equally likely, the task of the estimator becomes difficult and errors are made. The point of this discussion is that any estimator of λ or of s(t), even the best estimator, will have variance in the estimation. The mathematical description of the Markov model used here is provided in Appendix 5.A.

This variance leads to limitations in the discrimination performance of the auditory system (Delgutte, 1996), because, as should be obvious, two frequencies may be confused if the estimation variance is large enough that tone *A* might sometimes produce the same estimated spike rate λ^* as tone *B*. The theory of signal detection (Green and Swets, 1966) describes how estimation variance (which may be regarded as an internal noise source) and external environmental noise influence signal detectability and discriminability.

The standard deviation in estimation (the square root of the variance) can be shown to be equivalent to the just noticeable difference (jnd) for the parameter estimated (Siebert, 1970), and this observation is used in the present model. For example, if the stimulus was a 100 Hz pure tone, and the central auditory nervous system estimated this frequency with a standard deviation of 1 Hz, the just noticeable difference in frequency will be 1 Hz. So the task of our model of the central observer will be to observe the spike trains on the population of nerve fibres and (1) to estimate from this the rate λ of each neural channel and then (2) to combine this information to estimate the input frequency *f*. Although the auditory system need not have any explicit representation of the extracted tone, it is fair to design the central observer of the model to indeed extract the tone explicitly. The standard deviation of the estimated value of *f* is then calculated and this is used as the value for the frequency discrimination jnd, Δf.

4 MODELLING AND SIMULATION

A suitable model must be able to predict documented psychophysically measured frequency discrimination jnds (Zwicker and Fastl, 1990; Sek and Moore, 1995; Moore, 1973). With an adequate number of neural channels, it is expected that the variance of the estimation should be reduced relative to an estimation of the tone frequency f made on the basis of a single nerve fibre spike train only. How many nerve fibres should be included in a population coding model and what should their extent be? For the model discussed here, the extent of the neural channels was restricted to one critical band. The critical band is a band of frequencies within which loudnesses of all constituent tones are integrated (Moore, 1997). It is assumed that frequencies within one critical band are processed as a unit by the auditory system. It is known that the critical band is dynamically shifted to be symmetrical around the tone (Zwicker and Fastl, 1990). By restricting the model to one critical band, the following model of the signal processing at the auditory periphery is implicitly assumed: namely that there is a preliminary coarse filtering of the input signal into critical bands, after which the auditory system performs a fine filtering process that extracts the frequency by decoding the population code in this critical band.

There are around 3500 inner hair cells, which provide the primary source of afferent information to the central auditory system. A total of around 28000 afferent nerve fibres innervate these hair cells (Kim, 1984; Spoendlin and Schrott, 1989) so that each inner hair cell is innervated by about eight afferent nerve fibres. About 15% of these, i.e. around one nerve fibre per hair cell, have very low SR (Shofner and Sachs, 1986). As the model is restricted to these low SR nerve fibres, it will be assumed that only one or two nerve fibres per hair cell contribute to the rate-place coding of the input acoustic signal.

By modelling one critical band, a range of around 150 hair cells (Zwicker and Fastl, 1990), spaced 9 μm apart, is considered, i.e. a total range of about 1.3 mm along the basilar membrane. So there are 150 channels of spike trains from which the central observer has to estimate the tone frequency f.

Various techniques exist for combining the channels in a population coded model to decode the input signal (Bialek, 1991; Snippe and Koenderink, 1992; Pouget et al., 1998), the most common of these being to find the centre of gravity of the activities of the population of nerve fibres (Snippe and Koenderink, 1992). This may be explained as follows. Suppose a pure tone stimulus f is presented to the auditory system model. Denote by R_n the "response" of the nth neuron in the auditory nerve population. The response may be the average spike rate of the neuron (for example) in response to the tone. Each neuron has a characteristic frequency f_n, the stimulus frequency to which its response R_n is maximal. In centre-of-gravity estimation, the estimator judges the relative likelihood of all possible stimulus frequencies by using the response R_n of neuron n to weigh the contribution of a frequency at f_n to the received stimulus. For example, if the stimulus frequency was 100 Hz and neuron n had a characteristic frequency of 200 Hz, the response R_n would be lower than for a neuron with characteristic frequency of 100 Hz. Weighing is achieved by multiplying each characteristic frequency f_n by the response R_n of the corresponding neuron. The centre of gravity estimate is then

$$f^* = \frac{\sum_n f_n R_n}{\sum_n R_n}, \tag{5.1}$$

where f^* is the estimate of f and the summations are over all the neurons in the auditory nerve population. It is not known how real neural systems combine information from populations of nerve fibres, but this is not important for the model discussed here. The present model has to combine the 150 channels in an optimal way, as we are interested in determining the optimal frequency discrimination capability of the auditory system, given the available information. Because we are dealing with random variables, information is combined from the population of nerve fibres by calculating the most probable input frequency f using arguments from probability theory. This estimate f^* of the input tone frequency f is not necessarily the same as the estimate calculated by using the more common centre of gravity method.

5 METHODS

The modelling process has two main steps (figure 5.3). In the first step, the activities of the population of 150 nerve fibres which the central observer receives are generated (i.e. the 150 spike trains are generated). In the second step, the best estimate f^* of the input tone frequency f is calculated from the 150 spike trains observed by the central observer. The estimator assumes that the pure tone frequency evolves according to a Markov process. The mathematical details can be found in Appendix 5.A. The discussion below is intended to elucidate the principles.

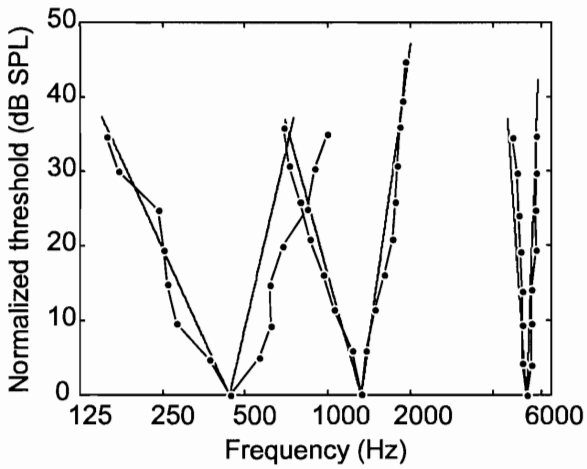

Figure 5.2.

Model tuning curves (solid lines) superimposed on tuning curve data (circles) from Kiang et al. (1965). Tuning curves are for three nerve fibres with characteristic frequencies (439 Hz, 1328 Hz and 5289 Hz) close to the stimulation frequencies used in the model, as tuning curve data were available from Kiang (1965) for nerve fibres with these characteristic frequencies. Tuning curves were normalized by substracting the threshold at the characteristic frequency from all the data points on a curve.

The equations describing the model were coded in Matlab 5.3. Simulations were run on a Pentium II personal computer under the Windows 95 operating system.

5.1 Step 1: Generation of auditory spike trains in response to a pure tone stimulus

1. Low frequency tones (500 Hz, 1500 Hz) and a high frequency tone (5000 Hz) were used. The choice of the 1500 Hz frequency was guided by the availability of data for the population response profiles at this frequency (Shofner and Sachs, 1986). Stimuli were always presented at 60 dB SPL. This intensity was chosen to be well above discrimination threshold of low SR fibres (Shofner and Sachs, 1986) to ensure that the peaks in the population response profiles were well-defined and because of the availability of data measured at intensities in this vicinity (Shofner and Sachs, 1986).

2. The input tones were passed through a set of 150 filters which approximated isorate tuning curves measured for auditory nerve fibres (Kiang et al., 1965). The isorate contours in figure 5.2 are plots of the intensity required at each frequency to achieve a given firing rate 20% above the spontaneous spike rate. The figure shows three representative isorate contours (plotted from data from Kiang et al., 1965) for nerve fibres with characteristic frequencies close to the pure tone input frequencies used in the model. The tuning curves of three model nerve fibres are superimposed on the data. The output of each of the filters was a single value for the average firing rate, λ, of the particular nerve fibre when the input was one of the three tones. The 150 filters were all placed within a range of one critical band, arranged symmetrically around the input tone. Table 5.1 provides information about the range of frequencies for these filters.

Table 5.1. Ranges of critical bands around the input tones used in the model.

Input tone	Range of filters which react to the input tone (1 critical band)	Number of filters in this range	Physical spacing of filters along the cochlea
500 Hz	450 - 570 Hz	150	9 μm
1500 Hz	1350 - 1650 Hz	150	9 μm
5000 Hz	4500 - 5500 Hz	150	9 μm

3. With the activities of each nerve fibre in the neural population known, spike trains were then generated for a 200 ms interval for all the fibres in the population. The spike trains were series of spikes occurring at random times according to a Poisson-distribution with rate parameter λ.

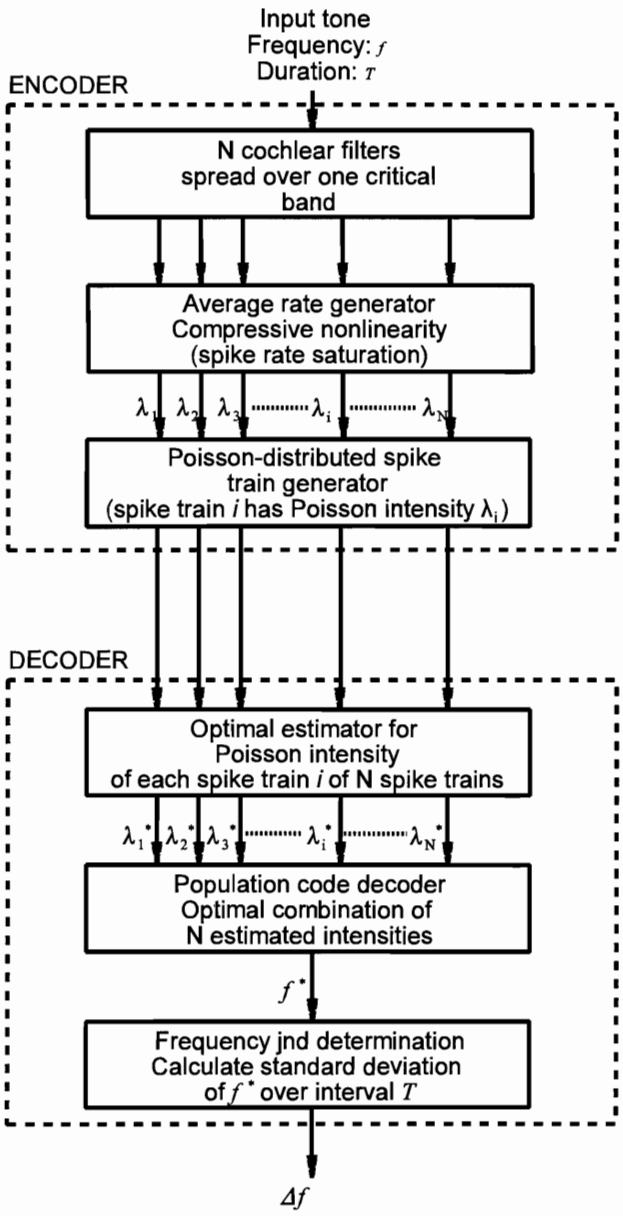

Figure 5.3.

Schematic diagram of the neural encoding and decoding operations implemented in the model. The encoder encodes stimulus parameters (in this case, frequency) in the spike train pattern. The cochlear filters are used to calculate activities of a neural population of N nerve fibres (centred approximately at the input tone frequency) in response to the input tone. The amount of activation of each cochlear filter is then used to calculate the appropriate average spike rate for each fibre, using parameters for low spontaneous rate fibres and taking spike rate saturation into account. The N spike trains now carry the neural code for the input frequency. The spike rate (Poisson process intensity) is estimated separately for each spike train. These estimations are then combined to provide an estimate for the input frequency. The frequency jnd Δf is determined by equating it to the standard deviation in estimated frequency over the duration of the stimulus.

5.2 Step 2: Estimation of the input by a central observer

1. The first calculation that the central observer had to perform, was to find an estimation λ^* of the rate λ on each neural channel. This calculation is in general quite complex as explained previously. A non-linear estimator that assumed a Markov model for the process that generates the pure tone input frequencies was used. The rate estimator in the central observer had 150 inputs (the spike trains from the neural population) and 150 outputs (the estimated value λ^* on each channel).

2. The second calculation was to combine the outputs from the 150 neural channels to form a single estimate f^* of the frequency f of the input tone. This was done in an optimal way as explained before. This estimate varies over the period of 200 ms of signal presentation. The standard deviation of this signal was calculated and this value was used as an approximation to the value of the frequency discrimination jnd Δf at the specific input frequency f.

6 RESULTS

A typical nerve fibre spike train as simulated is shown in figure 5.4(a). The output λ^* of one channel of the estimator is shown in figure 5.4(b). The estimator had to estimate the value of λ, but as can be seen from the figure, this is a quite difficult task and relatively large estimation errors are made.

Figure 5.5 shows a typical estimate of the input frequency as a function of time of a 500 Hz tone of 200 ms duration. This is the frequency estimate after combination of the spike rate estimates of each spike train. Only a small fragment of the frequency axis is shown. The variance in estimation of the frequency of the tone over the 200 ms period can be seen clearly. Fortunately the input tone is coded by a multiplicity of nerve fibres, and the variance of the collective estimation of the tone frequency is far smaller than what would have been expected from figure 5.4 for one nerve fibre alone.

Relative frequency discrimination jnds ($\Delta f/f$) for humans as measured by Sek and Moore (1995) and Moore (1973) (solid lines) are compared to the predictions from the model (markers) in figure 5.6. Clearly, although the task of the central observer seems very difficult (figure 5.4), the population code is quite effective in representing information about the frequency of a single tone. As was observed by Siebert (1970), we see here as well that, given enough channels in the population code, the ideal central observer does better than human

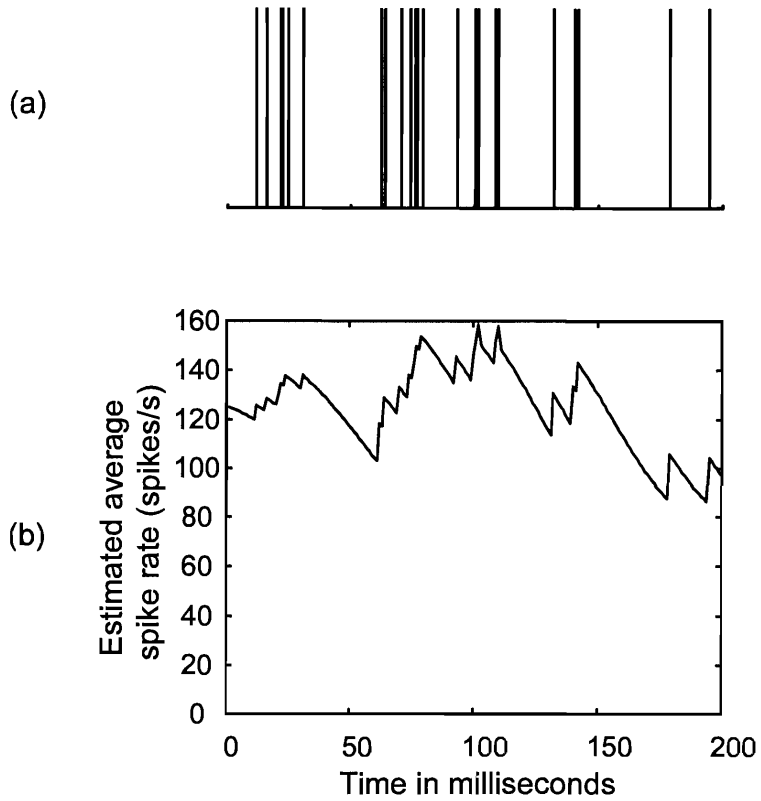

Figure 5.4.

A typical nerve fibre spike train as simulated is shown in (a). The frequency of the tone was 1500 Hz and the characteristic frequency of the nerve was 1503 Hz. The average rate λ was 125 spikes/second. The output λ^* of one channel of the estimator is shown in (b). The tone was presented for 200 ms.

observers in psychoacoustic frequency discrimination experiments. This is true when one or two low SR nerve fibres per hair cell were used and where it was assumed that the other nerve fibres do not contribute to frequency discrimination. When there is less than one low SR nerve fibre per hair cell, the central observer does not have enough information and fares worse than the human observer data of Moore (1973).

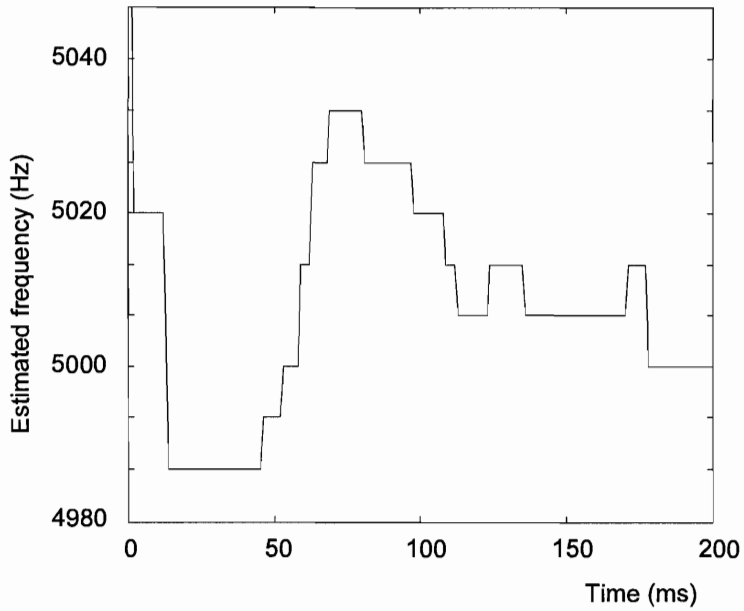

Figure 5.5.

Estimated frequency as a function of time for a 5000 Hz pure tone input of 200 ms duration. Only a small fragment of the frequency axis is shown to show the variance in estimation of the frequency of the tone clearly. The standard deviation is 13.3 Hz.

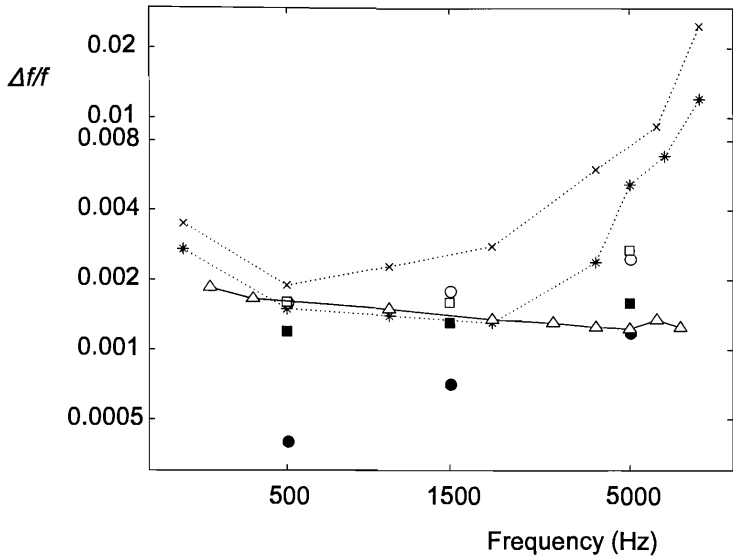

Figure 5.6.

Just noticeable difference in frequency (jndf or Δf) is normalized by frequency ($\Delta f/f$) and plotted as a function of frequency. Human frequency $\Delta f/f$ data (dotted lines) as measured by Sek and Moore (1995) (X) and Moore (1973) (*) are shown. The filled squares are the $\Delta f/f$ values predicted by the model when a population of 150 nerve fibres is used. Filled circles are for 300 fibres, open squares for 75 fibres and open circles for 30 fibres. The datapoints for 75 fibres and 30 fibres coincide at 500 Hz. In all cases the population of nerve fibres was spread over a range of one critical band around the tone. Predictions for $\Delta f/f$ obtained by Heinz et al. are also shown (triangles).

7 DISCUSSION

7.1 Characteristics of the model

The central observer in the model presented here is a single black box that models all the signal processing tasks of the central auditory system from the input of the auditory nerve signals right through to the final decision that the listener has to make in a typical two-alternative forced-choice psychoacoustical experiment. This model has a number of strengths, but also several shortcomings.

The trends in the predicted frequency discrimination just noticeable differences are correct. As frequency rises, the relative Δf's also rise. For a model with 150 nerve fibres representing the tone in a population code, the model predicts values for frequency discrimination jnds close to the data of Moore (1973) at lower frequencies. Furthermore, when human nerve fibre density is taken into account (Spoendlin and Schrott, 1989), around 200 low SR nerve fibres are expected in a critical band centred at 500 Hz, 300 fibres at 1500 Hz and 150 fibres at 500 Hz. A curve plotted through the relevant datapoints in figure 5.6 (the square at 500 Hz, the circle at 1500 Hz and the square at 5000 Hz) has a bowl shape similar to human frequency $\Delta f/f$ data.

It has to be taken into account that the choice of various model parameters influences the ability of the population code to present tonal information accurately. We completely ignored the role of the outer hair cells and also the role that the cochlear nucleus plays in sharpening tuning curves (Kim et al., 1991). Also, we assumed that high SR fibres do not contribute to the rate-place code. If they did, the predicted frequency discrimination jnds would decrease.

Furthermore, the role of phase-locking in coding the identity of a single tone was ignored in the present model. But phase-locking is ubiquitous in the peripheral auditory system and probably plays an important role in the coding of frequency. It was not the purpose of this chapter to prove the opposite, but rather to indicate how the rate-place code may be interpreted. If the central observer uses phase-locking information as well, listeners should

in theory be able to perform far better on frequency discrimination tasks than what has been observed in psychoacoustical experiments. This was also observed by Siebert (1970).

The approach followed by Heinz et al. (2001) is related to the approach described in this chapter in that both have a physiologically based computational model that provides the input to an optimal estimator. The primary difference between the approaches is that Heinz et al. use statistical estimation theory to calculate performance bounds, while the model in this chapter and in chapter 6 generates nonhomogeneous Poissonian spike trains. These spike trains are then used as input to an implementation of an optimal estimation mechanism. This is closer to reality and closer to identifying the underlying signal processing than the approaches followed in many other models.

The Heinz et al. model shows that rate-place information can predict the absolute values of human performance, but not the trend. Their analysis shows that rate-place information fails to predict human deterioration in frequency discrimination thresholds at high frequencies (figure 5.6). However, the frequency discrimination thresholds predicted by the rate-place model described in this chapter does show a trend of deterioration toward higher frequencies. The input to the central estimator in this model always is a set of 150 spike trains with maximum spike rate of 200 spike/s, irrespective of which critical band is the source. If the width of the critical band (which determines the frequency resolution if there are 150 model inputs per critical band) was the only factor that determined frequency discrimination thresholds, relative frequency discrimination thresholds ($\Delta f/f$) would not increase with frequency. Deterioration of frequency discrimination at higher frequencies may be ascribed to differences in the rate response profiles at different frequencies.

7.2 Hair cell loss

Apart from indicating that the assumption of more nerve fibres contributing to the rate-place code leads to better frequency resolution, figure 5.6 can also be interpreted as showing the effect of hair cell loss. With a smaller number of available hair cells, the model leads us to expect that frequency discrimination jnds will rise.

8 CONCLUSIONS

Siebert (1970) predicted the best achievable frequency jnd by other methods, but did not indicate which signal processing the auditory system requires to do to achieve this discrimination threshold. The model presented in this chapter proposes a mechanism for the signal processing required for frequency discrimination.

The model clearly shows (as was also remarked by Siebert, 1970), that the human observer does not make full use of all the information relevant to frequency which is available in the auditory nerve spike trains. The reasons for this are not clear. One possibility is that there are other sources of noise not taken into account in this model. It is concluded from the model that the rate-place code alone is adequate to account for frequency discrimination behaviour in humans.

The model employed an internal model for the process that generates pure tone input frequencies. This internal model takes the form of a Markov model in this chapter. It is shown that the model is able to predict the magnitude of frequency discrimination data for acoustic stimulation, and that some of the trends in the data can also be predicted.

APPENDIX 5.A
OPTIMAL ESTIMATOR FOR THE INPUT FREQUENCY

This appendix briefly describes the optimal nonlinear estimator that was used to estimate the rate parameter λ of a Poissonian spike train that evolves as a Markov process. It is assumed that a specific pure tone that a listener hears is the output of one state of a Markov process. This assumption is made, as it simplifies the design of an optimal estimator.

It is further assumed that a listener can discriminate a limited number of pure tones in a critical band. This number is taken as 150, the number of hair cells in a critical band (Zwicker and Fastl, 1990). The underlying assumption is that a physical spacing of one hair cell represents the auditory frequency resolution limit (see Zwicker and Fastl, 1990). For example, for the critical band ranging from 450 Hz to 570 Hz (table 5.1), the lower bound for frequency discrimination threshold should be 0.8 Hz. For the critical band around 5000 Hz, frequency discrimination thresholds should be larger than about 6.7 Hz, which is true (Moore, 1973; Sek and Moore, 1995).

The task of the optimal estimator is to observe 150 spike trains and then to estimate the state of a Markov process with 150 states. The estimated state then corresponds with the estimate of the pure tone frequency.

A continuous-time Markov process evolves according to the Kolmogorov equation (Ross, 1983),

$$\frac{dp_k(t)}{dt} = -q_k p_k(t) + \sum_{i \neq k} p_i(t) \cdot q_{ik}, \qquad (5.A.1)$$

where p_k is the probability that the Markov process is in state k at time t. The instantaneous rate of jumps from state i to state j is q_{ik}, and the rate of jumps out of state k is q_k.

This equation can be rewritten for a Poisson point process (Rudemo, 1972) as

$$d\hat{p}_k(t) = -q_k\hat{p}_k(t) + \sum_{i\neq k} \hat{p}_i(t)q_{ik} - [\lambda_k - \hat{\lambda}(t)]\,\hat{p}_k(t)\,, \tag{5.A.2}$$

where

$$\hat{\lambda}(t) = \sum_k \lambda_k\hat{p}_k(t)\,. \tag{5.A.3}$$

In these equations, λ_k is the Poisson rate parameter (the spike rate) in state k, $\hat{p}_k$ is the estimate of the state at time t, and $\hat{\lambda}$ is the estimate of the rate at time t. These equations have to be solved simultaneously and are valid between Poisson points (i.e. between spikes). When a spike occurs, the probability of being within state k at time t^+ (directly after the spike) is given by

$$\hat{p}_k(t^+) = \hat{p}_k(t)\,\frac{\lambda_k}{\hat{\lambda}(t)}\,. \tag{5.A.4}$$

The system of equations defined in equations 5.A.2 and 5.A.3 is quadratic. To solve these equations, Rudemo (1972) derives an associated linear system. The forward Kolmogorov equation in this linear system is

$$\frac{dp_k^*(t)}{dt} = -(q_k + \lambda_k)p_k^*(t) + \sum_{i\neq k} p_i^*(t)q_{ik}\,, \tag{5.A.5}$$

which may be written in matrix notation as

$$\frac{dp^*(t)}{dt} = p^*(t)\cdot Q\,, \tag{5.A.6}$$

where Q is a matrix with elements

$$Q_{ii} = -(q_i + \lambda_i)$$

$$Q_{ij} = q_{ij},\ i\neq j.$$

The solution of 5.A.6 is well known as

$$p^*(t) = p^*(t_0)e^{Q(t-t_0)}\,, \tag{5.A.7}$$

with $p_k^*(t_0) = \hat{p}_k(t_0)$.

The probability of being within state k at time t is then

$$\hat{p}_k(t) = \frac{p_k^*(t)}{\sum_i p_i^*(t)}. \qquad (5.A.8)$$

Equations 5.A.5 to 5.A.8 are valid between spikes. Equation 5.A.4 is used at the occurrence of a spike.

Equations 5.A.3 to 5.A.8 can then be used recursively to obtain the time-varying estimated value of spike rate $\lambda(t)$ for each of the 150 spike train inputs. An example is shown in figure 5.4. The estimated rates are then combined to obtain a single estimate of the state at time t. This is achieved through the Viterbi algorithm (Rabiner and Juang, 1993). The Viterbi algorithm uses the $\hat{\lambda}_k(t)$ as noisy observations of the underlying (hidden) Markov state to find the most probable sequence of hidden states. The algorithm is well known and is not repeated here. See Rabiner and Juang (1993) for details on the algorithm.

The time-varying estimate of the state is then used to find the (time-varying) estimate of the input pure tone frequency. This is a simple matter, as each critical band has 150 states, which are for simplicity assumed to be linearly spaced (although cochlear filter arrangement is based on Greenwood's (1990) frequency-position function). Finally, the frequency discrimination threshold is assumed to equate the standard deviation in the calculated frequency estimate as explained in the text.

Chapter 6

A TEMPORAL MODEL OF FREQUENCY DISCRIMINATION IN ACOUSTIC HEARING

The results in this chapter have previously been published: Hanekom, J.J. & Krüger, J.J. 2001, "A model of frequency discrimination with optimal processing of auditory nerve spike intervals", *Hearing Research*, vol .151 no. 1-2, pp. 188-204.

1 INTRODUCTION

Two mechanisms are hypothesized to be involved in the coding of frequency in the auditory system: rate-place coding and phase-lock coding (Dye and Hafter, 1980; Moore and Sek, 1996; Delgutte, 1997; Moller, 1999). Rate-place coding is a spectral analysis mechanism whereby the auditory system may combine firing rate information from nerves originating from spatially restricted sections of the cochlea to determine the stimulus frequency. Phase-lock coding is a temporal mechanism, wherein the auditory system presumably uses the synchronization of neural discharges to individual cycles of periodic stimuli as the primary cue to determine the frequency of a pure tone.

Rate-place coding operates over the entire stimulus frequency range, but is usually presumed to be dominant for the coding of high frequencies (above about 5000 Hz) (Moore, 1973; Kim and Parham, 1991). Phase-lock coding is usually presumed to operate primarily at lower frequencies, since phase-locking is progressively lost as stimulus frequency increases above about 2500 Hz (Delgutte, 1996). No phase-locking is observed above 5000 Hz (Rose et al., 1968; Johnson, 1980). It is possible that both coding mechanisms operate in parallel over a large range of frequencies, but it is not known yet to which extent the central auditory system uses either mechanism alone or both mechanisms simultaneously to determine the frequency

of a pure tone (Dye and Hafter, 1980; Johnson, 1980; Moller, 1999) and there are possibly also inter-species differences in the frequency ranges in which the two mechanisms operate (Hienz et al., 1993).

One motivation for the study of the mechanism used by the central auditory nervous system to code frequency is that understanding the mechanism will influence the stimulation strategies used in cochlear implant speech processors. It is important to know what information transmitted to the electrically stimulated cochlear nerve is perceptually important. Two strategies used in current cochlear implant systems reflect two different approaches. In the Spectral Peak (SPEAK) strategy (Skinner et al., 1994; Loizou, 1999), which is based on the rate-place mechanism, spectral peaks are extracted and presented to electrodes that are arranged tonotopically. In contrast, the Continuous Interleaved Sampling (CIS) strategy (Wilson et al., 1991; Loizou, 1999) uses high pulse-rate stimulation to conserve temporal waveform information.

Several models exist to explain psychoacoustic frequency difference limens (Δf's). These models are based on either the extraction of frequency directly from one or more neural spike trains (i.e. a temporal approach) (Goldstein and Srulovicz, 1977; Javel and Mott, 1988) or the rate-place code (Javel and Mott, 1988), or both mechanisms simultaneously (Siebert, 1970), which includes template matching models (Srulovicz and Goldstein 1983; Erell, 1988). All these models were based on available neurophysiological data (mostly from cat) and were intended to explain psychoacoustic data from neurophysiological data.

Plausible models should account for the absolute values of the Δf's and explain the origin of the bowl shape of the curve of the Weber fraction ($\Delta f/f$) plotted as a function of frequency (e.g. Moore, 1973; Moore, 1993; Sek and Moore, 1995), without the need to manipulate many free parameters to fit the psychoacoustic data. Moreover, Dye and Hafter (1980) have shown that for pure tone frequencies in noise at constant signal to noise ratios, Δf grows larger with increased signal intensity for frequencies below 2000 Hz, while for higher frequencies Δf becomes smaller.

A listener's ability to discriminate between two signals is limited by neural noise, i.e. the Poissonian nature of the neural spike train (Siebert, 1970; Colburn, 1973; Johnson, 1996). Siebert was first to propose the notion that the difference limen in a discrimination task (e.g. frequency or intensity discrimination) is equal to the standard deviation in estimating the magnitude of the stimulus variable (e.g. frequency or intensity). The implication is that estimators may be designed to extract a stimulus variable from its neurally encoded form. The difference limen can then be evaluated by applying known bounds on estimation variance or by calculating estimation variance. The Cramer-Rao Lower Bound (CRLB) (Kay, 1993) provides one such lower bound on the estimation variance of any estimator intended to estimate the magnitude of a stimulus variable, but it does not provide clues to the structure of the optimal estimator. Many authors (e.g. Siebert, 1970, Goldstein and Srulovicz, 1977, Srulovicz and Goldstein, 1983, Wakefield and Nelson, 1985 and Erell, 1988) have used this bound to calculate difference limens for various discrimination tasks. Thus, one shortcoming of many existing models is that they do not provide a neural mechanism by which the central auditory nervous system could implement the psychoacoustic task.

One conclusion from Siebert's work (1970) was that, using all the temporal information available in the set of spike trains of the entire auditory nerve population, the auditory system should be able to perform much better on frequency discrimination tasks than what is actually observed in psychoacoustic experiments. Goldstein and Srulovicz (1977) proposed a temporal model of frequency discrimination wherein frequency is encoded in inter-spike intervals only. They demonstrated that with the combination of a small number of fibers, sufficient information is available to account for perceptually measured frequency discrimination thresholds. Although their model provides acceptable predictions for both absolute magnitude of Δf and the shape of the curve of Weber fraction $(\Delta f/f)$ as a function frequency, Goldstein and Srulovicz did not consider the effect of intensity of stimulation on frequency discrimination.

An extension of their 1977 model (Srulovicz and Goldstein, 1983) accounts for a wider range of psychoacoustic phenomena. The more complex extended model is a template matching model including both temporal and rate-place cues. They concluded that phase-lock coding

is a more likely mechanism than rate-place coding for the frequency discrimination task. Wakefield and Nelson (1985) extended the temporal model of Goldstein and Srulovicz (1977) to include intensity effects. Erell (1988) built on the template matching approach to create a rate-place model that could account for frequency discrimination data in noise.

A recent model by Heinz et al. (2001) provides an important extension to the work of Goldstein and Srulovicz (1977) and Siebert (1970). This model combines computational auditory modelling and theoretical calculation of performance limits predicted by signal detection theory. A physiologically based computational model that can process an arbitrary stimulus is used to produce a time-varying discharge rate. This discharge rate is then used to calculate the CRLB or is used in a likelihood ratio test to calculate performance bounds for two situations. Frequency discrimination performance is predicted when all information in the spike trains are used, and when only rate-place information is used.

Several noteworthy findings emerged from the Heinz et al. (2001) study. First, optimal processing of rate-place information can predict the absolute values of frequency discrimination data, but not the trends. Rate-place predictions are especially poor at high frequencies, where the deterioration in human performance is not predicted. Second, performance predicted by using all available information (in the spike trains of all fibres) shows trends similar to that found in human listeners, although a discrepancy of two orders of magnitude exists. Third, the deterioration in human performance at high frequencies is predicted accurately when using all available information. As phase-locking is lost above 5000 Hz, it is usually assumed that rate-place information is responsible for high frequency behaviour, but Heinz et al. interpreted these results as suggesting that adequate temporal information to account for human frequency discrimination data is available up to at least 10000 Hz.

All of these models use statistical optimal processing arguments via the CRLB to arrive at closed form expressions for frequency discrimination thresholds. The CRLB gives the variance of the minimum variance unbiased estimator and holds for classical estimation problems, i.e. where the parameter to be estimated is unknown, but constant (Kay, 1993).

When the parameters are allowed to vary according to a known probability density function (pdf), Bayesian estimation approaches may provide better estimators as a priori knowledge is built into the estimator. The Bayesian estimation approach is used in this chapter.

Recent neurophysiological data measured by McKinney and Delgutte (1999) provide evidence in favour of an inter-spike interval based extraction of pitch or frequency estimation for pure tones. Their data show that low-order modes of inter-spike interval histograms (ISI histograms) are consistently offset from multiples of the stimulus period. Using this observation, they could predict the octave enlargement effect. The octave enlargement effect is the observation that listeners judge an octave as slightly larger than a 2:1 frequency ratio.

Based on the success of the simple inter-spike interval based model of Goldstein and Srulovicz (1977) in predicting the shape and magnitude of frequency discrimination thresholds, and motivated by the objective to construct a simple, but optimal frequency estimation mechanism that can account for psychoacoustic frequency discrimination thresholds, a new model for frequency discrimination is presented in this chapter. The objectives with this model are:

(1) to extend the well-known model of Goldstein and Srulovicz (1977) to account for intensity dependence and stimulus duration dependence of the frequency discrimination thresholds. The extension is similar to that of Wakefield and Nelson (1985), but we approach the problem from the viewpoint of providing an implementation of the frequency estimation mechanism, whereas Wakefield and Nelson used the statistical approach described earlier;

(2) to provide a simple descriptive model of the statistics of phase-locking;

(3) to construct a central estimation mechanism based on this simple model, by which frequency information can be extracted from one or more neural spike trains;

(4) to demonstrate the role of spatiotemporal integration (Bruce, Irlicht and Clark, 1998) or the volley principle (Wever, 1949) in frequency discrimination;

(5) to demonstrate the role of an internal model in frequency discrimination.

Of course, the auditory system does not have to extract the frequency of a tone explicitly, i.e.

there need not be an explicit representation of the tone somewhere in the central auditory nervous system. This paper does not present any hypotheses about the central representation of pure tones.

Also, the auditory system does not necessarily perform its operations in an optimal way. Even though the objective in the present paper is to describe a possible structure for an optimal frequency estimation mechanism, the emphasis is on the interpretation of the frequency discrimination performance of such an estimator and the factors affecting performance, rather than on suggesting that such a structure exists in the central auditory nervous system.

2 METHODS

2.1 Structure of an optimal processor

Goldstein and Srulovicz (1977) and Wakefield and Nelson (1985) modelled the spike train as a non-homogeneous Poisson process (Johnson, 1996) with the rate parameter being driven by a pure tone. The instantaneous spike rate $r(t)$ is given by

$$r(t) = a e^{kG(f,A)\cos(2\pi ft)}, \qquad (6.1)$$

which is similar to the equations used by Colburn (1973) and Srulovicz and Goldstein (1983). f is the stimulus frequency and t is time in this equation. The product factor $k\ G(f,A)$ is a synchrony parameter that depends on the degree of phase-locking to the stimulus. $G(f,A)$ is the synchronization index that has been defined by Johnson (1980). The synchronization index may take on values between one (all spikes occur on the same phase of the stimulus cycle) and zero (when there is no preferred phase for spikes), although the maximum value of $G(f,A)$ is 0.85 in the current model to fit Johnson's data. Scaling factors a and k are required to fit equation 6.1 to measured values of the instantaneous spike rate (Colburn, 1973). The choice $k = 7.5$ is used in the current model, so that the synchrony parameter $k\ G(f,A)$ has a maximum value of 6.4. This is close to the maximum synchrony value of 6.5 in Srulovicz and Goldstein (1983) and Wakefield and Nelson (1985).

The inter-spike interval distribution of a Poissonian spike train is exponential, and the exact form for this pdf is given in Goldstein and Srulovicz (1977) and in Wakefield and Nelson (1985). With the pdf known, the CRLB can be used to calculate the variance in estimation of the optimal estimator. By modelling the inter-spike intervals differently, constructing an optimal estimator for this problem is possible. Phase-locking is the tendency of the spikes to cluster around multiples of the stimulus period at a preferred phase. It is assumed that these clusters have Gaussian distributions (Javel and Mott, 1988) of which the variance depends on the amount of phase-locking. Perfect phase-locking occurs when spikes always occur at the same phase and when spikes are also entrained to the stimulus (i.e. spikes occur at each stimulus cycle), it is very simple to calculate the stimulus frequency perfectly. Thus the distribution of the spikes around the preferred stimulus phase is a source of noise. Measurements of inter-spike intervals used to estimate frequency are just noisy measurements of the actual period of the stimulus waveform. The problem is similar to the task of estimating the value of a dc signal embedded in noise, except that successive samples (each measured inter-spike interval is one sample) are correlated pairwise, as will be explained. This problem can be solved optimally with a Kalman filter (Kalman, 1960; Kay, 1993; Mendel, 1995). Thus, under these circumstances, the structure of the optimal estimator is known.

2.2 Model of phase-locking

At high stimulation intensities, for fibres with characteristic frequency (CF) at or close to the stimulus frequency, spikes may occur on each stimulus cycle for low frequencies (lower than about 1000 Hz), although this is usually not the case and cycles are often missed (Rose et al., 1968). Spikes can be very scarce at low intensities or when the stimulus frequency is far from the CF of a fibre. Two requirements for creating a realistic model of phase-locking are that (1) spikes should cluster around a specific phase of the stimulus cycle, and (2) the model should allow for cycles in which no spikes occur. Thus, the expected number of spikes in an interval should depend on the stimulus intensity and the closeness of the stimulus frequency to the fibre CF. Many complexities of neuronal responses to sound are not explicitly taken into account in this model. More than one firing per stimulus cycle is often observed at low

frequencies (below about 200 Hz) (Rose et al., 1968). This results in an additional skewed distribution in the ISI histogram, which occurs before the mode at the period of the stimulus. The shapes of the modes of the ISI histogram may be non-Gaussian or skewed, especially at low frequencies (Rose et al., 1968). The influence of these idealizations is discussed below. Accomodation is not taken into account.

In the current model, neural spike trains are produced by a spike generator. The number of spikes in an interval follows a Poisson distribution, with the average spike rate determined by the amount of activation at a specific cochlear place as a result of the stimulus. The average spike rate is determined by a model of peripheral auditory filtering (Colburn, 1973; Goldstein and Srulovicz, 1983). While the actual number of stimulus cycles receiving spikes is calculated according to a Poisson distribution, this does not mean that the spikes are Poisson-distributed. Only the number of spikes in an interval is calculated according to a Poisson distribution and the spike generator then randomly (with a uniform distribution) places spikes on the correct number of stimulus cycles, clustered at a preferred phase. The distribution of spikes is Gaussian with standard deviation

$$\sigma_n = \sqrt{k}\,\frac{1}{2\pi f}\arccos\left(\frac{G(f,A)-1/2}{G(f,A)}\right), \tag{6.2}$$

where $G(f,A)$ is the synchronization index, k is the scaling factor as explained before, and f is the stimulus frequency. Equation 6.2 was derived from equation 6.1 by equating the value of a Gaussian distribution at one standard deviation to $r(t)$, solving for t and equating this t to σ_n (Appendix 6.A). The factor $\sqrt{k}$ is required to rescale σ_n to appropriate values and would not have been required if the scale factor k was not used in equation 6.1. Figure 6.1 shows σ_n as a function of frequency together with data from Javel and Mott (1988). This standard deviation in spike position grows from below 20% of the period of the pure tone stimulus at low frequencies to 35% at 5000 Hz. (Typical spike trains for pure tone stimuli of 1000 Hz and 5000 Hz are shown in figure 6.4).

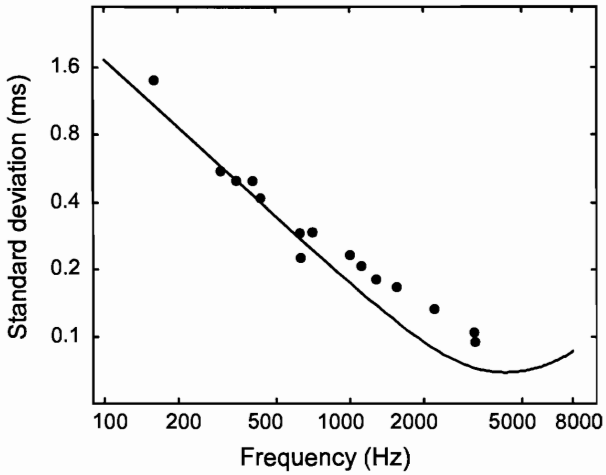

Figure 6.1.

Standard deviation σ_n of spike clusters around the preferred stimulus phase (solid line) as calculated from equation 6.2 is shown together with data on the standard deviations of peaks of inter-spike interval histograms (filled circles) from Javel and Mott (1988).

The synchronization index $G(f,A)$ is a function of both frequency and intensity. $G(f,A)$ may be written as the product of two factors, $G(f,A) = G_1(f)\, G_2(A)$, where A is intensity in dB SL and f is frequency in Hz. $G_1(f)$ is given by

$$G_1(f) = \frac{0.85}{1 + \left(\frac{f}{3500}\right)^3}, \tag{6.3}$$

and $G_2(A)$ is given by

$$G_2(A) = \frac{1.1\, A^{0.3} H}{\sqrt{0.5 (A^{0.3})^2 H^2 + K}} - 0.6\ . \tag{6.4}$$

Equation 6.3 and equation 6.4 are curve fits to typical values of synchronization index as a function of frequency and intensity respectively. In equation 6.4, K is a sensitivity constant which controls the threshold of the model fibre. H is a tuning constant that takes on a

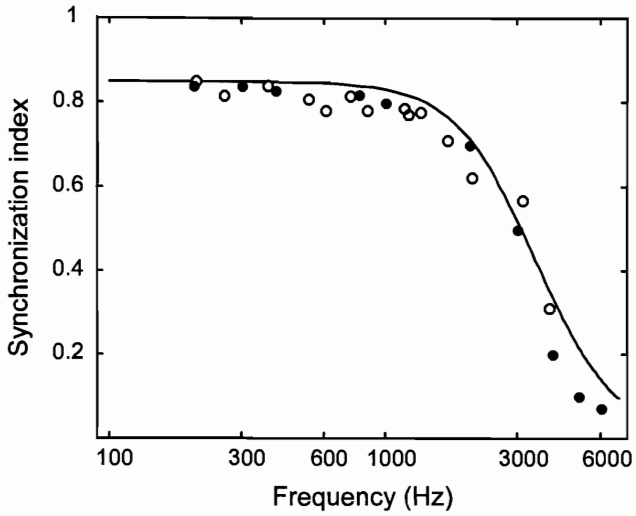

Figure 6.2.

Synchronization index as a function of frequency. The solid curve was calculated from equation 6.3. Filled circles are data from Johnson (1980) and open circles are data from Javel and Mott (1988).

maximum value of 1 when the model fibre has CF at the stimulus frequency. It is assumed that the auditory system uses fibres tuned to the stimulus for temporal estimates of the stimulus frequency, so that H=1 in the current model. $G(f,A)$ is shown in figure 6.2 as a function of frequency at maximum $G_2(A)$, along with measurements of the synchronization index by two authors. $G(f,A)$ is shown in figure 6.3 as a function of intensity at maximum $G_1(f)$, along with neurophysiological data.

2.3 Model of the pooling of spike trains

It is assumed that the auditory system has a way in which to combine spike trains from a number of fibres to obtain a single spike train that has one spike per stimulus cycle. This assumption is an idealization and was made to obtain a simple Kalman filter model, as is explained below. This idea is essentially the same as the volley principle of Wever (1949). Javel (1990) speculated that the great redundancy in auditory nerve fibre innervation of the inner hair cells may exist to ensure that a spike occurs on every stimulus cycle. Superimposing

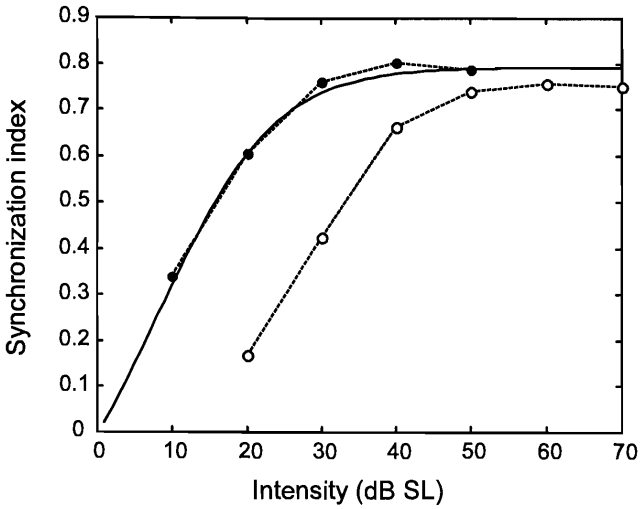

Figure 6.3.

The synchronization index *G(f,A)* is shown as a function of intensity (solid curve) at a fixed frequency of 1000 Hz, using equation 6.4 with *H*=1 and *K*=0.0045. Data from Johnson (1980) is shown for a fibre with CF of 809 Hz. Filled circles are for a pure tone stimulus at CF, while open circles are for a 1162 Hz stimulus. This fibre had a detectable synchronized response at -15 dB SPL, and this was used as the threshold value.

a number of spike trains results in clusters of spikes, with cluster centers spaced approximately 1/*f* apart. If more spike trains are superimposed, estimates of the cluster centers become more accurate, resulting in more accurate estimates of the actual stimulus period.

A very simple model of the combining of spike trains across fibres is proposed. The task of a pooling or integrating neuron is to recognize clusters of spikes and to generate a spike to "mark" each cluster. This may be achieved by an integrating neuron which fires when a number of incoming spikes arrive on its dendrites (as postsynaptic potentials) within a certain time window. This neuron model is part of the family of integrate-and-fire neuron models (e.g. Gabbiani and Koch, 1996). The model incorporates several idealizations as explained below. The model assumes that several auditory nerve fibres synapse with the dendrites of a single integrating neuron located in the cochlear nucleus (CN).

Typical values of postsynaptic potentials, refractory periods and membrane time constants were used in the integrate-and-fire neuron model. The membrane time constant τ can be calculated from the membrane leakage resistance R and the membrane capacitance C as $\tau=RC$. These parameters vary across a wide range and are dependent on the function, location and size of the nerve fibre (Deutsch and Deutsch, 1993). Membrane time constants for onset units in the CN may be very short (Rhode and Greenberg, 1992). Values for R and C from Rattay (1999) reduce to a membrane time constant of 100 μs for myelinated auditory nerve fibres, while the membrane time constant for a motoneuron in cat may be 2 ms (Aidley, 1998). A membrane time constant of 0.5 ms was chosen for the neuron model.

The generation of postsynaptic potentials is a highly non-linear process. Non-linearities include that the post-synaptic potential is a function of the amplitude of the presynaptic action potential (Aidley, 1998) and that the amplitude of the dendritic potential reaching the soma is dependent on the travelling distance from the synapse to the soma and the number of dendrite branchings (Deutsch and Deutsch, 1993). The summation of postsynaptic potentials is also non-linear (Aidley, 1998). These non-linearities were ignored and the model assumed that dendritic potentials reaching the soma have the same amplitude and add linearly. Postsynaptic potentials of CN onset units are small with a maximum amplitude of 4 mV (Rhode and Greenberg, 1992). A value of 1 mV was used for the dendritic potential at the soma as the result of a single spike arriving at a presynaptic terminal.

The absolute refractory period for cat auditory nerve fibres is no shorter than 0.5 ms, and is typically around 0.75 ms (Rose et al., 1968; Gaumond, Molnar and Kim, 1982). A figure of 0.5 ms for the refractory period is also consistent with the responses of fibres in the CN (Rhode and Greenberg, 1992). This refractory period corresponds to a maximum spike rate of 2000 spikes/s, which is higher than the spike rates that auditory nerve fibres are known to be able to sustain. Auditory nerve fibres may attain these high firing rates in the first 10 ms after a stimulus, after which the rate declines (Rattay, 1990). The model assumes an absolute refractory period of 0.5 ms, but does not incorporate a relative refractory period.

The input to an integrating neuron in the CN is a number of postsinaptic potentials arriving

on its dendrites as a result of presynaptic spikes. The response properties of onset units in the CN are thought to be the result of convergence of several auditory nerve fibres (Rhode and Greenberg, 1992). The current model has 40 fibres converging onto the integrating neuron. Although phase-locked auditory nerve fibres do not necessarily fire at the same preferred phase, it is assumed here that the dendritic potentials arriving at the soma have approximately the same preferred phase (or that mean arrival times from different inputs do not differ too much). Synchronization may be achieved by variations in dendritic architecture and properties. Passive properties like dendrite branching patterns, dendrite length, and location of synapses are thought to support information processing operations (Koch, Poggio and Torre, 1982). Voltage-gated channels in dendrites (Cook and Johnston, 1999) may also play a role in supporting or counteracting synapse location-dependent properties of dendritic potential propagation. The fibre model used has less output spike jitter than input spike jitter, even for moderately different mean arrival times. This is consistent with the study of Maršálek et al (1997), who found that output jitter is less than input jitter under a wide range of conditions.

Each input spike is represented by a dendritic potential of 1 mV at the soma that decays exponentially with the membrane time constant of 0.5 ms. A fixed fibre threshold is assumed 10 mV above the resting potential (Johnston and Wu, 1995) and when the threshold is reached, the fibre generates a spike with probability one. During the absolute refractory period of 0.5 ms after the generation of a spike, the membrane potential decays according to its time constant of 0.5 ms and input spikes are ignored.

Simulations with this model at 60 dB SPL showed that the spike train at the output of the integrating neuron has a Gaussian distribution around the preferred phase. The maximum firing frequency is around 2000 Hz for this model. Across the frequency range up to 2000 Hz, the model generates a spike train with exactly one spike per stimulus cycle on most cycles, but in some cycles two spikes occur and in others none. The probability of obtaining two or more consecutive stimulus cycles without spikes is less than 10% and cycles with more than two spikes did not occur in simulations. Simulations show that at 300 Hz, two or more consecutive stimulus cycles without spikes or two spikes per cycle occur for less than 2% of

stimulus cycles. The probability of single cycles with no spikes is 20% at 300 Hz, 2% at 600 Hz and below 1% at 1000 and 2000 Hz, while no dual spikes occur at 600 Hz or above.

These results suggest that it should be possible for the central auditory system to obtain spike trains which fire regularly on each stimulus cycle across a frequency range limited to a maximum of 1000-2000 Hz, with only a small percentage of cycles in which no spikes or dual spikes occur. Candidates for the function of pooling spike trains are the onset locker cells in the CN, which can fire once per stimulus cycle for frequencies up to 1100 Hz (Langner, 1992; Rhode and Greenberg, 1992).

As will be shown, when the proposed model of spike train pooling was used to generate an input spike train for the proposed Kalman filter model, it was found that at low frequencies the standard deviation in estimation is much larger than for the condition of exactly one spike per stimulus cycle. The reason is simply that the state space model as formulated below only allows for the one spike per stimulus cycle condition. However, it is possible to formulate more complex Kalman filter models than the model proposed here to contend with missing spikes or dual spikes. As will be explained in the discussion, a more complex Kalman filter structure with more realistic spike train input will lead to the same conclusions than a simpler Kalman filter that has to contend with the idealized situation of exactly one spike per stimulus cycle. This idealization was used in the current model. As further motivation for using this assumption in subsequent calculations, we remark that a simplifying assumption like this is often made to circumvent extraneous issues that may obscure understanding of the primary signal processing task that the system being modelled has to perform.

An additional motivation for not using the pooling model in subsequent calculations is that the model is constrained to frequencies below 2000 Hz. As phase-locking is known to operate up to 5000 Hz, it was assumed for modelling purposes that it is possible to obtain exactly one spike per stimulus cycle up to and beyond 5000 Hz. It must be emphasized that no known fibres can fire at this rate. Nonetheless, it is interesting to consider the Kalman filter model results at higher frequencies where phase-locking still operates. As elaborated later, the

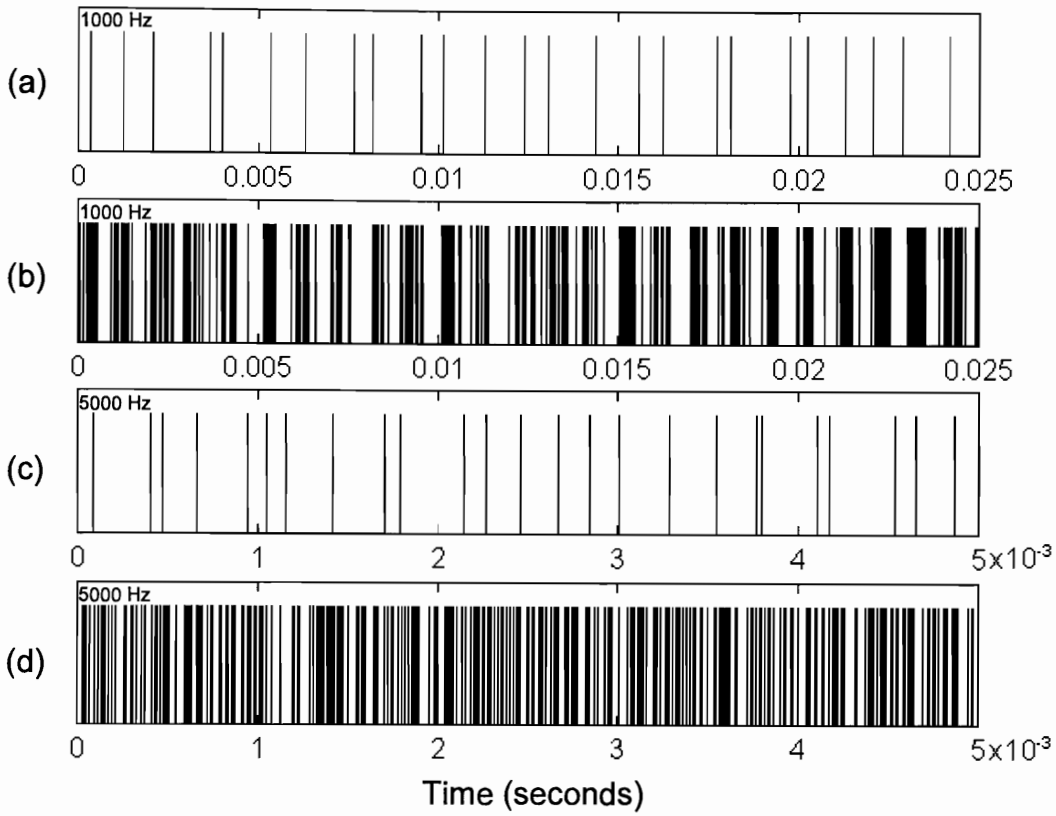

Figure 6.4.

Panels (a) and (c) show spike trains typical of those used as inputs to the Kalman filter for stimuli of 1000 Hz and 5000 Hz respectively. Note that different scales are used on the time axes. The time window is 25 cycles of the pure tone stimulus for both frequencies. One spike was generated per stimulus cycle. Spikes had a Gaussian distribution around a preferred phase of the stimulus. The standard deviation in spike position is 18% of the period of the pure tone stimulus at 1000 Hz and 35% at 5000 Hz. Pooled spike trains from 15 fibres are shown in (b) and (d). Phase locking is evident in the pooled spike train for the 1000 Hz stimulus (b) but is difficult to see for the 5000 Hz stimulus (c) because of the large spike position jitter around the preferred phase.

phase-lock code may be transformed directly into a rate-place code without the need for fibres firing at high rates.

To summarize, spike trains were not combined explicitly for the results presented in subsequent sections. It was assumed that one spike per stimulus cycle was available. Furthermore, improvement of estimates because of superposition was not taken into account, i.e. the spike standard deviation specified in equation 6.2 was used. Typical spike trains are shown in figure 6.4.

2.4 Design of the optimal estimator

When the simplifying assumption of one spike per stimulus cycle is used, the only difficulty in formulating the Kalman filter arises from the fact that the measurement noise is coloured, i.e. there is correlation between samples. This is demonstrated below. The state equations describing the system and measurement are:

$$x(k+1) = ax(k) + bw(k) \tag{6.5}$$

$$z(k+1) = x(k+1) + v(k+1)\ , \tag{6.6}$$

where equation 6.5 is the system equation and equation 6.6 is the measurement equation. Here $x(k)$ is the current inter-spike interval, $x(k+1)$ is the next, and $w(k)$ is the system noise. The system equation models the dynamics of the "signal" $x(k)$. If we expect the inter-spike interval to remain constant, we may assume a=1 and b=0. $z(k)$ is the noisy observation of the period $x(k)$, with $v(k)$ the measurement noise.

The current inter-spike interval clearly depends on both the placement of the current spike and the previous spike:

$$v(k) = n(k) - n(k-1)\ , \tag{6.7}$$

where $n(k)$ is the noise in the placement of the spike around the preferred stimulus phase. This is consistent with the neurophysiological data of McKinney and Delgutte (1999), which show a clear dependence between consecutive inter-spike intervals. The variance of $n(k)$ is σ_n^2. Noise is correlated between consecutive samples, i.e. the value of $v(k)$ depends on the value

of *v(k-1)*. Correlated noise is dealt with by augmenting the system and measurement equations. *v(k)* is eliminated and the system and measurement equations are rewritten in terms of the noise *n(k)* of which the statistics are assumed to be known. If we let

$$x_1(k) = -n(k-1) \ , \tag{6.8}$$

the system equation can be rewritten as a set of two equations:

$$\begin{bmatrix} x(k+1) \\ x_1(k+1) \end{bmatrix} = \begin{bmatrix} a & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x(k) \\ x_1(k) \end{bmatrix} + \begin{bmatrix} bw(k) \\ -n(k) \end{bmatrix} . \tag{6.9}$$

The measurement equation becomes

$$z(k+1) = [1 \quad 1] \begin{bmatrix} x(k+1) \\ x_1(k+1) \end{bmatrix} + n(k+1) \ . \tag{6.10}$$

Clearly, the system and measurement have correlated noise. With the augmented system and measurement equations having been obtained, the Kalman filtering equations are defined in the usual way to obtain recursive estimates for the period *x(k)*. The Kalman filtering equations are well-known (Kay, 1993; Mendel, 1995) and are not repeated here.

The Kalman filter is characterized by two parameters, the system noise σ_w^2 and the measurement noise σ_n^2. The choice of these parameters is based on physiological considerations as described below and the model then predicts frequency discrimination performance very close to perceptual performance. As will be explained later, other choices of these two parameters may lead to frequency discrimination performance far exceeding that observed in humans.

2.5 Choice of Kalman filter parameters

The measurement noise σ_n^2 is simply the variance of the spike distribution around the

preferred phase of a stimulus cycle. The system noise σ_w^2 models the dynamics of the process that generates the stimulus. If the Kalman filter is optimized for a pure tone stimulus (a dc value), σ_w^2 may be set to zero. This, however, makes the Kalman filter slow to react to variations in stimulus frequency as the filter does not "expect" changes. Gap detection thresholds (Zhang and Salvi, 1990; Eddins and Green, 1995) provide a clue of how to choose more realistic values of σ_w^2. The usual explanation of gap thresholds is that the gap is filled, perhaps by the ringing of a cochlear filter (Shailer and Moore, 1987), but gap thresholds are not determined by processing at the auditory periphery alone (Forrest and Formby, 1996). A central observer may not be the primary factor limiting gap detection performance, but at least, the Kalman filter tracking response should be faster than that shown by the neural response as determined by Zhang and Salvi (1990), or otherwise the central observer will introduce even longer gap thresholds.

The variance of the frequency estimate depends on the system noise. A gap can be detected only when the frequency estimate (during the gap) changes by a value greater than the standard deviation σ_w. Variance σ_w^2 may be chosen as zero, but then the response of the filter is too slow and the filter response fills gaps longer than the 2 to 3 ms observed in humans (Eddins and Green, 1995). A tradeoff exists between frequency discrimination thresholds and gap detection thresholds. Larger system noise variance σ_w^2 will allow shorter gaps to be detected, but introduces more estimation variance, which leads to larger estimates for Δf, inconsistent with measurements. Simulations indicate values of σ_w^2 to the order of 10^{-12} to be consistent with both Δf measurements and gap detection thresholds.

2.6 Simulations

In simulations, inter-spike intervals were used as the noisy observations $z(k)$ of the period $x(k)$ of the stimulus. These inter-spike intervals were used as input samples to the Kalman filter. Estimates were obtained for frequency by observing the spike train from a single fibre under the assumption that one spike per stimulus cycle was available. Spikes were placed according to a Gaussian distribution with standard deviation σ_n, the standard deviation of the measurement noise $n(k)$.

Δf was assumed to be equal to the standard deviation in the frequency estimate, following Siebert (1970) and several other authors after him. The standard deviation in the frequency estimate was obtained by repeating the pure tone stimulus of duration T many (typically 200) times and calculating the standard deviation of the frequency estimate at a specific time. This time was either at the end of the interval T or after 50 observations of $z(k)$, as will be explained in the discussion. Values of Δf were obtained as a function of stimulus frequency, intensity and duration.

3 RESULTS

3.1 Δf/f as a function of frequency

Figure 6.5 shows the normalized frequency difference limen ($\Delta f/f$) as a function of frequency as predicted by the model. Parameters are indicated in the caption of the figure. Frequency discrimination data as measured by Sek and Moore (1995) are plotted on the same axes. The shapes of the two curves are very similar, and both reach minima at 500 Hz. The absolute values of $\Delta f/f$ as predicted by the model correspond well to measured values across the entire frequency range, except at 10000 Hz, where the model predicts frequency discrimination that is superior to the psychoacoustic data.

3.2 Δf as a function of intensity

Figure 6.6 shows the model predictions for Δf as a function of intensity A. For intensities below 30 dB SL, Δf decreases monotonically with increasing intensity. As intensity grows above 30 dB SL, the curves level off. For these simulations, it was assumed that the auditory system has access to one spike per stimulus cycle at all intensities down to threshold. Model

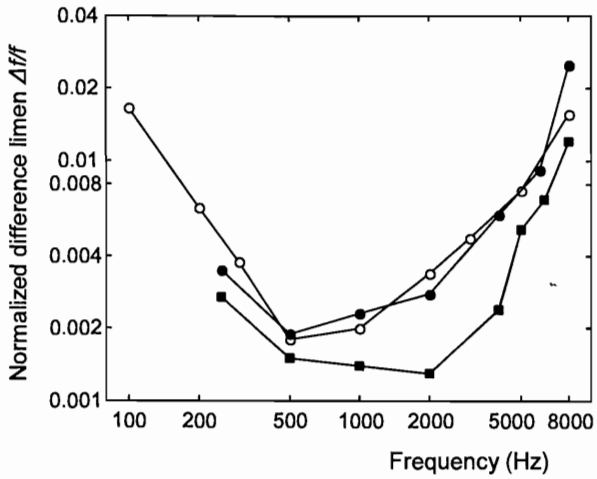

Figure 6.5.

Values of the frequency difference limen Δf expressed as a proportion of frequency ($\Delta f/f$) are plotted as a function of the frequency of a pure tone stimulus on logarithmic axes. Open circles are model predictions, while closed circles are the perceptual frequency discrimination data of Sek and Moore (1995) and closed squares are the data of Moore (1973). Parameters of the Kalman filter were $\sigma_w^2 = 10^{-12}$ and A=60 dB SPL. The measurement noise variance σ_n^2 was a function of frequency as shown in figure 6.1.

predictions are compared with data from Wier et al. (1977) at two frequencies. The model predictions are consistent with the psychoacoustic data in both absolute values and in shapes of the curves. The model prediction at 300 Hz was shifted right by 4 dB to fit the data of Wier et al. at 200 Hz, and the prediction was shifted to the right by 8 dB for the 1000 Hz stimulus, but no other scaling was done on either curve.

The shape of the Δf intensity curve is sensitive for the slope of the synchronization index as a function of intensity (figure 6.3), especially at lower frequencies where fewer observations are available for a given stimulus duration T. This is because σ_n decreases monotonically as synchronization index increases. To account for high Δf 's at low intensities, it is a requirement that the synchronization index approaches zero as intensity approaches threshold. This is consistent with the data in Johnson (1980).

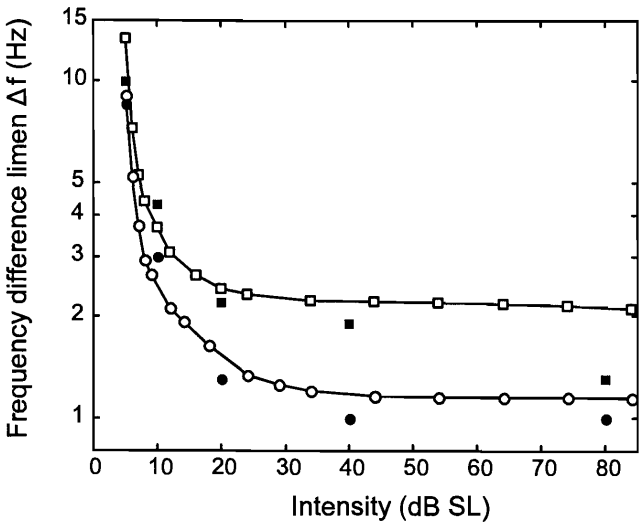

Figure 6.6.

The effect of stimulus intensity (dB SL) on the frequency difference limen Δf is shown for two frequencies. Open circles and open squares are model predictions at 200 Hz and 1000 Hz respectively, while filled circles and filled squares are perceptual frequency discrimination thresholds (Wier et al., 1977). The system noise parameter of the Kalman filter was $\sigma_w^2 = 10^{-12}$, while measurement noise parameter σ_n^2 was a function of frequency as shown figure 6.1.

3.3 Δf/f as a function of duration

The effect of duration on the relative frequency difference limen ($\Delta f/f$) is shown in figure 6.7. This is compared with psychoacoustic data obtained by Moore (1973). The model does not fit the data perfectly, but demonstrates the same trends. At short durations, model predictions for frequency discrimination thresholds are inferior to psychoacoustical performance.

The slopes of the curves are steeper than the psychoacoustic data at short durations, but slope decreases with higher frequencies, which is consistent with the data. Both the model and the data show that an increase in signal duration results in improved performance, until a limit in duration is reached after which performance levels off. The models of Srulovicz and

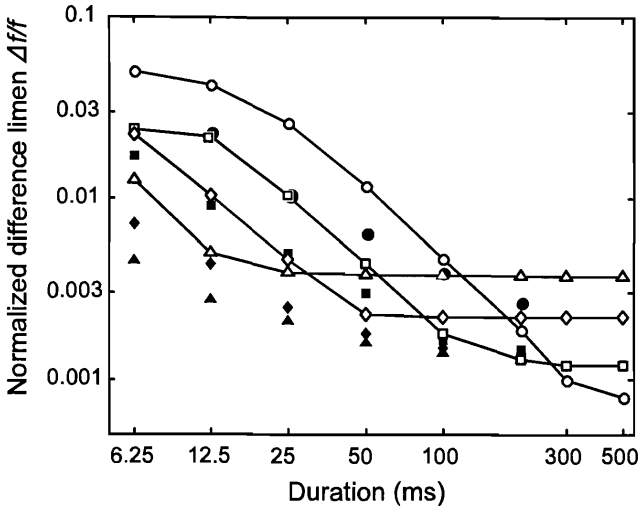

Figure 6.7.

Frequency difference limen (Δf), expressed as a proportion of frequency ($\Delta f/f$), is plotted as a function of duration T of the pure tone stimulus on logarithmic axes. The parameter is frequency. Open symbols are model predictions, while filled symbols are perceptual frequency discrimination data of Moore (1973). The frequencies used are 250 Hz (● and ○), 500 Hz (■ and □), 1000 Hz (♦ and ◊) and 2000 Hz (▲ and △). Kalman filter parameters are the same as in the previous figures.

Goldstein (1983) and Wakefield and Nelson (1985) do not predict this effect, but continue to improve with longer stimulus duration. These authors had to introduce a cutoff time for the maximum duration useful to the central auditory nervous system for estimating the signal frequency, but were not able to assign a single value for cutoff time. The results from the current model suggest that a fixed number of observations (a fixed number of inter-spike intervals, assuming one spike per stimulus cycle), and not a fixed duration, are required to achieve an optimal estimate of each frequency. The required number of observations is estimated at around 50 from the results presented here. This is why longer durations are required to achieve optimal frequency discrimination at lower frequencies. The duration required for optimal frequency discrimination decreases monotonically with an increase in frequency.

3.4 Performance when the one spike per cycle assumption is violated

The Kalman filter model as formulated assumes that one spike occurs per stimulus cycle. This assumption is built into the state space equations. Violations of this assumption, e.g. when cycles are skipped, constitute modelling errors rather than additional noise. If the neural model of spike train pooling as presented above is used, the possibility exists that this may happen. The Kalman filter is very sensitive to modelling errors. When these kinds of errors occur the estimator may lock onto an incorrect frequency and the variance in estimation will grow. Using the neural spike train pooling model (which allows for missed cycles or more than one spike per cycle) over its valid range (up to 2000 Hz), it is found that the shape of the $\Delta f/f$ curve does not change for frequencies below 1000 Hz, but $\Delta f/f$ values are generally larger by an order of magnitude. As the spike train pooling model generates (with high probability) one spike per stimulus interval at frequencies in the range 1000 Hz to 2000 Hz, $\Delta f/f$ values are comparable to the values obtained under the one spike per interval assumption.

4 DISCUSSION

4.1 Nature of the model

An attractive attribute of the model proposed here is the simplicity. Firstly, Goldstein and Srulovicz (1977) used an exponential model (equation 6.1) for the instantaneous spike rate, and from this obtained a pdf for the inter-spike intervals from which the CRLB could be calculated. Instead of this exponential model the current model simply models the distribution of spikes as clusters with Gaussian distributions with standard deviation σ_n around the preferred stimulus phase. Secondly, only first order spike intervals (inter-spike intervals) are required to obtain accurate predictions of psychoacoustic data, as has been shown previously by Goldstein and Srulovicz (1977) and Wakefield and Nelson (1985).

Thirdly, a major difference arises with the implementation of the optimal processor; the current model provides an explicit mathematical implementation of the optimal processor

while previous authors, including Goldstein and Srulovicz (1977 and 1983) and Wakefield and Nelson (1985) used the CRLB to calculate Δf without providing an implementation. It has to be emphasized that we are not referring to a biological implementation here, but rather to the capability of the model to calculate numerical estimates of the stimulus frequency with any given spike train as input. This extends previous models which operate on spike train statistics only, rather than on the spike trains themselves. Comments on biological implementation are given below.

A primary difference between the approaches in the Heinz et al. (2001) model and the model in this chapter is that these authors used statistical estimation theory to calculate performance bounds, whereas the models in this chapter and chapter 5 generate non-homogeneous Poissonian spike trains as input to an implementation of an optimal estimation mechanism. This approach is closer to identifying the actual signal processing that may be performed by the auditory system. Furthermore, spike trains need not be observed for long times to gather statistical information (e.g. to form ISI histograms), but new estimates are formed in real time as new spikes arrive.

4.2 Significance of the Kalman filter model

The Kalman filter model for frequency discrimination is a very simple example of what may be a more general principle of perception, namely an "analysis-by-synthesis" mechanism for perception. The Kalman filter is one of a more general class of estimators which possess an internal model of the system that generates the variable to be estimated (i.e. an internal model of the signal source). In these estimators, input measurement data are compared to predictions generated by the internal model and the prediction error, weighed by a gain, is used to improve estimates of the state of the external system,

$$\hat{x}(k/k) = \hat{x}(k/k-1) + K\left(z(k) - \hat{x}(k/k-1)\right) \tag{6.11}$$

In equation 6.11 $\hat{x}(k/k)$ is the estimated state at the current sample k, given measurements up to this sample. $z(k)$ is the measurement at sample k, while $\hat{x}(k/k\text{-}1)$ is the prediction by the

internal model of the state at sample k, given data up to the previous sample $k-1$. The error in the prediction $z(k)-\hat{x}(k/k-1)$ is weighed by a gain K.

The notions of an internal representation and analysis-by-synthesis have been applied to models of thinking and the brain (Maron, 1965), speech perception (Stevens and Halle, 1967; Lewis, 1996), rhythm and time perception (Todd, O'Boyle and Lee, 1999) and the integration of sensory input and motor control (Wolpert, Ghahramani and Jordan, 1995). As applied to speech perception, the analysis-by-synthesis model contends that when a speech sound is received, the listener attempts to reproduce it by using an internal model of his own production of the sound. This internally generated signal does not activate the musculoskeletal system. If the external and internally generated signals match, the perception is accepted as correct.

The Kalman filter is an explicit implementation of an analysis-by-synthesis mechanism which provides the ability to produce numerical predictions. The class of estimators to which the Kalman filter belongs have a number of characteristics in common. An internal model of an external system is present, often in the form of a state space model (such as Eqs. 6.5 and 6.6). Calculations are recursive in nature, so that there is no need to store previous values of measurements. The gain used to weigh the prediction error follows an exponential-like decay profile until it reaches a steady-state value. For linear state space models, the optimal gain profile as a function of time is given by the Kalman gain, and the optimal estimator is the Kalman filter. Initially, when the first measurement is made, the gain is large, which means that the estimator has little confidence in the internal model's estimates and relies primarily on the measurements to produce estimates of the unknown variable. While the gain is large, the estimator adapts quickly and (if the model is correct) locks onto the variable to be estimated. The internal model is trusted increasingly and the measured data are given less weight as the gain is reduced. Thus, the estimator "grows in confidence". When the gain has reached the steady state value, the noisy measurements still contribute to the estimates, but the estimator primarily trusts its internal estimates. This allows the estimator to suppress noise in the measurements.

The Kalman filter incorporates two sources of noise. System dynamics not explicitly included in the state space model are often represented as a system noise parameter, while measurement noise characterizes imperfections in the measurement process. The steady state value of the Kalman gain is determined by these noise parameters. If the system state to be estimated has large variance or fast dynamics, the steady state gain is larger to allow rapid tracking of changes. However, larger steady state gain results in larger estimation variance.

4.3 Comparison between different classes of models of frequency discrimination

The current model (which provides a numerical implementation) and the statistical models of investigators including Siebert (1970), Goldstein and Srulovicz (1977, 1983) and Wakefield and Nelson (1985) may be regarded as two different classes of models. A third class of models, of which a recent model of McKinney and Delgutte (1999) is an example, operates on ISI histograms. The frequency of a pure tone may be extracted from the interval between the modes of the ISI histogram. Although this class of models is related to the model described in this chapter, they still operate on a statistical representation of spike trains.

To implement a histogram-based frequency estimation mechanism in neural "hardware" would require the central estimator to be able to create and store histograms. Three possibilities exist for the central creation of histograms. Either the central estimator will have to store the values of a large number of inter-spike intervals over a relatively long period, or it will have to pool ISI histograms or spike trains across many fibres. To obtain ISI histograms smooth enough to make reasonably reliable frequency judgements will require long spike records or the pooling of many ISI histograms across fibres. Pooling ISI histograms still requires storage of a large number of inter-spike interval values to form a histogram.

However, as a stable pitch sensation is formed within only around 6 stimulus cycles for low frequencies or 10 ms for high frequencies (Pollack, 1967; McKinney and Delgutte, 1999), it is unlikely that the auditory system creates ISI histograms to estimate frequency. It is more likely that spike trains are combined directly across fibres to form a histogram-like representation. Pooling spike trains across fibres creates a many-cycle period histogram rather

than an ISI histogram. Pooling may occur where many fibres converge onto a single integrating neuron, for example at the onset locker cells in the CN. Note that combining spike trains across fibres in this way does present the additional requirement that phase-locked spikes have the same preferred phase, while the pooling of ISI histograms does not have this requirement.

To summarize, histogram-based frequency estimation mechanisms (Schroeder, 1968; McKinney and Delgutte, 1999) can substantiate that neural spike train statistics are sufficient to enable the auditory system to estimate frequency. However, histogram-based schemes for frequency estimation are only feasible when relatively short periods of spike trains are pooled across many fibres, and not when long spike records are required to construct a histogram.

Histogram-based models are similar to the current model in some respects. Histogram-based models may pool histograms across fibres to form an ISI histogram, while the current model pools spike trains across fibres. Although this was not necessary in the current model, a many-cycle period histogram could then be constructed if an adequate number of spike trains were pooled.

Histogram-based models calculate the stimulus frequency from the distance between mode peaks. Although mode offsets may occur for lower order modes of an ISI histogram, higher order modes may be included in the calculation to estimate the stimulus frequency (McKinney and Delgutte, 1999) more accurately. The current model assumed that the pooling of spike trains resulted in a new spike train with one spike per stimulus cycle, with spikes randomly occurring close to a preferred phase. The Kalman filter model does not estimate the mode peak positions, but uses the spike times directly to estimate the stimulus frequency.

Herein lies an important dissimilarity between the Kalman filter model and the histogram-based models. The standard deviation in spike position results in estimation variance that is used to explain frequency discrimination data in the Kalman filter model. On the other hand, the mode widths (or standard deviations) in an ISI histogram-based model have no influence on the estimated stimulus frequency, so that these models offer no clear-cut explanation for

the variance in estimation of frequency. Mode offsets in ISI histograms vary between different fibres with the same characteristic frequency (McKinney and Delgutte, 1999), which may explain estimation variance if ISI histograms are pooled across fibres.

4.4 The influence of ISI histogram mode offsets

Unlike ISI histograms which may have mode offsets, multi-cycle period histograms formed by combining spike trains across fibres cannot exhibit mode offsets. If one spike occurs a little before the preferred phase, the next spike interval must be a little longer if phase-locking is maintained. Successive inter-spike intervals are correlated (equation 6.7), as also shown by joint first order histograms (McKinney and Delgutte, 1999). Mode offsets in ISI histogram-based models may bias the frequency estimates of these models, but do not play a role in the Kalman filter model. At any rate, as frequency discrimination is assumed to be related to the variance in estimation only, biases in frequency estimates will not influence frequency discrimination in the current model.

4.5 The influence of peak splitting

Peak splitting occurs in ISI histograms when two or more spikes occur per stimulus cycle (Ruggero and Rich, 1989; McKinney and Delgutte, 1999). Less than 10% of the fibres demonstrated peak splitting in the extensive data of McKinney and Delgutte (1999). The authors argue that only a small fraction of fibres will exhibit peak splitting at a specific stimulus intensity, as the intensity at which peak splitting occurs is a function of stimulus frequency as well as fibre CF. Peak splitting does not influence the current model if the proposed model of spike train pooling is used. Peak splitting will result in dual spikes in a small number of stimulus cycles at the input to an integrating neuron. The temporal and spatial integration of spikes that occur at the integrating neuron ensures that dual spikes on the input usually do not result in dual spikes on the neuron output. Peak splitting may still occur at the output of an integrating neuron, as explained in section 2.3.

4.6 Robustness with respect to the number of spikes per stimulus cycle

The Kalman filter model as formulated is very sensitive to modelling errors. A tenfold increase in $\Delta f/f$ occurs at frequencies below 1000 Hz because of the way the state space model was formulated, i.e. the model does not permit the possibility of either more or less than one spike per stimulus cycle. This problem may be overcome by creating more elaborate Kalman filter models. One example of a slightly more complex Kalman filter model would be a model that assumes that either one spike occurs in every stimulus cycle, or not more than one cycle is skipped. Even more elaborate Kalman filters may allow more realistic spike train patterns.

These more complex estimators will have a relationship between the variance in spike position (around the preferred phase) and the estimation variance similar to the original Kalman filter. Development of such models is beyond the scope of this chapter, but they will lead to similar conclusions as have been reached from the results with the simple Kalman filter model presented here.

4.7 Robustness with respect to spike distribution

One of the assumptions of the model is that spike clusters around the preferred stimulus phase have Gaussian distributions (Javel and Mott, 1988). Although it was a natural idealization that simplified the equations, this assumption was not necessary. The Kalman filter is based on second order statistics and any distribution with the correct mean and variance will give the same results. Thus, the model is not sensitive to non-Gaussian or skewed ISI histogram modes.

4.8 Parameter sensitivity and the origin of the shape of the Δf/f frequency curve

The Δf obtained is a tradeoff between three parameters of the model: the number of observations, the system noise and the measurement noise. The choice of the system noise parameter σ_w is least obvious. When the system noise variance $\sigma_w^2 = 0$ and the appropriate variation is used for the measurement noise σ_n (equation 6.2), the shape of the $\Delta f/f$ frequency

curve obtained is similar to the psychoacoustic curve, but flatter at the high and low frequency ends. The $\Delta f/f$ curve shifts downwards with increasing number N of observations. The same absolute values of $\Delta f/f$ as found in the psychoacoustic data are achieved with 40 to 50 observations. It is possible for the model to significantly outperform human observers with correct parameter choice. One possibility is to use zero system noise and to increase the number of observations N.

If σ_w^2 is chosen as a constant but non-zero value, the expected standard deviation of stimulus period from observation to observation is a constant over frequency, which implies that the ratio of spike standard deviation to stimulus signal period grows, or in other words the signal (stimulus period) to noise (standard deviation) ratio decreases with increasing frequency. This results in growth in $\Delta f/f$ towards higher frequencies. As shown before, to be consistent with gap detection data, $\sigma_w^2=10^{-12}$ is a good choice. This results in a high frequency $\Delta f/f$ slope consistent with psychoacoustic data. For this choice of σ_w^2 it is also found that the number of observations needs to be close to N=50 to achieve the same $\Delta f/f$ values as the psychoacoustic data. Larger N results in little further decrease in $\Delta f/f$.

For these parameter choices, i.e. N=50 and $\sigma_w^2=10^{-12}$, the model predictions are consistent with psychoacoustic data at frequencies above 500 Hz. To account for psychoacoustic data below 500 Hz, stimulus duration T is limited to 100 ms so that the number of observations decreases with lower frequencies, which results in a growth in $\Delta f/f$ at lower frequencies consistent with psychoacoustic data. This choice for T is consistent with known auditory integration times. Longest integration time for pure tones has been estimated to be in the 100 ms to 300 ms range (Green, 1973; Eddins and Green, 1995). Although perceptions of loudness or pitch emerge well before 200 ms, computations of loudness and pitch, as required in discrimination experiments, continue to improve up to approximately 200 ms (Lewis, 1996).

The assumption could also have been made that the auditory system uses a constant frequency deviation criterion, i.e. the system noise σ_w^2 should be chosen such that a constant frequency deviation rather than a constant period deviation is expected across frequency. This results in a growth in σ_w^2 towards higher frequencies, resulting in even smaller signal to noise ratios

at high frequencies, in turn resulting in larger values of $\Delta f/f$ and a steep slope towards high frequencies, which is inconsistent with psychoacoustic data. This suggests that the auditory system uses estimates of stimulus period rather than frequency to obtain frequency estimates.

The significance of the choice of a 1 μs standard deviation ($\sigma_w=10^{-6}$) in period is not clear. Coincidence detectors in the CN (Delgutte, 1997) can respond to spike timing differences which is an order larger than σ_w. CN cells are probably not able to react to differences in period as small as 1 μs, which means that this is in the neural noise bed. As it is possible that the auditory system may be able to regulate the internal noise (Tomlinson and Langner, 1998), the auditory system may have chosen to work with a non-zero system noise that is in the noise bed, as opposed to choosing $\sigma_w^2=0$, to avoid divergence in the estimate (Mendel, 1995).

4.9 Frequency range

The current model provides accurate predictions of frequency discrimination over the entire frequency range up to at least 6000 Hz. However, several studies have shown that no observable phase-locking is present above about 5000 Hz (Rose et al., 1968; Johnson, 1980; Palmer and Russell, 1986) and model predictions should be disregarded at frequencies where no phase-locking exists. Within the frequency range of 100 Hz to 5000 Hz, the model predicts psychoacoustic frequency discrimination thresholds accurately, suggesting the possibility that the phase-lock code operates across this entire frequency range. Other investigators (Dye and Hafter, 1980; Javel and Mott, 1988; Javel, 1990) have suggested that phase-locking is used at frequencies below 1000 Hz while rate-place coding is used for higher frequencies.

Model predictions for high frequencies can be explained by the standard deviation in spike distribution around the preferred phase (the spike jitter). It is interesting that model predictions are accurate up to 5000 Hz, even though cells that can sustain entrained firing at rates higher than around 1000 Hz have not been found. Many auditory afferent fibres converge on onset cells in the CN. Onset locker cells fire once per stimulus cycle for frequencies up to 1100 Hz, on a very precise phase of every stimulus cycle (Langner, 1992; Rhode and Greenberg, 1992), with better precision than found in the auditory nerve. This is

consistent with the notion that a volley principle may be operational in the CN, at least for low frequencies. It may be reasonable to assume that a model based on phase-locking and the volley principle can hold up to 1000-1500 Hz (Rhode and Greenberg, 1992) only.

However, why are model predictions still accurate at higher stimulation frequencies? One explanation may be that the auditory system may have a mechanism to estimate frequency from fibres that fire at integer multiples of the stimulus period rather than on each stimulus cycle. Chopper units in the CN can lock onto integer multiples of the stimulus period very precisely (Wiegrede and Winter, 2000). The constraints under which the central estimator has to perform are still the same, i.e. estimation variance is limited by neural noise and the number of available observations.

A second explanation may be the gradual transformation of temporal information on the auditory nerve into a rate-place code at higher levels of the central auditory system. It is possible that this transformation takes place at the level of the CN (Rhode and Greenberg, 1992), although it is not known how such a transformation takes place (Brugge, 1992). A large number of auditory afferents carrying a phase-lock code converge on CN cells. These fibres should, as a population, provide at least one spike per stimulus cycle on the input to a CN neuronal assembly. The possibility exists that the phase-lock code may be transformed directly into a rate-place code without the need for fibres firing at rates up to 5000 Hz, but the accuracy of such a transformation would still be dependent on auditory afferent spike jitter. Such a mechanism could operate over the entire frequency range of phase-locked activity.

However, evidence is available that suggests that the upper limit for the encoding of frequency by phase-locking is below 1000 Hz. Cochlear implant users cannot discriminate changes in sinusoidal electrical stimulation frequency above about 300-500 Hz (Shannon, 1983a), while modulation detection performance decreases monotonically above 100 Hz for normal-hearing listeners (Bacon and Viemeister, 1985) and for cochlear implant users (Shannon and Otto, 1990).

To summarize, the Kalman filter model results are consistent with an optimal central estimator that is constrained by limitations in the number of observations at low frequencies (below 500 Hz) and by spike position jitter at higher frequencies (above 500 Hz). Although the model can predict frequency discrimination data over the entire frequency range in which phase-locking is observed (up to 5000 Hz), not enough neurophysiological evidence is available to support a claim that phase-locking is used for the encoding of frequency across this entire range, and evidence exists which suggests that phase-lock coding is used only at low frequencies.

4.10 Number of fibres required

Previous models predict considerably better human frequency discrimination performance than measured perceptually. Siebert (1970) used the entire array of nerve fibres and the occurrence times of all spikes in an optimal processing model to obtain predictions for Δf far surpassing human frequency discrimination ability. Goldstein and Srulovicz (1977) and Wakefield and Nelson (1985) used inter-spike intervals only and required only nine nerve fibres to account for human frequency discrimination data. If one spike were available for each stimulus cycle, the current model would require only a single nerve fibre to account for human frequency discrimination data. However, even at high intensities, firings do not occur at every stimulus cycle (Rose et al., 1968). The current model is for the availability of spikes on every stimulus cycle and estimation errors grow rapidly when spikes are missed as shown previously.

At high intensities, the combination of spike information from just a few nerve fibres will ensure the availability of at least one spike per stimulus cycle. At lower intensities, the combination of more nerve fibres is required to account for human frequency discrimination data. If too few fibres are pooled, it cannot be ensured that at least one spike is available per stimulus cycle. As mentioned previously, more complex Kalman filter models can contend with the condition where spikes do not occur on every stimulus cycle. Calculating the least number of fibres to be combined to have a combined *average* of at least one spike per stimulus cycle is simple. There is no guarantee that spikes will be available on each stimulus cycle when fibres are combined, but the probability of missing cycles decreases as the number

of fibres to be combined increases. It is estimated from simulations that the current model requires the combination of not more than 100 fibres to ensure at least one spike per stimulus cycle at all frequencies and supra-threshold intensities.

4.11 Behavior of the model in noise

The Kalman filter model incorporates noise in the system noise and measurement noise parameters. The system noise characterizes the variability of the stimulus frequency because of signal dynamics. The measurement noise is a result of imprecise measurement of the stimulus period because of the Gaussian distribution of spikes around the preferred stimulus phase. These noise sources exist even when the stimulus is a pure tone in quiet. Additive external noise may be incorporated into the measurement noise parameter, as addition of noise tends to desynchronize the neural spikes so that phase-locking becomes less precise (Dye and Hafter, 1980). The effect is that for all frequencies, Δf grows and the Δf intensity curves flatten.

This is contrary to the observation by Dye and Hafter (1980) that Δf in humans is frequency dependent at constant signal to noise ratios. For pure tone frequencies at 3000 Hz or above, Δf grows larger with increased signal intensity, while at 1000 Hz or lower frequencies Δf becomes smaller. The crossover point is around 2000 Hz. The current model cannot predict these effects. The current model, or any model based on phase-locking, predicts decreasing Δf with increasing intensity, as the temporal dispersion of spikes around the preferred stimulus phase will become smaller with increased intensity. To predict increasing Δf with intensity, the synchronization index should become smaller with increasing intensity. This possibility exists. The cat data presented in Johnson (1980) show examples of a fibre tuned to intensity, i.e. for which the synchronization index grows with intensity at lower intensities and declines again at higher intensities.

4.12 Comments on the use of cat neurophysiological data to predict human performance

Neurophysiological data used in the current model to predict human psychoacoustic data were obtained in cat. Several authors have cautioned that cross-species comparisons are subject to interpretational difficulties, especially because of a lack of neurophysiological stimulus encoding data from humans (Hienz et al., 1993). Also, humans discriminate smaller frequency increments than monkeys (Prosen et al., 1990) or cats (Javel and Mott, 1988; Hienz et al., 1993) and frequency discrimination dependence on intensity differs in these species (Hienz et al., 1993). Previous studies suggested that differences between frequency discrimination thresholds in humans and other animals may be a result of different frequency encoding mechanisms in different species (Prosen et al., 1990; Hienz et al., 1993). From several studies it is clear that frequency information is available in both rate-place codes and phase-lock codes. The accurate prediction of human frequency discrimination thresholds with the current model suggests that encoding of stimuli in neural spike trains may be very similar in cat and human, although different animals may still use different decoding strategies to extract frequency information at different frequencies, resulting in differences in discrimination thresholds.

4.13 Comments on neural implementation

The Kalman filter model proposed here is a purely mathematical operation, but is believed to have biological significance. Although a possible biological implementation has been proposed for a volley principle, no attempt has been made to suggest a potential biological implementation for the Kalman filter or to speculate where in the auditory system such a mechanism may exist. This is outside the scope of this chapter. It is not the intention with the current model to claim that the central auditory nervous system implements a Kalman filtering mechanism to extract frequency information, but rather to demonstrate by example how an analysis-by-synthesis principle may be used for the estimation of a biological parameter. Two primary intentions were (1) to demonstrate that most of the psychoacoustic frequency discrimination data may be explained by the statistics of a simple spike generation model, (2)

using a recursive optimal processor that operates on spike trains with these statistics.

A recursive mechanism for noise suppression and parameter estimation may be attractive from a biological implementation viewpoint. The neural implementation of a Kalman estimator requires that the central auditory system has an internal model of the signal source (as reflected in the state space model equations and the Kalman filter parameters) and the ability to perform recursive calculations. Instead of explicitly storing information from the entire duration of a stimulus, the history of a sequence can be stored in the internal states of fibres (Lewis, 1996). Computations can then be carried out so that the output of a neural calculation is a function of past inputs and the present input. This will limit the amount of data that have to be stored in short term memory, which is attractive from a biological implementation viewpoint. Although these kinds of computations are of theoretical use, more support from neurophysiological and psychophysical work is needed to establish the biological relevance. Plausible biological implementations for recursive calculations have been discussed in literature, e.g. McLaren (1989), but there is little biological evidence for the existence of these mechanisms. Wolpert, Ghahramani, and Jordan (1995) present experimental data that support the notion of the existence of an internal model and recursive calculations in sensorimotor integration.

5 CONCLUSIONS

(1) It was shown that an analysis-by-synthesis type of mechanism may be used to explain frequency discrimination data. The Kalman filter model described here is an explicit implementation of an analysis-by-synthesis mechanism which provides the ability to produce numerical predictions.

(2) This recursive implementation of a frequency estimation mechanism can account for most of the psychoacoustic data for frequency discrimination in quiet.

(3) The particular Kalman filter model constructed in this chapter depends on the availability of one spike per stimulus cycle, which may be provided by the operation of a volley principle. More complex recursive estimators will free the model from the one spike per stimulus cycle constraint.

(4) The temporal information in inter-spike intervals is sufficient to account for human frequency discrimination performance up to 5000 Hz.

(5) The number of observations of spike intervals, rather than the integration time used for estimates, is probably fixed.

(6) Under the assumption of constant system noise across frequency, the number of observations accounts for the low frequency part of the $\Delta f/f$ frequency curve, while the measurement noise accounts for the high frequency part of the curve. Thus model results are consistent with an ideal observer that is limited by the number of available observations at low frequencies (below 500 Hz) and by spike position jitter at higher frequencies (above 500 Hz).

APPENDIX 6.A
DERIVATION OF EQUATION 6.2

The instantaneous spike rate $r(t)$ for a fiber that is phase-locked to a pure tone stimulus is given by equation 6.1, which is also interpreted as the envelope of a multi-cycled period histogram. This envelope is bell-shaped and close to being Gaussian for arguments of the cosine function in the range $(-\pi, \pi)$. To approximate $r(t)$ by a Gaussian, it is necessary to calculate the standard deviation of a Gaussian with the same width than this period histogram envelope. Ignoring temporarily the scale factor k in equation 6.1 gives

$$r_1(t) = a e^{G(f,A)\cos(2\pi f t)} . \tag{6.A.1}$$

Equation 6.A.1 is first normalized to have a maximum value of 1 at t=0,

$$a e^{G(f,A)\cos(2\pi f t)} = 1 . \tag{6.A.2}$$

Solving for a,

$$a = e^{-G(f,A)} . \tag{6.A.3}$$

The normalized period histogram envelope is now equated to a Gaussian which has also been normalized to height 1 at the origin. To calculate the width of the Gaussian that fits $r_1(t)$, the heights of $r_1(t)$ and the Gaussian are equated at $t=\sigma$,

$$e^{-G(f,A)} e^{G(f,A)\cos(2\pi f \sigma_n)} = e^{-\sigma_n^2/2\sigma_n^2} \tag{6.A.4}$$

Solving for σ in terms of the synchronization index $G(f,A)$,

$$\sigma_n = \frac{1}{2\pi f} \arccos\left(\frac{G(f,A) - 1/2}{G(f,A)}\right) \tag{6.A.5}$$

Equation 6.A.5 is a good fit to the standard deviations of peaks of inter-spike interval histograms from Javel and Mott (1988). Reintroducing the scale factor k into equation 6.A.1 to obtain equation 6.1, it is necessary to scale equation 6.A.5 as well to fit the data. Scaling equation 6.A.5 by $\sqrt{k}$, equation 6.2 is obtained, which provides a good fit to the data from Javel and Mott (1988).

Chapter 7
A TEMPORAL MODEL OF FREQUENCY DISCRIMINATION IN ELECTRIC HEARING

The results in this chapter have previously been published: Hanekom, J.J. 2000, "What do cochlear implants teach us about the encoding of frequency in the auditory system?", *South African Journal of Communication Disorders*, vol. 47, pp. 49-56.

1 INTRODUCTION

A long-standing question about frequency analysis in the auditory system is how frequency information is represented: is frequency coded as a temporal code or as a place code (Moller, 1999) or as both? Pure tones are represented as both rate-place information (rate-place coding) and temporal information (phase-lock coding) in the discharge patterns of auditory nerve fibres and the central auditory nervous system, but, as discussed in the previous chapter, the extent to which the auditory system uses either representation is unknown.

It is possible that both coding mechanisms operate in parallel over a large range of frequencies. So far, neither neurophysiological studies in animals, nor psychoacoustic experiments in humans have been able to determine to which extent the central auditory system uses either mechanism alone or both mechanisms simultaneously to determine the frequency of a pure tone (Moller, 1999; Johnson, 1980). Whilst previously this may have been regarded as a purely academic question, the development of cochlear implants has made it important to understand how information is coded in the auditory nervous system. This knowledge will influence the stimulation strategies used in cochlear implant speech processors.

Specifically, we need to know what information transmitted to the electrically stimulated cochlear nerve is perceptually significant. Two strategies used in current cochlear implant systems reflect two different approaches. In the Spectral Peak (SPEAK) strategy (Skinner et al., 1994; Loizou, 1999), which is based on the rate-place mechanism, spectral peaks are extracted and presented to electrodes that are arranged tonotopically. In contrast, the Continuous Interleaved Sampling (CIS) strategy (Wilson et al., 1991; Loizou, 1999) uses high pulse-rate stimulation to conserve temporal waveform information.

Moller (1999) reviews the roles of temporal and rate-place coding of frequency. He presents convincing arguments in favour of the phase-lock code. His arguments are based on the robustness of the code and on the effects of various kinds of pathology on the impairment of frequency discrimination and pitch perception. It is well established that frequency tuning in the auditory system is a function of sound intensity (Moller, 1999; Johnstone et al., 1986). The location of maximal vibration of the basilar membrane shifts at higher intensities. However, the perception of pitch of pure tones is relatively insensitive to changes in intensity over large intensity ranges. Some models conjecture that it is not the spectral peaks, but rather the complete spectral profile, or the edges of the spectral profile that are used in frequency discrimination (Moore and Glasberg, 1986). Moller (1999) argues that these are just as sensitive to intensity. Moller also reviews data that suggests that impairment of spectral analysis in the cochlea does not affect speech discrimination noticeably, which suggests that spectral analysis might not be important for speech perception.

Cochlear implants provide the opportunity to study the coding of frequency by the rate-place code and by the phase-lock code separately. The perception of frequency as encoded in the rate-place code alone may be studied by using a fixed stimulation frequency and varying the site of stimulation in the cochlea. To explore the role of phase-lock coding alone, stimulation at a fixed position in the cochlea may be used while varying the frequency of stimulation (Townshend et al., 1987; Blamey et al., 1996; Dorman et al., 1994).

This chapter further explores the coding of frequency in the phase-lock code, using

neurophysiological and psychoacoustic data from auditory electrical stimulation as instrument. One motivation for this study stems from the strong arguments by Moller (1999) in favour of phase-lock coding, but the way in which frequency information should be encoded in cochlear implants is also of interest.

1.1 Approach

The model of frequency discrimination in the acoustically stimulated auditory system (chapter 6; Hanekom and Krüger, 2001) is extended to include the electrically stimulated auditory system. This model provides satisfactory predictions of frequency discrimination in the normal (acoustically stimulated) auditory system. If, in addition, it is found that the extended model can predict psychoacoustic data for the electrically stimulated auditory system, strengthening the arguments in favour of a phase-lock code for frequency may be possible. The applicability to cochlear implant stimulation strategies will be explained in the discussion.

The objective of the extension to this model is to include the statistics of spike trains evoked by electrical stimulation in order to make predictions about frequency discrimination thresholds in cochlear implants. The statistics of spike trains that result from electrical stimulation are quite different from acoustically evoked spike trains, as has been discussed in chapter 4. It should be noted that this chapter does not present any hypotheses about the central representation of pure tones. The emphasis is on the interpretation of the frequency discrimination performance of an optimal estimator, presumably located somewhere in the central auditory nervous system, given the statistics of acoustically and electrically evoked spike trains.

2 METHOD

The phase-lock model of chapter 6 is extended to incorporate electrical stimulation. The proposed extension is quite simple and detail is provided below.

2.1 Estimator structure

Phase-locking is the tendency of nerve spikes (action potentials) to cluster around multiples of the stimulus cycle at a preferred phase (a specific time relative to the onset of the stimulus cycle). It is assumed that these clusters have Gaussian distributions (Javel and Mott, 1988) of which the variance depends on the amount of phase-locking. Perfect phase-locking occurs when spikes always occur at the same phase. When spikes are also entrained to the stimulus (i.e. spikes occur at each stimulus cycle), calculating the stimulus frequency perfectly is very simple. Thus, the distribution of the spikes around the preferred stimulus phase is a source of noise.

It is assumed that the auditory system can combine spike trains from a number of fibres to obtain a single spike train that has one spike per stimulus cycle. This idea is essentially the same as the volley principle of Wever (1949). Javel (1990) speculated that the great redundancy in auditory nerve fibre innervation of the inner hair cells may exist to ensure that a spike occurs on every stimulus cycle. Superimposing a number of spike trains results in clusters of spikes, with cluster centres spaced approximately $1/f$ apart. If spike trains from more fibres are superimposed, estimates of the cluster centres become more accurate, resulting in more accurate estimates of the actual stimulus period. This is on the condition that fibres fire on exactly the same preferred phase, which is not necessarily true. Different fibres, tuned to slightly different frequencies, will all phase-lock to the stimulus, but each at its own preferred phase (Javel, 1990). Neurons have been found in the cochlear nucleus that may be able to implement a volley principle by combination of several auditory nerve inputs (Moller, 1999). To achieve this, the integration centre has to compensate for differences in the preferred firing phase. The auditory system may achieve this by variation in fibre length and variation in the strength of synapses (Cook and Johnston, 1999).

In the implementation of the model described in this chapter, spike trains were not combined explicitly. It was assumed that one spike per stimulus cycle was available. Under this condition, measurements of inter-spike intervals used to estimate frequency are just noisy measurements of the actual period of the stimulus waveform. The problem of obtaining a good estimate of

the stimulation frequency from these noisy measurements can be solved with an estimator that is often used in engineering applications, the Kalman filter (Kalman, 1960).

2.2 Number of fibres combined

At high intensities, the combination of spike information from just a few nerve fibres will ensure the availability of one spike per stimulus cycle. At lower intensities, the combination of more nerve fibres is required to account for human frequency discrimination data. The probability of missing cycles decreases as the number of fibres to be combined increases. It is estimated from simulations that the current model requires the combination of around 100-200 fibres to ensure one spike per stimulus cycle at all frequencies for acoustical stimulation.

2.3 Model of phase-locking for acoustic stimulation

At high stimulation intensities, for fibres with characteristic frequency (CF) at or close to the stimulus frequency, spikes may occur on each stimulus cycle for low frequencies (lower than about 1000 Hz), although this is usually not so and cycles are often missed (Rose et al., 1968). Spikes can be very scarce at low intensities or when the stimulus frequency is far from the CF of a fibre.

It is assumed in the current model that the central processor integrates spike train information from a restricted area in the cochlea where the strongest activity is found. This corresponds to the average localized synchronized rate (ALSR) model of Sachs and Miller (1985). Thus, for a stimulus well above threshold as used in the current model, it is assumed that the most strongly stimulated fibres fire at their maximum rates.

A further assumption is that spikes cluster around a specific phase of the stimulus cycle according to a Gaussian distribution. The distribution of spikes is Gaussian with standard deviation given by equation 6.2 (chapter 6). The synchronization index was given by equations 6.3 and 6.4.

2.4 Model of phase-locking for electrical stimulation

Some important differences exist between neural synchronization to acoustical and electrical stimulation. First, electrically stimulated fibres exhibit phase-locked responses with a much higher degree of synchronization (Shepherd and Javel, 1999; Javel, 1990; Van den Honert and Stypulkowski, 1987a; Hartmann et al., 1984). This is demonstrated in period histograms where the peak is much narrower for electrical stimulation than for acoustic stimulation (e.g. Javel, 1990). Thus phase locking occurs on a very precise phase of the stimulus signal (Hartmann et al., 1984). For high frequencies (4-8 kHz), phase-locking is weaker (Dynes and Delgutte, 1992). The fibre still discharges regularly, but many stimulus cycles may be skipped, similar to the acoustic case. Phase locking is maintained at high frequencies (10 kHz) for electrical stimulation, unlike acoustical stimulation which demonstrates no phase-locking above 5 kHz. The synchronization index for electrical stimulation is described by

$$G_1(f) = \frac{0.92}{1+\left(\frac{f}{3500}\right)}, \tag{7.1}$$

which is a curve fit to the synchronization index data in Dynes and Delgutte (1992) as shown in figure 7.1.

Second, the degree of entrainment is much higher for electrical stimulation than for acoustic stimulation. Spikes may occur on each stimulus cycle for low frequencies (below 1 kHz) (Javel and Shepherd, 2000; Javel, 1990).

Third, there is no frequency selectivity for electrical stimulation (there is no basilar membrane filtering), so that all areas activated by the electrical stimulus generate action potentials, regardless of the stimulus frequency or the CF of the stimulated site. In addition, the statistical

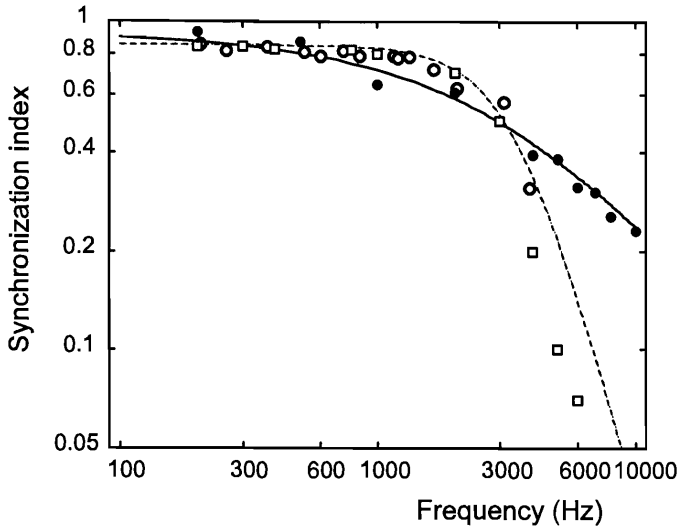

Figure 7.1.

Synchronization index as a function of frequency for electrical and acoustic stimulation. The solid curve (electrical stimulation) was calculated from equation 7.1. The dotted curve (acoustic stimulation) was calculated from equation 6.3. Open squares are data from Johnson (1980) and open circles are data from Javel and Mott (1988) for acoustic stimulation. Filled circles are data for sinusoidal electrical stimulation from Dynes and Delgutte (1992).

independence among spike trains in different fibres is lost, so that many fibres fire exactly in phase (Javel, 1990; O'Leary et al., 1995).

Finally, synchronization increases with stimulus intensity (Shepherd and Javel, 1999). Spike latencies become smaller for more intense stimuli, and spike latencies are shorter for pulsatile stimuli than for sinusoidal stimuli. The probability of firing is a function of the stimulus strength, but the slope of this function is very steep (Van den Honert and Stypulkowski, 1987a) so that for pulsatile stimulation at just around 6 dB above threshold, neurons fire at their maximum (entrained) rate. It is assumed that fibres are stimulated well above threshold, so that a value of $G_2(A)=1$ is used for electrical stimulation in the current model.

2.5 Combination of fibres for electrical stimulation

If spike trains are combined according to a volley principle, then presumably the auditory system must compensate for differences in the preferred firing phase of different fibres as explained before. The assumption was made in the previous chapter that phase-locked spike trains arriving at the central integration centre have the same preferred phase. This may be achieved by a mechanism that compensates for differences between fibres, as explained in chapter 6. The mean arrival times of spikes on different inputs were the same, and the time of spike occurrence was determined by a Gaussian distribution with standard deviation σ_n around the preferred phase. If linear summation at the central integration centre is assumed and the pdfs of the different input spike occurrence times are the same, the pdf of the output spike times is the same as the input pdfs (as was assumed in chapter 6). Nonlinear processing of input spike trains by the neural model in chapter 6 results in smaller output standard deviation. This is consistent with other neural models (Burkitt and Clark, 1996) and physiological data (Rhode and Greenberg, 1992).

Fibres fire exactly in phase in electrical stimulation, and that should lead to improved frequency discrimination ability. However, frequency discrimination is poorer in cochlear implantees than in normal-hearing listeners (e.g. Townshend et al., 1987) and mechanisms that may bring this about are modelled in this chapter.

So for electrical stimulation, it is assumed that phase-locked spike trains arriving at the central integration centre are desynchronized relative to each other. It is assumed that the central integration centre still generates one spike per stimulus cycle, but with larger variance in spike position around the preferred phase than for the acoustic case. Thus

$$\sigma_n = s\sqrt{k}\,\frac{1}{2\pi f}\arccos\left(\frac{G_1(f)-0.5}{G_1(f)}\right) \tag{7.2}$$

is used for electrical stimulation, with $G_1(f)$ as given in equation 7.1. Increased standard deviation is reflected by the factor s in this equation.

Two cases are considered. First, any mechanism that compensates for differences between fibres in acoustic stimulation may in fact disperse spikes around the preferred phase in electrical stimulation as measured at a central integration centre. Assuming linear summation, the pdfs of the spikes on the different inputs will add. If spikes on different inputs have different mean arrival times, the output variance will increase. As spike jitter is small for pulsatile electrical stimulation, output variance will be determined primarily by the degree of dispersion in mean arrival times.

Cochlear nerve fibres are myelinated and have diameters of typically 1 μm (Rattay, 1990) to 3 μm (Frijns et al., 1995), so that action potential conduction velocity is estimated to be between 6 m/s and 18 m/s (Deutsch and Deutsch, 1993). The length of the cochlear nerve is approximately 5 mm, so that maximum conduction time from the cochlea to the CN should be around 0.8 ms for 1 μm fibres. If this is assumed to set an upper limit to delays that the central integration centre has to compensate for, output spike standard deviation may increase by 4 times (s=4 in equation 7.2) relative to the standard deviation of 0.18 ms at 1000 Hz predicted by equation 6.2.

Second, neurogenic disease or injury to nerve fibre causes decreased conduction velocity (Kandel, Schwartz and Jessell, 1991). A nonregenerative or weakly regenerative section of fibre or demyelination can result in a tenfold increase in neural conduction time (Deutsch and Deutsch, 1993). Demyelination occurs in long-term deaf fibres, resulting in prolonged refractory periods (Shepherd and Javel, 1997). If some auditory nerve fibres are injured, temporal dispersion of neural activity will result, so that spike trains arriving on different inputs to the central integration centre will be desynchronized. Ignoring that decreased output standard deviation may be achieved by nonlinear processing at the central integration centre, output standard deviation may increase tenfold (s=10 in equation 7.2). Interestingly, it is known that injuries to the auditory nerve affect speech discrimination ability more than cochlear injuries (Moller, 1999).

2.6 Implementation of the estimator and simulations

Details about the derivation of the Kalman filter equations may be found in Hanekom and Krüger (2001) or chapter 6. Spike trains were computer-generated using the model. Two different approaches were used to obtain the output spike train of a central integration centre. This output was the input to the Kalman filter for both approaches.

In the first approach, as in chapter 6, estimates were obtained for frequency by observing the spike train from a single modelled fibre under the assumption that one spike per stimulus cycle was available. Spikes were placed according to a Gaussian distribution with standard deviation σ_n. For acoustic stimulation, $G_1(f)$ of equation 6.3 was used in equation 6.2 to calculate σ_n, while $G_1(f)$ as in equation 7.1 was used in equation 7.2 for electrical stimulation.

The frequency difference limen Δf was then obtained by assuming it to be equal to the standard deviation in the frequency estimate, following Siebert (1970) and several other authors after him. The standard deviation in the frequency estimate was obtained by repeating the pure tone stimulus of duration T many (typically 200) times and calculating the standard deviation of the frequency estimate at a specific time. This time was either at the end of the interval T or after 50 observations of inter-spike intervals, as will be explained in the discussion. Values of Δf were obtained as a function of stimulus frequency for both acoustic and electrical stimulation.

The second approach was to use the nerve fibre model of chapter 6 as a model for the central integration centre. The integration centre was assumed to have 40 inputs, and the output spike train was then used directly as input to the Kalman filter. Mean arrival time dispersion on the inputs was assumed to be 0.8 ms, which is taken as the maximum desynchronization between spikes travelling on different fibres from the cochlea to the CN as explained above. The frequency difference limen Δf was then obtained as explained above.

A number of differences exist between these two approaches. The explicitly modelled central integration centre has less temporal jitter on the output spike train than on the input spike

trains, especially at higher frequencies where the input spike rates are higher. However, dual spikes or no spikes may occur in response to some cycles of the stimulation waveform.

Furthermore, model nerve fibre cannot generate output spike trains at a rate higher than 2000 spikes/s due to the refractory period.

The equations describing the model were coded in Matlab 5.3. Simulations were run on a Pentium II personal computer under the Windows 95 operating system.

3 RESULTS

3.1 Δf/f as a function of frequency for acoustic stimulation

Figure 7.2 shows the normalized frequency difference limen ($\Delta f/f$) as a function of frequency for electrical stimulation as predicted by the model when equation 7.2 is used to predict the temporal jitter at the output of a central integration centre. For comparison, $\Delta f/f$ for acoustic stimulation is also shown. Frequency discrimination data for acoustic stimulation as measured by Sek and Moore (1995) are plotted on the same axis. The shapes of the two curves for acoustic stimulation are very similar, as discussed in chapter 6, and both reach minima at 500 Hz.

The electrical stimulation curves show trends similar to the acoustic stimulation curves at higher frequencies, but $\Delta f/f$ is generally larger. However, in contrast to acoustic stimulation, the $\Delta f/f$ curve for electrical stimulation decreases monotonically toward lower frequencies. This trend is explained in the discussion. Two situations of desynchronization are investigated as explained above. In the first (filled triangles) the assumption was that compensation for differences between fibres in acoustic stimulation is responsible for dispersion of mean arrival times in electrical stimulation. A smaller increase in $\Delta f/f$ than for cochlear nerve injury (the second case; filled circles) is observed.

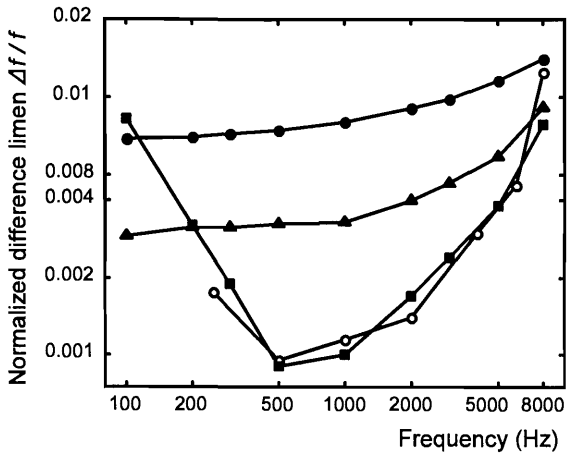

Figure 7.2.

Values of the frequency difference limen Δf expressed as a proportion of frequency ($\Delta f/f$) are plotted as a function of the frequency of a pure tone stimulus (acoustic stimulation) or frequency of electrical stimulation on logarithmic axes. Filled squares are model predictions (from figure 6.5) for acoustic stimulation, while open circles are the perceptual frequency discrimination data of Sek and Moore (1995). Filled circles (s=10) and filled triangles (s=4) are model predictions for electrical stimulation when equation 7.2 is used to calculate the jitter at the output of a central integration centre.

3.2 Δf as a function of frequency for electrical stimulation

Figure 7.3 shows the frequency difference limen (Δf) as a function of frequency for electrical stimulation as predicted by the model, using both approaches explained previously. Simulation predictions are not shown as the Weber fraction $\Delta f/f$ in this figure, as the measurement data were available as Δf. Frequency discrimination data as documented in Pfingst (1988) and Townshend et al. (1987) are plotted on the same axes. The trends of the model prediction and the psychoacoustic data are similar when equation 7.2 is used to obtain the output temporal jitter. Slopes are similar at higher frequencies. The absolute values of Δf as predicted by the model are smaller than measured values across most of the frequency range. At high frequencies, model predictions may be an order smaller than the psychoacoustic data.

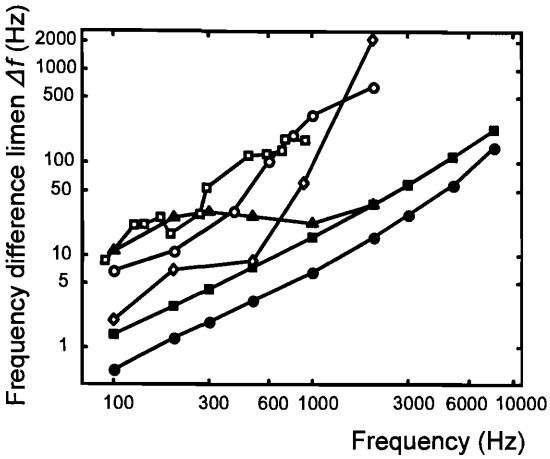

Figure 7.3.

Values of the frequency difference limen Δf are plotted as a function of the frequency of electrical stimulation on logarithmic axes.

Open circles and open diamonds are perceptual frequency discrimination data from two studies (for sinusoidal electrical stimulation) as reported in Pfingst (1988). Open squares are data for pulsatile electrical stimulation (Townshend et al., 1987).

Filled symbols are model predictions for electrical stimulation. Two of the model-predicted curves were obtained using equation 7.2 to estimate the output spike jitter, with s=4 (filled circles) and s=10 (filled squares) respectively. The third modelled curve (filled triangles) was obtained by using the fibre model of chapter 6 as a model for a central integration centre, assuming mean arrival time dispersion of 0.8 ms among the inputs.

To bring the model and data into closer agreement, larger standard deviations may be used for spike position jitter to shift the curves upward. Alternatively, higher stimulus intensities shift the measured curves downwards (Pfingst, 1988). Unfortunately, it is unknown at which stimulus intensities the data were measured.

In contrast to the predictions obtained with equation 7.2, Δf saturates at higher frequencies with the explicitly modelled fibre output used as input to the Kalman filter. The trend and

magnitude of Δf at lower frequencies are similar to those of the data. Larger Δf is obtained at low frequencies than in the curves predicted with equation 7.2. This is due to increased Kalman filter estimation variance, because dual or no spikes occur on some cycles. At higher frequencies, temporal jitter at the central integration centre output is smaller than predicted by equation 7.2 so that the Kalman filter estimates improve. When the mean arrival times on different input fibres to the central integration station differ by more than about 0.8 ms, phase-locking is increasingly lost. Many cycles are then skipped and many dual spikes occur, resulting in rapidly increasing Kalman filter estimation variance.

4 DISCUSSION

4.1 Justification of assumptions

The current model rests on two important assumptions. First, it was assumed that the auditory system has some way to ensure one spike per stimulus interval across the entire frequency range. This assumption idealizes known neurophysiological data. More than one spike may occur per stimulus cycle in acoustic stimulation (Rose et al., 1968) and multiple spikes may occur in electrical stimulation (Javel and Shepherd, 2000). Both phases of the electrical stimulation waveform may evoke spikes (Van den Honert and Stypulkowski, 1987b) and multiple spikes per phase may occur at higher frequencies of pulsatile stimulation (Javel and Shepherd, 2000).

Because of the way that the current model was formulated, the Kalman filter expects exactly one spike per stimulus cycle. More than one spike per stimulus interval will be regarded as a source of noise. If a small percentage of cycles have either more than one spikes per cycle, or some cycles are skipped, the dominant inter-spike interval is still the period of the stimulus waveform and the central estimator will make the correct estimate (although with larger standard deviation in the estimate). With many cycles not conforming to the one spike per cycle assumption, the central estimator may make an incorrect estimate of the input frequency. A higher likelihood exists that this will happen for electrical stimulation, as spikes may occur on both phases of the stimulus waveform.

Nonetheless, the model may still explain the observed frequency difference limens, because frequency discrimination measurements are differential and do not measure the absolute frequency perceived. The close correlation between the predicted and measured acoustic frequency discrimination thresholds suggests the possibility that a central representation of the pure tone exists that is equivalent to the one spike per stimulus interval assumption. This, however, is not what the model intended to prove. Rather, the intention was to show that frequency discrimination thresholds for both acoustic and electrical stimulation could be explained by spike position jitter in a phase-locked response.

The second assumption was that, because many fibres fire in phase as a result of electrical stimulation, the net result at the central auditory estimator would be a desynchronization of spike trains, rather than improved synchronization. It is unknown whether data exist which support this hypothesis. Available data seem to refute this notion. The cochlear nucleus (CN) exhibits greater response diversity than the auditory nerve (O'Leary et al., 1995). Some fibres display phase-locking to the stimulus, while the responses of other fibres are more complex. CN fibres that do phase-lock exhibit very little temporal dispersion of spikes for electrical stimulation (Javel and Shepherd, 2000).

However, as it is not known what the central representation of frequency is, to search for spike trains at the CN output that exhibits larger spike position jitter for electrical stimulation than for acoustic stimulation may be fallacious. It is known that temporal information on the auditory nerve is gradually transformed into a rate-place code at higher levels of the central auditory system, possibly at the level of the CN (Rhode and Greenberg, 1992). Many auditory afferents carrying a phase-lock code converge on CN cells. These fibres should provide at least one spike per stimulus cycle on the input to a CN neuronal assembly. The possibility exists that the phase-lock code may then be transformed directly into a rate-place code without the need for fibres firing at rates up to 5000 Hz. So, not enough is known to be able to prove or disprove the second assumption.

Neither assumption is unrealistic in terms of biological implementation and the results justify the two assumptions to some extent.

As a final comment, the possibility that the model predictions only hold for frequencies below 5000 Hz needs to be pointed out, as no phase-locking is observed at higher (acoustic) stimulation frequencies.

4.2 The origin of the shape of the Δf/f frequency curve

The Δf obtained is primarily a tradeoff between two parameters of the model: the number of observations and the spike jitter around the preferred phase of the stimulus cycle.
As explained in chapter 6, to account for psychoacoustic data below 500 Hz, stimulus duration T is limited to 100 ms so that the number of observations decreases with lower frequencies, which results in a growth in $\Delta f/f$ at lower frequencies consistent with psychoacoustic data. This choice for T is consistent with known auditory integration times (Eddins and Green, 1995). At these frequencies, $\Delta f/f$ is determined primarily by the number of observations available. At higher frequencies the number of observations in the 100 ms time interval grows. For acoustic stimulation, it was found that the number of observations needs to be close to N=50 to achieve the same $\Delta f/f$ values as the psychoacoustic data. Larger N results in little further decrease in $\Delta f/f$. To account for the data in figure 7.3, the restriction of a 100 ms time window was lifted and N=50 observations were allowed throughout the frequency range. This required a stimulus duration of 500 ms at 100 Hz.

At higher frequencies (above 500 Hz), the spike jitter becomes a systematically growing percentage of the stimulus period. This plays the primary role in the growth of $\Delta f/f$ at these frequencies. It seems that, as long as at least 50 observations are available, the tradeoff between the number of observations and spike jitter is not important in electric hearing. Spike jitter (because of desynchronization) may be the primary factor that controls frequency discrimination. This suggests that auditory integration times may be longer (at low frequencies) in electric hearing than in acoustic hearing, but this notion will require further investigation.

Estimates of Δf using the explicitly modelled central integration centre are realistic at low frequencies, suggesting that desynchronization may play a role in determining frequency

discrimination thresholds as hypothesized, and that this desynchronization may be in the 0.8 ms region as modelled. The saturation of the predicted Δf curve is because of reduced jitter at the output of the central integration centre and may indicate shortcomings in the fibre model. The integration centre receives more spikes on its inputs at higher frequencies, but it is possible that not all of these are useful in the generation of the output spike train, and that the model does not allow for this adequately. It is also known that spike position jitter increases with higher frequencies in electrical stimulation (Javel and Shepherd, 2000), which would result in increased Δf at high frequencies, but the model did not take this into account.

4.3 What do cochlear implants teach us about the coding of frequency in the auditory system?

Psychoacoustic data from cochlear implants seem to refute the idea that temporal coding mechanisms are utilized by the central auditory system to extract frequency information from the neural spike train, as the frequency difference limens are much poorer than for normal-hearing listeners, even though there is much more synchronization to the stimulus waveform in electrical stimulation. The current model demonstrates (with reasonable assumptions) that a central auditory estimator that uses inter-spike intervals to calculate frequency may fare worse with electrical stimulation than with acoustic stimulation. This is consistent with psychoacoustic data. So at least the current model indicates that we cannot rule out temporal mechanisms as a mechanism for frequency coding.

It is known that cochlear implant signal processing strategies based on preserving the temporal pattern (e.g. CIS) are generally more successful than strategies based on vocoders (e.g. SPEAK) (Loizou, 1999), which supports the argument in favour of phase-lock coding.

Also, recent studies have shown that fewer channels in a speech processor can lead to equally good or better speech discrimination (Fishman et al., 1997), but if fewer than 4 to 6 channels are used, performance drops. The interpretation is that the actual number of independent information channels in an implant is probably not more than 4 to 6. Also, because higher stimulation rates can be achieved with fewer activated electrodes (Shannon et al., 1990), the

temporal characteristics of the signal are preserved better. Thus, evidence suggests that good spatial resolution is not achieved in cochlear implants, but also that preservation of the temporal waveform is important in cochlear implants.

Conversely, it has been shown in many pitch discrimination or electrode discrimination experiments (Nelson et al., 1995; Pfingst et al., 1999), where a fixed stimulation frequency was used on various electrodes, that cochlear implant users can discriminate between electrodes. Furthermore, pitch estimation experiments show that implant users can assign pitch to electrodes in a systematic fashion (Dorman et al., 1994) which follows the tonotopical arrangement of the cochlea. Spikes are entrained to the stimulus in electrical stimulation (Javel, 1990), so if the phase-lock code was the only mechanism operating in frequency discrimination or pitch perception, stimuli on all electrodes would have had the same pitch. So electrode discrimination and pitch estimation experiments provide convincing arguments in favour of the rate-place code.

It is concluded that cochlear implants have not yet provided the final answers to the question of the coding of frequency in the auditory system.

4.4 Implications for cochlear implants

It is far easier to get high temporal resolution in electrical stimulation than it is to get high spectral resolution. Current spread from electrodes limit spectral resolution (Kral et al., 1998). New electrode designs may limit current spread (Cords et al., 2000), but certain physical limitations on electrode design remain. For example, maximum safe levels of charge density exist (Shannon, 1992). On the other hand, there are no basic technological limitations on increasing the stimulation rate. However, neural threshold adaptation may occur for high stimulation rates (above 400 pulses per second per channel), which suggests that higher stimulation rates may not be beneficial and may even degrade speech recognition performance (Javel and Shepherd, 2000). Still, the success of temporal pattern based strategies for cochlear implants like CIS is encouraging and warrants further study.

5 CONCLUSIONS

(1) To be able to predict frequency difference limens for acoustic stimulation, an important assumption is that one spike per stimulus cycle is available, which may be provided by the existence of a volley principle. The volley principle may be implemented by the cochlear nucleus, where neurons have been found that can improve temporal precision by combination of a number of auditory nerve inputs (Moller, 1999).

(2) An additional assumption is required in order to predict frequency difference limens for electrical stimulation of the auditory system. It is assumed that because many fibres fire on exactly the same phase of the electrical stimulation waveform, desynchronization results at a central auditory nervous system integration centre, which in turn leads to degradation in frequency discrimination.

(3) Psychoacoustic data from cochlear implants show that both mechanisms for the coding of frequency information in the auditory system are equally likely. Thus, though cochlear implants may provide a tool to solve this problem, they have not yet provided the final answer to the question of coding of frequency in the auditory system.

Chapter 8
CONCLUSION

This thesis intended to contribute to the understanding of the processing of sound in the central auditory nervous system, for both acoustic and electrical stimulation of the auditory system, through cochlear implant psychoacoustic research and modelling of the auditory system. A central hypothesis was that the same underlying mechanisms are responsible for sound perception in electric and acoustic hearing.

Both temporal (chapters 6 and 7) and spatial (chapters 3, 4 and 5) models were considered for acoustic (chapters 3, 5 and 6) and electric (chapters 4 and 7) hearing. In simple terms, the idea was to create a more comprehensive and integrated "picture" of what is known about the auditory system under both acoustic and electrical stimulation. The picture painted in this thesis is far from complete (see figure 3.1) and is really only a starting point in this mammoth task. The problem is multidimensional, as was illustrated in figure 1.3. The gaps in the "picture" are immediately evident, and even where chapters in this thesis have started to fill these gaps, it is not comprehensive by any means.

1 DISCUSSION OF HYPOTHESES

Qualitative proof of the hypotheses made in chapter 1 have been demonstrated in chapters 2 to 7. It is not possible prove each hypothesis conclusively, as is dicussed below.

1.1 Hypothesis 1

The primary hypothesis was that the same underlying mechanisms are responsible for sound perception in electric and acoustic hearing. It was shown that an acceptable model for acoustic hearing can predict psychoacoustic data from electric hearing when the model input is adjusted

in an appropriate way. Spike train statistics were different for the models of acoustic and electric gap detection. It was assumed that strong phase-locking occurs in electrical stimulation, while Poissonian statistics were assumed for acoustic stimulation. Furthermore, tuning (as reflected by H_m in equation 3.5) in the acoustic model was replaced by a model for current spread (paragraph 2.4.1 of chapter 4).

A classical signal detection theoretical model for acoustic gap discrimination and gap detection was created in chapter and equation 3.37 was obtained as the final expression for the gap detection threshold. The same principles used in chapter 3 were then applied in chapter 4 to obtain an expression for the gap detection threshold. A general result is given by equation 4.8, which may be compared to equation 3.37, but the final result is not in closed form. Paragraph 2.3 of chapter 4 gives the procedure for calculating the gap threshold for electric hearing.

Paragraph 3 in chapter 3 and paragraph 3 in chapter 4 show that gap detection data in both acoustic and electrical stimulation can be predicted by the same model if appropriate adjustments are made to the model input. This provides qualitative proof of hypothesis 1.

1.2 Hypothesis 2

It is conceivable that the brain employs prior knowledge to estimate input signals. A secondary hypothesis of this thesis was that an internal model (or an analysis-by-synthesis mechanism) may be used to implement detection and discrimination tasks in the auditory system. The Kalman filter discussed in chapters 6 and 7 is one example of such an internal model, while the related Markov model approach in chapter 5 is also a form of an internal model. The Kalman filter model is particularly attractive for several reasons. First, it is a simple recursive mechanism that does not require storage of information (as opposed to template-matching models, e.g. Gerson and Goldstein, 1978) but requires only prior knowledge of the dynamics of expected speech or environmental sounds. This may be learnt by learning to create speech sounds and by being exposed to speech and environmental sounds. Second, the Kalman filter

provides optimal tracking ability, i.e., when a sound has rapid transitions and given that the auditory system cannot react instantaneously to changes (see, for example, Zhang et al., 1990), the Kalman filter will lock onto new sounds faster than any other tracking mechanism.

This thesis has not proven conclusively that the auditory system uses an internal model to estimate sounds, but it has been shown by example (chapter 6) that a Kalman filter model that uses temporal information in spike trains can predict many of the characteristics of frequency discrimination data (figures 6.5 to 6.7).

Chapter 5 showed that a second kind of internal model, where signal dynamics are modelled by a Markov process, can also predict some of the characteristics of frequency discrimination. The Markov model assumes prior knowledge of state transition probabilities, where the states in this instance was the frequency of the pure tone input to the auditory system. This model was based on spatial information and its predictions are shown in figure 5.6.

1.3 Hypothesis 3

It was hypothesized that the same underlying mechanisms may be responsible for what is usually interpreted as "temporal" or "spatial" mechanisms. This is proved by examples. Gap detection thresholds (chapter 2), which have traditionally been explained in terms of temporal mechanisms, are explained in terms of spatial mechanisms (chapters 3 and 4; see for example figures 3.6 and 4.22) in this thesis. The auditory interpretation of frequency information, for which the underlying mechanisms are more controversial (e.g. Moller, 1999), is perhaps explained more often in terms of spatial mechanisms. However, here it is shown that frequency discrimination data may be explained in terms of either spatial (chapter 5) or temporal mechanisms (chapters 6 and 7; see for example figures 5.6 and 6.5 to 6.7).

1.4 Hypothesis 4

The hypothesis that electrode interaction can be measured with gap detection was investigated

in chapters 2 and 4. It was shown that gap thresholds increase as two electrode pairs that generated the marker stimuli were separated (figures 2.2 to 2.4). This was ascribed to increasingly disjunct excitation of neural populations as the electrode pairs were separated. Although these data should be interpreted with some care (as discussed in the introduction of chapter 4), it was shown in a model (chapter 4, paragraph 2.3) that across-channel spatial mechanisms may be responsible for gap detection thresholds under these conditions (e.g. figure 4.22). This does not prove that gap detection thresholds provide a reliable measure for electrode interaction, but the similarities between measured electrode discrimination data and electrode discrimination predicted from estimated current distributions (paragraph 3.3 in chapter 4) provide some support for the notion.

2 RESEARCH CONTRIBUTION

Apart from the hypotheses discussed in the previous paragraph that provided the central themes of this thesis, many findings were described in the self-contained chapters. Each chapter investigated a particular aspect, and the most important findings are summarized in this paragraph. These findings constitute a large part of the research contribution of this thesis.

(1) It was found that gap detection thresholds in electric hearing are a function of the physical separation of the electrode pairs used for the two stimuli that bound the gap. Gap thresholds increase from a minimum when the two stimuli are presented on the same electrode pair to a maximum when the two stimuli are presented on widely separated electrode pairs. This change may be due to a change-over from a peripheral, within-channel gap detection process for closely spaced electrode pairs to a central across-channel process for widely spaced electrode pairs. This suggests that gap detection thresholds may be used to measure electrode interaction, a notion that was made explicit in paragraph 3.3 of chapter 4.

(2) The extent of neural activation by each electrode probably varies across subjects and

across electrodes. For the three subjects in chapter 2, it seems the better implant users exhibit sharper tuning as measured with gap detection, although not enough data was available to confirm this (paragraph 4.1, chapter 2).

(3) Using stimulation modes with larger separation between active and reference electrodes has limited effect on spatial selectivity (paragraph 3.3 in chapter 2)

(4) The Cramer Rao Lower Bound for the Poisson change-point problem, which appears not to be documented in literature, was derived in chapter 3 (equation 3.28). The result is intuitively satisfying.

(5) A model for acoustic hearing that can predict the U-shaped curves found in across-channel gap detection (chapter 3, paragraph 3) has been created. The model is based on statistical signal detection theory considerations.

(6) This model shows that spatial mechanisms, as opposed to temporal mechanisms, may contribute to gap detection thresholds in the across-channel condition (paragraph 4.2, chapter 3). This is important in cochlear electrical stimulation, where spike trains are strongly phase-locked to the stimulus.

(7) The acoustic gap detection model of chapter 3 could predict realistic gap detection thresholds in auditory electrical stimulation (figure 4.22) when appropriate input spike train statistics (paragraph 2.3, chapter 3) and model parameters (paragraph 2.4, chapter 3) were used .

(8) Predictions for the current distribution in the cochlea of a specific implant user can be obtained from the gap detection tuning curves (paragraph 3.2, chapter 4). It is not known how accurate these predictions are, but they do show trends similar to those found in current distribution measurements (figure 4.32) and the predicted length constants of 0.5 mm to 3 mm correlate well with measured data quoted in literature.

Predictions of electrode discrimination show some similarity with electrode discrimination data (paragraph 3.3, chapter 4), although these predictions do not correlate well with measured data in two out of three subjects. It was shown that electrodes spaced closer than 1.5 mm are not discriminable.

(9) The sharp tip seen in some gap tuning curves in electric hearing is possibly obtained when entrainment is close to 100%, when the primary factor determining gap thresholds is probably the temporal dispersion of spike placement in response to a stimulus pulse (paragraph 2.3.1, chapter 4).

(10) The shallow bowl portion of the gap tuning curve in electric hearing is probably obtained when entrainment is not close to 100%, when gap thresholds are probably determined by standard deviation of the inter-spike interval pdf (paragraph 2.3.2.2, chapter 4).

(11) The primary factor determining gap thresholds in electric hearing is probably the shape of the current distribution (paragraph 3.1, chapter 4). Modelling results suggest that exponential current decay is not a good model of current distribution in the cochlea (figures 4.26 to 4.28). Possibly, for bipolar stimulation, sharper current peaks are obtained close to the electrode (figures 4.26 to 4.28; figure 4.32).

(12) It was shown that an internal model based or analysis-by-synthesis mechanism may be used to explain frequency discrimination data in acoustic hearing (paragraph 3, chapter 6). The Kalman filter model described here is an explicit implementation of an analysis-by-synthesis mechanism which provides the ability to produce numerical predictions of frequency discrimination data.

(13) An additional assumption is required to predict frequency difference limens for electric hearing with the temporal model of chapter 6. It has to be assumed that (because many fibres fire on the same phase of the electrical stimulation waveform), desynchronization

results at a central auditory nervous system integration centre. This will lead to degradation in frequency discrimination (paragrap 3, chapter 7).

(14) Overall, all the models show that human listeners do not make full use of all the information relating to frequency or time that is available in the auditory nerve spike trains. With certain choices of model parameters, the models can perform much better than human listeners (for example, see paragraph 4.8 of chapter 6).

3 IMPLICATIONS FOR COCHLEAR IMPLANTS

3.1 Speech processors for cochlear implants

A trade-off between maximum pulse rate and the number of electrodes exist in cochlear implants, and it is still not clear which is more important for improved speech recognition: more electrodes, or higher pulse rates? Psychoacoustic data from cochlear implants show that temporal and spatial mechanisms for the coding of frequency information in the auditory system are both plausible (chapter 7), although a general finding of this thesis is that temporal models are not good predictors of some of the psychoacoustic data for electric hearing. Specifically, temporal models cannot predict the across-channel gap detection thresholds of chapter 2. Although cochlear implants may provide a tool to investigate which mechanisms the central auditory nervous system uses to interpret sound, they have not yet provided the final answer to the question of coding of frequency information (chapter 7) or time information (chapter 4) in the auditory system.

However, the results of chapters 2 to 4 suggest that it is important to find ways to improve spatial resolution of cochlear implants. New electrode designs with improved current focussing ability (e.g. Cords et al., 2000) may provide an indication as to whether less channel interaction can be achieved, and whether this will in turn lead to improved speech recognition.

3.2 Improved individualized programming of cochlear implants

It is possible that the number of distinct channels in cochlear implants are small, as shown in chapter 2 and as can be deduced from the predicted current distributions in chapter 4. Speech processors may be programmed to use a subset of the available electrodes to achieve independent channels. Gap detection data (chapter 2) or electrode discrimination data may be used to guide decisions on which electrodes to include in a processor. Current distributions may be predicted from gap detection data, as shown in chapter 4. If specific electrodes exhibit larger current distributions than others, they may be excluded from a processor. Hanekom and Shannon (1996) were able to achieve improved speech recognition scores in cochlear implant users by judicious choice of electrodes for a reduced electrode processor. Thus, appropriate models that explain how performance in psychoacoustic experiments are related to physical parameters, may assist audiologists and researchers in achieving optimal individualized settings of cochlear implants.

4 FUTURE RESEARCH DIRECTIONS

The gaps in figure 1.3 provide a summary of some research work that needs to be done in future studies.

(1) Spatial models of frequency discrimination in acoustic hearing need to be expanded to electric hearing, especially since temporal models do not adequately describe frequency discrimination in the electrically-stimulated auditory system.

(2) The work in this thesis suggests that spatial mechanisms are more likely to explain the interpretation of frequency information in cochlear implants than temporal mechanisms. Temporal mechanisms cannot predict the gap detection data in chapter 2, and the model assumption required in chapter 7 to predict cochlear implant frequency discrimination data is disputable. Appropriate models for pitch discrimination (or electrode discrimination) and pitch perception in electric hearing need to be developed.

Although not discussed in this thesis, it is still not clear how spatial and temporal mechanisms co-operate to create the pitch sensation in either acoustic or electric hearing, but cochlear implants provide a unique opportunity to isolate these two mechanisms.

(3) Current distributions estimated from gap detection tuning curves have an intuitive appeal, but it is necessary to find ways to prove that predicted current distributions are fair reflections of the actual current distributions. Current distributions need to be estimated from other psychoacoustic data and compared with distributions estimated from gap detection data, but especially current distribution data from cochlear implant users are required. Voltage distributions that are often measured (see the references in paragraph 3.2 of chapter 4) do not adequately reflect current distributions in the cochlea.

(4) Although it seems that temporal models will not be able to predict gap detection data in electric hearing for the across-channel condition, this needs to be investigated further, especially since it is possible to predict within-channel gap detection data (not shown in this thesis) with temporal models.

(5) Techniques to implement detection and estimation mechanisms when the spike trains statistics are known, is required. Especially, the applicability of the internal model or analysis-by-synthesis approach needs to be investigated for more complex signals than was done in this thesis.

(6) An appropriate model for acoustic hearing can predict psychoacoustic data from electric hearing when the model input is correctly adjusted. This has been shown to be true for gap detection and (perhaps less convincingly) for frequency discrimination. This principle has to be applied to expand models to include other psychoacoustic phenomena (e.g., forward masking).

(7) Finally, as discussed in the previous paragraph, an important outflow of the modelling results is that it is necessary to focus research efforts on obtaining better spatial resolution in cochlear implants.

REFERENCES

Aidley, D. J. 1998, *The physiology of excitable cells*, Cambridge University Press, Cambridge.

Bacon, S. P. & Viemeister, N. F. 1985, "Temporal modulation transfer functions in normal-hearing and hearing-impaired listeners", *Audiology*, vol. 24, pp. 117-134.

Baldi, P. & Heiligenberg, W. 1988, "How sensory maps could enhance resolution through ordered arrangements of broadly tuned receivers", *Biological Cybernetics*, vol. 59, pp. 313-318.

Bialek, W., Rieke, F., de Ruyter van Steveninck, R. R., & Warland, D. 1991, "Reading a neural code", *Science*, vol. 252, pp. 1854-1857.

Black, R. C., Clark, G. M., Tong, Y. C., & Patrick, J. F. 1983, "Current distributions in cochlear stimulation", *Annals of the New York Academy of Sciences*, vol. 405 , pp. 137-145.

Blamey, P. J., Dooley, G. J., Parisi, E. S., & Clark, G. M. 1996, "Pitch comparisons of acoustically and electrically evoked auditory sensations", *Hearing Research*, vol. 99, pp. 139-150.

Bower, J. M. 1990, "Reverse engineering the nervous system: an anatomical, physiological, and computer-based approach," in *An introduction to neural and electronic networks*, S.F. Zornetzer, L.L. Davis & C. Lau, eds., Academic Press, San Diego.

Bremaud, P. 1981, *Point processes and cues*. Springer-Verlag, New York.

Bruce, I. C. 1997, *Spatiotemporal coding of sound in the auditory nerve for cochlear implants*. PhD Thesis, University of Melbourne.

Bruce, I. C., Irlicht, L. S., & Clark, G. M. 1998. A mathematical analysis of spatiotemporal summation of auditory nerve firings. *Information Sciences,* vol. 111, pp. 303-334.

Bruce, I. C., Irlicht, L. S., White, M. W., O'Leary, S. J., Dynes, S., Javel, E., & Clark, G. M. 1999a, "A stochastic model of the electrically stimulated auditory nerve: single-pulse response", *IEEE Transactions on Biomedical Engineering*, vol. 46 no 6, pp. 617-629.

Bruce, I. C., Irlicht, L. S., White, M. W., O'Leary, S. J., Dynes, S., Javel, E., & Clark, G. M. 1999b, "A stochastic model of the electrically stimulated auditory nerve: pulse-train response", *IEEE Transactions on Biomedical Engineering*, vol. 46 no 6, pp. 630-637.

Bruce, I. C., Irlicht, L. S., White, M. W., O'Leary, S. J., & Clark, G. M. 2000, "Renewal process approximation of a stochastic threshold model for electrical neural stimulation", *Journal of Computational Neuroscience*, vol. 9, pp. 119-132.

Brugge, J. F. 1992, "An overview of central auditory processing," in *The mammalian auditory pathway: neurophysiology*, A. N. Popper & R. R. Fay, eds., Springer-Verlag, New York, pp. 1-33.

Busby, P. A., Whitford, L. A., Blamey, P. J., Richardson, L. M., & Clark, G. M. 1994, "Pitch perception for different modes of stimulation using the cochlear muliple-electrode prosthesis", *Journal of the Acoustical Society of America*, vol. 95 no 5, pp. 2658-2669.

Burkitt, A.N. & Clark, G.M. 1999, "Analysis of integrate-and-fire neurons: synchronization of synaptic input and spike output", *Neural Computation*, vol. 11, pp. 871-901.

Busby, P. A. & Clark, G. M. 1999, "Gap detection by early-deafened cochlear-implant subjects", *Journal of the Acoustical Society of America*, vol. 105 no 3, pp. 1841-1852.

Chatterjee, M. & Shannon, R. V. 1998, "Forward masked excitation patterns in multielectrode electrical stimulation", *Journal of the Acoustical Society of America*, vol. 103 no 5, pp. 2565-2572.

Chatterjee, M., Fu, Q.-J., & Shannon, R. V. 1998, "Within-channel gap detection using dissimilar markers in cochlear implant listeners", *Journal of the Acoustical Society of America*, vol. 103 no 5, pp. 2515-2519.

Clark, G. M., Tong, Y. C., & Patrick, J. F. 1990, *Cochlear prostheses*, Churchill Livingstone, Edinburgh.

Clark, G. M. 1993, "The development of the Nucleus multiple-channel cochlear implant", *Australian Journal of Otolaryngology*, vol. 1(4), pp. 310-316.

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*, vol. 23, pp. 766-776.

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*, vol. 54 no 6, pp. 1458-1470.

Collins, L. M., Zwolan, T. A., & 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*, vol. 101 no 1, pp. 440-455.

Cook, E. P. & Johnston, D. 1999, "Voltage-dependent properties of dendrites that eliminate location-dependent variability of synaptic input", *Journal of Neurophysiology*, vol. 81, pp. 535-543.

Cords, S. M., Reuter, G., Issing, P. R., Sommer, A., Kuzma, J., & Lenarz, T. 2000, "A silastic positioner for a modiolus-hugging position of intracochlear electrodes: Electrophysiologic effects", *American Journal of Otology*, vol. 21, pp. 212-217.

Creelman, C. D. 1962, "Human discrimination of auditory duration", *Journal of the Acoustical Society of America*, vol. 34 no 5, pp. 582-593.

Davis, M. H. A. 1976, "A note on the Poisson disorder problem", Proceedings of a conference, Zakopane, January 1974, *Mathematical Control Theory*, pp. 65-72.

Delgutte, B. 1996, "Physiological models for basic auditory percepts.," in *Auditory Computation*, H. L. Hawkins et al., eds., Springer-Verlag, New York, pp. 157-220.

Delgutte, B. 1997, "Auditory neural processing of speech," in *The handbook of phonetic sciences*, W. J. Hardcastle & J. Laver, eds., Blackwell Publishers, Oxford, pp. 507-538.

Deutsch, S. & Deutsch, A. 1993, *Understanding the nervous system. An engineering perspective,* IEEE Press, New York.

Divenyi, P. L. & Danner, W. F. 1977, "Discrimination of time intervals marked by brief acoustic pulses of various intensities and spectra", *Perception & Psychophysics*, vol. 21 no. 2, pp. 125-142.

Divenyi, P. L. & Sachs, R. M. 1978, "Discrimination of time intervals bounded by tone bursts", *Perception & Psychophysics*, vol. 24 no. 5, pp. 429-436.

Donaldson, G. S. & Nelson, D. A. 2000, "Place-pitch sensitivity and its relation to consonant recognition by cochlear implant listeners using the MPEAK and SPEAK speech processing strategies", *Journal of the Acoustical Society of America*, vol. 107 no. 3, pp. 1645-1658.

Dorman, M. F., Smith, M., Smith, L., & Parkin, J. L. 1994, "The pitch of electrically presented sinusoids", *Journal of the Acoustical Society of America*, vol. 95 no. 3, pp. 1677-1679.

Dye, R. H. & Hafter, E. R. 1980, "Just-noticeable differences of frequency for masked tones", *Journal of the Acoustical Society of America*, vol. 67 no. 5, pp. 1746-1753.

Dynes, S. B. C. & Delgutte, B. 1992, "Phase-locking of auditory-nerve discharges to sinusoidal electric stimulation of the cochlea", *Hearing Research*, vol. 58, pp. 79-90.

Dynes, S. B. C. 1996, *Discharge characteristics of auditory nerve fibers for pulsatile electrical stimuli*, Doctoral dissertation, Massachusetts Institute of Technology, MIT Press, Cambridge.

Eddins, D. A. & Green, D. M. 1995, "Temporal integration and temporal resolution," in *Hearing*, B. C. J. Moore, ed., Academic Press, San Diego, pp. 207-242.

Erell, A. 1988, "Rate coding model for discrimination of simple tones in the presence of noise", *Journal of the Acoustical Society of America*, vol. 84 no. 1, pp. 204-214.

Evans, E. F. 1975, "Cochlear nerve and cochlear nucleus", in *Handbook of sensory physiology vol V/2*, W.D. Keidel & W.D. Neff, eds., Springer-Verlag, Berlin, pp. 1-108;

Fayad, J., Linthicum, F. H. Jr., Otto, S. R., Galey, F. R., & House, W. F. 1991, "Cochlear implants: Histopathologic findings related to performance in 16 human temporal bones", *Annals of Otology, Rhinology and Laryngology (United States)*, vol. 100 no. 10, pp. 807-811.

Finley, C. C., Wilson, B. S., & White, M. W. 1990, "Models of neural responsiveness to electrical stimulation," in *Cochlear Implants. Models of the electrically stimulated ear*, J. M. Miller & F. A. Spelman, eds., Springer-Verlag, New York, pp. 55-96.

Fishman, P. M., Shannon, R. V., & Slattery, W. H. 1997, "Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor", *Journal of Speech and Hearing Research*, vol. 40, pp. 1201-1215.

Fitzgibbons, P. J. 1983, "Temporal gap detection in noise as a function of frequency, bandwidth, and level", *Journal of the Acoustical Society of America*, vol. 74 no. 1, pp. 67-72.

Fitzgibbons, P. J. 1984, "Temporal gap resolution in narrow-band noises with center frequencies from 6000-14000 Hz", *Journal of the Acoustical Society of America*, vol. 75 no. 2, pp. 566-569.

Florentine, M. & Buus, S. 1984, "Temporal gap detection in sensorineural and simulated hearing impairments", *Journal of Speech and Hearing Research*, vol. 27, pp. 449-455.

Formby, C. & Forrest, T. G. 1991, "Detection of silent temporal gaps in sinusoidal markers", *Journal of the Acoustical Society of America*, vol. 89 no. 2, pp. 830-837.

Formby, C., Morgan, L. N., Forrest, T. G., & Raney, J. J. 1992, "The role of frequency selectivity in measures of auditory and vibrotactile temporal resolution", *Journal of the Acoustical Society of America* , vol. 91, pp. 293-305.

Formby, C., Sherlock, L. P., & Forrest, T. P. 1996, "An assymmetric roex filter model for describing detection of silent temporal gaps in sinusoidal markers", *Auditory Neuroscience*, vol. 3, pp. 1-20.

Forrest, T. G. & Formby, C. 1996, "Detection of silent temporal gaps in sinusoidal markers simulated with a single-channel envelope detector model", *Auditory Neuroscience*, vol. 3, pp. 21-33.

Friesen, L.M., Shannon, R.V., Baskent, D., & Wang, X., "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*, vol. 110, pp. 1150-1163.

Frijns, J. H. M., de Snoo, S. L., & Schoonhoven, R. 1995, "Potential distributions and neural excitation patterns in a rotationally symmetric model of the electrically stimulated cochlea," in *Cochlear Implants. A Modelling Approach*, J. H. M. Frijns, ed., CIP-Data Koninklijke Bibliotheek, Den Haag, pp. 93-124.

Fu, Q.-J., Shannon, R. V., Zeng, F.-G., & Chatterjee, M. 1996, "Electrode interactions measured by loudness summation in cochlear implant users", *Journal of the Acoustical Society of America*, vol. 100, p. 2631.

Fu, Q.-J. 1997. "Comparison of electrode interaction measures in multichannel cochlear implants" in *Abstracts of the 1997 annual midwinter meeting, Association for Research in Otolaryngology*, Abstract #305.

Fu, Q.-J. & Shannon, R. V. 1999, "Effects of electrode location and spacing on phoneme recognition with the Nucleus-22 cochlear implant", *Ear and Hearing*, vol. 20 no. 4, pp. 321-331.

Fujita, S. & Ito, J. 1999, "Ability of nucleus cochlear implantees to recognize music", *Annals of Otology, Rhinology and Laryngology*, vol. 108, pp. 634-640.

Gabbiani, F. & Koch, C. 1996, "Coding of time-varying signals in spike trains of integrate-and-fire neurons with random threshold", *Neural Computation*, vol. 8, pp. 44-66.

Gal'chuk, L. I. & Rozovskii, B. L. 1971, "The "disorder" problem for a poisson process", *Theory of probability and its applications*, vol. 16, pp. 712-716.

Gaumond, R. P., Molnar, C. E., & Kim, D. O. 1982, "Stimulus and recovery dependence of cat cochlear nerve fibre spike discharge probability", *Journal of Neurophysiology*, vol. 48 no. 3, pp. 856-873.

Goldstein, J. L. & Srulovicz, P. 1977, "Auditory-nerve spike intervals as an adequate basis for aural frequency measurement," in *Psychophysics and physiology of hearing*, E. F. Evans & J. P. Wilson, eds., Academic Press, London, pp. 337-347.

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

Green, D. M. 1973, "Minimum integration time," in *Basic mechanisms in hearing*, A. R. Moller, ed., Academic Press, New York, pp. 829-843.

Greenwood, D. D. 1990, "A cochlear frequency-position function for several species - 29 years later", *Journal of the Acoustical Society of America*, vol. 87 no. 6, pp. 2592-2605.

Grill, W. M. & Mortimer, J. T. 1994, "Electrical properties of implant encapsulation tissue", *Annals of Biomedical Engineering*, vol. 22, pp. 23-33.

Hall, J. W., Grose, J. H., & Joy, S. 1996, "Gap detection for pairs of noise bands: effects of stimulus level and frequency separation", *Journal of the Acoustical Society of America*, vol. 99 no. 2, pp. 1091-1095.

Hanekom, J. J. & Shannon, R. V. 1996, "Place pitch discrimination and speech recognition in cochlear implant users", *South African Journal of Communication Disorders*, vol. 43, pp. 27-40.

Hanekom, J. J. & Shannon, R. V. 1997. "Gap detection thresholds as a measure of spread of excitation in cochlear electrical stimulation" in *Abstracts of the 1997 annual midwinter meeting, Association for Research in Otolaryngology,* Abstract #307.

Hanekom, J. J. & Shannon, R. V. 1998, "Gap detection as a measure of electrode interaction in cochlear implants", *Journal of the Acoustical Society of America*, vol. 104 no. 4, pp. 2372-2384.

Hanekom, J. J. 1999, "A Model of frequency coding in the central auditory nervous system", *South African Journal of Communication Disorders*, vol. 46, pp. 81-89.

Hanekom, J. J. 2000, "What do cochlear implants teach us about the encoding of frequency in the auditory system?", *South African Journal of Communication Disorders*, vol. 47, pp. 49-56.

Hanekom, J. J. & Krüger, J. J. 2001, "A model of frequency discrimination with optimal processing of auditory nerve spike intervals", *Hearing Research*, vol. 151, pp. 188-204.

Hanekom, T. 2001, "Three-dimensional spiraling finite element model of the electrically stimulated cochlea", *Ear and Hearing*, vol. 22, pp. 300-315.

Hartmann, R., Topp, G., & Klinke, R. 1984, "Electrical stimulation of the cat cochlea - discharge pattern of single auditory fibres", *Advances in Audiology*, vol. 1, pp. 18-29.

Heinz, M. G., Goldstein, M. H., & Formby, C. 1996, "Temporal gap detection in sinusoidal markers simulated with a multi-channel, multiresolution model of the auditory periphery", *Auditory Neuroscience*, vol. 3, pp. 35-56.

Heinz, M. G., Colburn, H.S., & Carney, L.H. 2001, "Evaluating Auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve", *Neural Computation*, vol. 13, pp. 2273-2316.

Henry, B. A., McKay, C. M., McDermott, H. J., & Clark, G. M. 1997. The relationship between speech information perceived by cochlear implantees in different spectral regions and electrode discrimination. *Program and Abstracts of the 1997 Conference on implantable auditory prostheses,* p. 74.

Hienz, R. D., Sachs, M. B., & Aleszczyk, C. M. 1993, "Frequency discrimination in noise: comparison of cat performances with auditory nerve models", *Journal of the Acoustical Society of America*, vol. 93 no 1, pp. 462-469.

Hirsch, H. G. 1993, "Intelligibility improvements of noisy speech for people with cochlear implants", *Speech Communication*, vol. 12, pp. 261-266.

Ifukube, T. & White, R. L. 1987, "Current distributions produced inside and outside the cochlea from a scala tympani electrode array", *IEEE Transactions on Biomedical Engineering*, vol. BME-34, pp. 883-890.

Javel, E. & Mott, J. B. 1988, "Physiological and psychophysical correlates of temporal processes in hearing", *Hearing Research*, vol. 34, pp. 275-294.

Javel, E. 1990, "Acoustic and electrical encoding of temporal information," in *Cochlear implants: models of the electrically stimulated ear*, J. M. Miller & F. A. Spelman, eds., Springer-Verlag, New York, pp. 247-295.

Javel, E. & 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*, vol. 107, pp. 908-921.

Javel, E. & Shepherd, R. K. 2000, "Electrical stimulation of the auditory nerve. III. Response initiation sites and temporal fine structure", *Hearing Research*, vol. 140, pp. 45-76.

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*, vol. 68 no. 4, pp. 1115-1122.

Johnson, D. H. & Swami, A. 1983, "The transmission of signals by auditory-nerve fiber discharge patterns", *Journal of the Acoustical Society of America*, vol. 74 no. 2, pp. 493-501.

Johnson, D. H. 1996, "Point process models of single-neuron discharges", *Journal of Computational Neuroscience*, vol. 3, pp. 275-299.

Johnston, D. & Wu, S. M. S. 1995, *Foundations of cellular neurophysiology,* MIT Press, Cambridge.

Johnstone, B. M., Patuzzi, R., & Yates, G. K. 1986, "Basilar membrane measurements and the travelling wave", *Hearing Research*, vol. 22, pp. 147-153.

Kalman, R. E. 1960, "A new approach to linear filtering and prediction problems", *Transactions of the ASME. Journal of Basic Engineering*, vol. 82, pp. 35-45.

Kandel, E.R., Schwartz, J.H., & Jessell, T.M. 1991, *Principles of neural science*, third edition, Appleton and Lange, Norwalk.

Karr, A. F. 1986, *Point processes and their statistical inference*, Marcel Dekker, New York, pp. 202-203.

Kay, S. M. 1998, *Fundamentals of statistical signal processing. Volume 2. Detection Theory* Prentice Hall, Upper Saddle River.

Keidel, W. D., Kallert, S., & Korth, M. 1983, *The Physiological Basis of Hearing*, Thieme-Stratton Inc., New York.

Kiang, N. Y.-S., Watanabe, T., Thomas, E. C., & Clark, L. F. 1965, *Discharge patterns of single fibers in the cat's auditory nerve*, Research Monograph no. 35, MIT Press, Cambridge.

Kileny, P. R., Zimmerman-Phillips, S., Zwolan, T. A., & Kemink, J. L. 1992, "Effects of channel number and place of stimulation on performance with the Cochlear Corporation multichannel implant", *American Journal of Otology*, vol. 13 no. 2, pp. 117-123.

Kim, D. O. & Molnar, C. E. 1979, "A population study of cochlear nerve fibers: comparison of spatial distributions of average-rate and phase-locking measures of responses to single tones", *Journal of Neurophysiology*, vol. 42 no. 1, pp. 16-30.

Kim, D. O. 1984, "Functional roles of the inner- and outer-hair-cell subsystems in the cochlea and brainstem," in *Hearing Science*, C. Berlin, ed., College-Hill Press, San Diego, pp. 241-262.

Kim, D. O., Chang, S. O., & Sirianni, J. G. 1990, "A population study of auditory-nerve fibers in unanesthetized decerebrate cats: response to pure tones", *Journal of the Acoustical Society of America*, vol. 87 no. 4, pp. 1648-1655.

Kim, D. O., Parham, K., Sirianni, J. G., & Chang, S. O. 1991, "Spatial response profiles of posteroventral cochlear nucleus neurons and auditory-nerve fibers in unanesthetized decerebrate cats: response to pure tones", *Journal of the Acoustical Society of America*, vol. 89 no. 6, pp. 2804-2817.

Kim, D. O. & 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*, vol. 52, pp. 167-180.

Kistler, W. M., Gerstner, W., & van Hemmen, J. L. 1997, "Reduction of the Hodgkin-Huxley equations to a single-variable threshold model", *Neural Computation*, vol. 9, pp. 1015-1045.

Koch, C., Poggio, T., & Torre, V. 1982, "Retinal ganglion cells: a functional interpretation of dendritic morphology", *Philosophical Transactions of the Royal Society of London, part B*, vol. 298, pp. 227-264.

Kou, B. S., Shipp, D. B., & Nedzelski, J. M. 1994, "Subject benefits reported by adult Nucleus 22-channel cochlear implant users", *Journal of Otolaryngology (Canada)*, vol. 23, no. 1, pp. 8-14.

Kral, A., Hartmann, R., Mortazavi, D., & Klinke, R. 1998, "Spatial resolution of cochlear implants: the electrical field and excitation of auditory afferents", *Hearing Research*, vol. 121, pp. 11-28.

Laming, D. 1986, *Sensory analysis,* Academic Press, London, chapter 4.

Langner, G. 1992, "Periodicity coding in the auditory system", *Hearing Research*, vol. 60, pp. 115-142.

Lawson, D., Wilson, B. S., & Finley, C. C. 1993, "New processing strategies for multichannel cochlear prostheses," in *Natural and Artificial Control of Hearing and Balance*, J. A. Allum et al., eds., Progress in Brain Research, vol. 97, Elsevier, Amsterdam, pp. 313-321.

Lawson, D., Wilson, B. S., & Finley, C. C. 1996, *Speech processors for auditory prostheses*, Third Quarterly Progress Report, NIH Contract N01-DC-5-2103.

Lecar, H. & Nossal, R. 1971, "Theory of threshold fluctuations in nerves. I. Relationships between electrical noise and fluctuations in axon firing", *Biophysical Journal*, vol. 11, pp. 1048-1067.

Levitt, H. 2001, "Transformed up-down methods in psychoacoustics", *Journal of the Acoustical Society of America*, vol. 49, pp. 467-477.

Lewis, E. R. 1996, "Further computations involving time," in *Auditory Computation*, H. L. Hawkins et al., eds., Springer-Verlag, New York, pp. 469-504.

Lim, H. H., Tong, Y. C., & Clark, G. M. 1989, "Forward masking patterns produced by intracochlear electrical stimulation of one and two electrode pairs in the human cochlea", *Journal of the Acoustical Society of America*, vol. 86 no. 3, pp. 971-980.

Linthicum, F. H. Jr., Fayad, J., Otto, S. R., Galey, F. R., & House, W. F. 1991, "Cochlear implant histopathology", *American Journal of Otology*, vol. 12, no. 4, pp. 245-311.

Loizou, P. C. 1999, "Signal-processing techniques for cochlear implants", *IEEE Engineering in Medicine and Biology Magazine*, vol. 18 no. 3, pp. 34-46.

Maron, M. E. 2001, "On cybernetics, information processing and thinking," in *Cybernetics of the nervous system*, N. Wiener & J. P. Schade, eds., Elsevier, Amsterdam, pp. 118-138.

Maršálek, P., Koch, C. & Maunsell, J. 1997, "On the relationship between synaptic input and spike output jitter in individual neurons", *Proceedings of the National Academy of Science*, vol. 94, pp. 735-740.

McDermott, H. 1989, "An Advanced Multiple Channel Cochlear Implant", *IEEE Transactions on Biomedical Engineering*, vol. 36, pp. 789-797.

McDermott, H. J., McKay, C. M., Vandali, A. E., & Clark, G. M. 1992, "A comparison of speech perception of cochlear implantees using the Spectral Maxima Sound Processor (SMSP) and the MSP (Multipeak) processor", *Acta Otolaryngologica (Stockholm)*no. 112, pp. 752-761.

McDonough, R. N. & Whalen, A. D. 1995, *Detection of signals in noise*, 2 edn, Academic Press, San Diego.

McKinney, M. F. & Delgutte, B. 1999, "A possible neurophysiological basis of the octave enlargement effect", *Journal of the Acoustical Society of America*, vol. 106 no. 5, pp. 2679-2692.

McLaren, I. 1989, "The computational unit as an assebly of neurones: an implementation of an error correcting learning algorithm," in *The computing neuron*, R. Durbin, C. Miall, & G. Mitchison, eds., Addison-Wesley, Wokingham, pp. 160-179.

Mendel, J. M. 1995, *Lessons in estimation theory for signal processing, communications and control,* Prentice Hall, Englewood Cliffs.

Miller, J. M. & Spelman, F. A. 1990, *Cochlear Implants: models of the electrically stimulated ear,* Springer-Verlag, New York.

Moller, A. R. 1999, "Review of the roles of temporal and place coding of frequency in speech discrimination", *Acta Otolaryngologica (Stockholm)*, vol. 119, pp. 424-430.

Moore, B. C. J. 1973, "Frequency difference limens for short-duration tones", *Journal of the Acoustical Society of America*, vol. 54 no. 3, pp. 610-619.

Moore, B. C. J. & Glasberg, B. R. 1986, "The role of frequency selectivity in the perception of loudness, pitch and time," in *Frequency selectivity in hearing*, B. C. J. Moore, ed., Academic Press, London, pp. 251-308.

Moore, B. C. J. 1993, "Frequency analysis and pitch perception," in *Human psychophysics*, W. A. Yost, A. N. Popper, & R. R. Fay, eds., Springer-Verlag, New York, pp. 56-115.

Moore, B. C. J. & Sek, A. 1996, "Detection of frequency modulation at low modulation rates: evidence for a mechanism based on phase locking", *Journal of the Acoustical Society of America*, vol. 100 no. 4, pp. 2320-2331.

Moore, B. C. J. 1997, "Aspects of auditory processing related to speech perception," in *The handbook of phonetic sciences*, W. J. Hardcastle & J. Laver, eds., Blackwell Publishers, Oxford, pp. 539-565.

Nelson, D. A., Van Tasell, D. J., Schroder, A. C., Soli, S., & Levine, S. 1995, "Electrode ranking of "place pitch" and speech recognition in electrical hearing", *Journal of the Acoustical Society of America* , vol. 98 no. 4, pp. 1987-1999.

O'Leary, S. J., Tong, Y. C., & Clark, G. M. 1995, "Responses of dorsal cochlear nucleus single units to electrical pulse train stimulation of the auditory nerve with a cochlear implant electrode", *Journal of the Acoustical Society of America*, vol. 97 no. 4, pp. 2378-2393.

Oxenham, A. J. 2000, "Influence of spatial and temporal coding on auditory gap detection", *Journal of the Acoustical Society of America*, vol. 107 no. 4, pp. 2215-2223.

Palmer, A. R. & Russell, I. J. 1986, "Phase-locking in the cochlear nerve of the guinea pig and its relation to the receptor potential of inner hair-cells", *Hearing Research*, vol. 24, pp. 1-15.

Patterson, R. D. & Moore, B. C. J. 1986, "Auditory filters and excitation patterns as representations of frequency resolution," in *Frequency selectivity in hearing*, B. C. J. Moore, ed., Academic Press, London, pp. 123-177.

Penner, M. J. 1976, "The effect of marker variability on the discrimination of temporal intervals", *Perception & Psychophysics*, vol. 19 no. 466, p. 469.

Penner, M. J. 1977, "Detection of temporal gaps in noise as a measure of the decay of auditory sensation", *Journal of the Acoustical Society of America*, vol. 61, pp. 552-557.

Pfingst, B. E. 1988, "Comparisons of psychophysical and neurophysiological studies of cochlear implants", *Hearing Research*, vol. 34, pp. 243-252.

Pfingst, B. E., Holloway, L. A., Zwolan, T. A., & Collins, L. M. 1999, "Effects of stimulus level on electrode-place discrimination in human subjects with cochlear implants", *Hearing Research*, vol. 134, pp. 105-115.

Phillips, D. P., Taylor, T. M., Hall, S. E., Carr, M. M., & Mossop, J. E. 1997, "Detection of silent intervals between noises activating different perceptual channels: Some properties of "central" auditory gap detection", *Journal of the Acoustical Society of America*, vol. 101 no. 6, pp. 3694-3705.

Phillips, D. P. & Hall, S. E. 2000, "Independence of frequency channels in auditory temporal gap detection", *Journal of the Acoustical Society of America*, vol. 108 no. 6, pp. 2957-2963.

Plomp, R. 1964, "Rate of decay of auditory sensation", *Journal of the Acoustical Society of America*, vol. 36, pp. 277-282.

Pollack, I. 1967, "Number of pulses required for minimal pitch", *Journal of the Acoustical Society of America*, vol. 40, p. 895.

Pouget, A., Zhang, K., Deneve, S., & Latham, P. E. 1998, "Statistically efficient estimation using population coding", *Neural Computation*, vol. 10, pp. 373-401.

Preece, J. P. & Tyler, R. S. 1989, "Temporal-gap detection by cochlear prosthesis users", *Journal of Speech and Hearing Research*, vol. 32, pp. 849-856.

Prosen, C. A., Moody, D. B., Sommers, M. S., & Stebbins, W. C. 1990, "Frequency discrimination in the monkey", *Journal of the Acoustical Society of America*, vol. 88 no. 5, pp. 2152-2158.

Rabiner, L. & Juang, B-H. 1993, *Fundamentals of speech recognition*, Prentice Hall, Englewood Cliffs.

Raftery, A. E. & Akman, V. E. 1986, "Bayesian analysis of a Poisson process with a change-point", *Biometrika*, vol. 73 no. 1, pp. 85-89.

Rattay, F. & Motz, H. 1986, "A study of the application of the Hodgkin-Huxley and the Frankenhauser-Huxley model for electrostimulation of the acoustic nerve", *Neuroscience*, vol. 18 no. 3, pp. 699-712.

Rattay, F. 1990, *Electrical nerve stimulation. Theory, experiments and applications,* Springer-Verlag, Vienna.

Rattay, F. 1999, "The basic mechanism for the electrical stimulation of the nervous system", *Neuroscience*, vol. 89, pp. 335-346.

Rebscher, S. J., Snyder, R. L., Brasell, J., Bruszewski, W., Leake, P. A., & Merzenich, M. M. 1991, "Studies on pediatric auditory prosthesis implants", *Neuroprosthesis Contract NO-DC-0-2401*, 6th Quarterly Progress Report, pp. 1-26.

Reza, A. M. & Doroodchi, M. 1996, "Cramer-Rao Lower Bound on locations of sudden changes in a steplike signal", *IEEE Transactions on Signal Processing*, vol. 44 no. 10, pp. 2551-2556.

Rhode, W. S. & Greenberg, S. 1992, "Physiology of the cochlear nuclei," in *The mammalian auditory pathway: neurophysiology*, A. N. Popper & R. R. Fay, eds., Springer-Verlag, New York, pp. 94-152.

Rieke, F.,Warland, D., de Ruyter van Steveninck, R., & Bialek, W. 1997, *Spikes. Exploring the neural code*, MIT Press, Cambridge MA.

Rose, J. E., Brugge, J. F., Anderson, D. J., & Hind, J. E. 1968, "Patterns of activity in single auditory nerve fibres of the squirrel monkey," in *Hearing mechanisms in vertebrates*, A. V. S. de Reuck & J. Knight, eds., J. & A. Churchill Ltd., London, pp. 144-157.

Ross, S.M. 1983, *Stochastic processes*, John Wiley & Sons, New York.

Rudemo, M. 1972, "Doubly stochastic poisson processes and process control", *Advances in Applied Probability*, vol. 4 no. 2, pp. 318-338.

Ruggero, M. A. & Rich, N. C. 1989, " "Peak splitting": intensity effects in cochlear afferent responses to low frequency tones," in *Cochlear Mechanics: structure, function and models*, J. P. Wilson & D. T. Kemp, eds., Plenum Press, New York, pp. 259-267.

Ruggero, M. A. 1992, "Physiology and coding of sound in the auditory nerve," in *The mammalian auditory pathway: neurophysiology* , A. N. Popper & R. R. Fay, eds., Springer-Verlag, New York, pp. 34-93.

Sachs, M. B. & Miller, M. I. 1985, "Pitch coding in the auditory nerve: possible mechanisms of pitch sensation with cochlear implants," in *Cochlear Implants*, R. A. Schindler & M. M. Merzenich, eds., Raven Press, New York, pp. 185-194.

Schindler, R. A. & Kessler, D. K. 1993, "Clarion cochlear implant: phase I investigational results", *American Journal of Otology*, vol. 14, pp. 263-272.

Schroeder, M. R. 1968, "Period histogram and product spectrum: new methods for fundamental-frequency measurement", *Journal of the Acoustical Society of America*, vol. 43 no. 4, pp. 829-834.

Sek, A. & Moore, B. C. J. 1995, "Frequency discrimination as a function of frequency, measured in several ways", *Journal of the Acoustical Society of America*, vol. 97 no. 4, pp. 2479-2486.

Shadlen, M. N. & Newsome, W. T. 1994, "Noise, neural codes and cortical organization", *Current Opinion in Neurobiology*, vol. 4, pp. 569-579.

Shailer, M. J. & Moore, B. C. J. 1983, "Gap detection as a function of frequency, bandwidth, and level", *Journal of the Acoustical Society of America*, vol. 74, pp. 467-473.

Shailer, M. J. & Moore, B. C. J. 1987, "Gap detection and the auditory filter: phase effects using sinusoidal stimuli", *Journal of the Acoustical Society of America*, vol. 81 no. 4, pp. 1110-1117.

Shannon, R. V. 1983a, "Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics", *Hearing Research*, vol. 11, pp. 157-189.

Shannon, R. V. 1983b, "Multichannel electrical stimulation of the auditory nerve in man. II. Channel interaction", *Hearing Research*, vol. 12, pp. 1-16.

Shannon, R. V. 1985, "Loudness summation as a measure of channel interaction in a cochlear prosthesis," in *Cochlear implants*, R. A. Schindler & M. M. Merzenich, eds., Raven Press, New York, pp. 323-334.

Shannon, R. V. 1989, "Detection of gaps in sinusoids and pulse trains by patients with cochlear implants", *Journal of the Acoustical Society of America*, vol. 85 no. 6, pp. 2587-2592.

Shannon, R. V. & Otto, S. R. 1990, "Psychophysical measures from electrical stimulation of the human cochlear nucleus", *Hearing Research*, vol. 47, pp. 159-168.

Shannon, R. V., Adams, D. D., Ferrel, R. L., Palumbo, R. L., & Grandgenett, M. 1990, "A computer interface for psychophysical and speech research with the Nucleus cochlear implant", *Journal of the Acoustical Society of America*, vol. 87 no. 2, pp. 905-907.

Shannon, R. V. 1992, "A model of safe levels for electrical stimulation", *IEEE Transactions on Biomedical Engineering*, vol. 39 no. 4, pp. 424-426.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. 1995, "Speech recognition with primarily temporal cues", *Science*, vol. 270, pp. 303-304.

Shepherd, R. K., Hatsushika, S., & Clark, G. M. 1993, "Electrical stimulation of the auditory nerve: the effect of electrode position on neural excitation", *Hearing Research*, vol. 66, pp. 108-120.

Shepherd, R. K. & Javel, E. 1997, "Electrical stimulation of the auditory nerve. I. Correlation of physiological responses with cochlear status", *Hearing Research*, vol. 108, pp. 112-144.

Shepherd, R. K. & Javel, E. 1999, "Electrical stimulation of the auditory nerve: II. Effect of stimulus waveshape on single fibre response properties", *Hearing Research*, vol. 130, pp. 171-188.

Shofner, W. P. & Sachs, M. B. 1986, "Representation of a low-frequency tone in the discharge rate of populations of auditory nerve fibers", *Hearing Research*, vol. 21, pp. 91-95.

Siebert, W. M. 1970, "Frequency discrimination in the auditory system: place or periodicity mechanisms?", *Proceedings of the IEEE*, vol. 58 no. 5, pp. 723-730.

Skinner, M., Clark, G.M., Whitford, L., Seligman, P., Staller, S. 1994, "Evaluation of a new spectral peak coding strategy for the Nucleus 22 channel cochlear implant system", *American Journal of Otology*, vol. 15 (suppl. 2), pp. 15-27.

Smith, R. L. 1988, "Encoding of sound intensity by auditory neurons," in *Auditory Function. Neurobiological Bases of Hearing*, G. M. Edelman, W. E. Gall, & W. M. Cowan, eds., College-Hill Press, San Diego, pp. 243-274.

Snippe, H. P. & Koenderink, J. J. 1992, "Discrimination thresholds for channel-coded systems", *Biological Cybernetics*, vol. 66, pp. 543-551.

Snyder, D. L. 1975, *Random point processes,* Wiley, New York.

Spelman, F. A., Pfingst, B. E., Clopton, B. M., Jolly, C. N., & Rodenhiser, K. L. 1995, "Effects of electrical current configuration on potential fields in the electrically stimulated cochlea: field models and measurements", *Ann Otol Rhinol Laryngol Suppl*, vol. 104 no. 9 II, pp. 131-136.

Spoendlin, H. & Schrott, A. 1989, "Analysis of the human auditory nerve", *Hearing Research*, vol. 43, pp. 25-38.

Srulovicz, P. & 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*, vol. 73 no. 4, pp. 1266-1276.

Stevens, K. N. & Halle, M. 1967, "Remarks on analysis by synthesis and distinctive features," in *Models for the perception of speech and visual form*, W. Wathen-Dunn, ed., MIT Press, Cambridge, Massachusetts, pp. 88-102.

Sutter, M. L. 2000, "Shapes and level tolerances of frequency tuning curves in primary auditory cortex: quantitative measures and population codes", *Journal of Neurophysiology*, vol. 84, pp. 1012-1025.

Todd, N. P. M., O'Boyle, D. J., & Lee, C. S. 1999, "A sensory-motor theory of rhythm, time perception and beat induction", *Journal of New Music Research*, vol. 28 no. 1, pp. 5-28.

Tomlinson, R. W. W. & Langner, G. 1998. *Temporal processing in the auditory system: the functional significance of neural noise,* NATO Advanced Study Institute, pp. 123-128.

Townshend, B. & White, R. L. 1987, "Reduction of Electrical Interaction in Auditory Prostheses", *IEEE Transactions on Biomedical Engineering*, vol. 34 no. 11, pp. 891-897.

Townshend, B., Cotter, N., Van Compernolle, D., & White, R. L. 1987, "Pitch perception by cochlear implant subjects", *Journal of the Acoustical Society of America*, vol. 82 no. 1, pp. 106-115.

van den Honert, C. & Stypulkowski, P. H. 1987a, "Temporal response patterns of single auditory nerve fibers elicited by periodic electrical stimuli", *Hearing Research*, vol. 29, pp. 207-222.

van den Honert, C. & Stypulkowski, P. H. 1987b, "Single fiber mapping of spatial excitation patterns in the electrically stimulated auditory nerve", *Hearing Research*, vol. 29, pp. 195-206.

van Wieringen, A. & Wouters, J. 1999, "Gap detection in single- and multiple-channel stimuli by LAURA cochlear implantees", *Journal of the Acoustical Society of America*, vol. 106 no. 4, pp. 1925-1939.

Wakefield, G. H. & 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*, vol. 77 no. 2, pp. 613-619.

Wang, G., Vannier, M. W., Skinner, M. W., Kalender, W. A., Polacin, A., & Ketten, D. R. 1996, "Unwrapping cochlear implants by spiral CT", *IEEE Transactions on Biomedical Engineering*, vol. 43 no. 9, pp. 891-900.

West, R. W. & Ogden, R. T. 1994. *Continuous-time estimation of a change-point in a Poisson process*. Technical Report no. 185, Dept of Statistics, University of South Carolina.

Westerman, L. A. & Smith, R. L. 1984, "Rapid and short-term adaptation in auditory nerve responses", *Hearing Research*, vol. 15, pp. 249-260.

Wever, E. G. 1949, *Theory of hearing,* Wiley, New York.

Wiegrede, L. & Winter, I. M. 2000, "Psychophysics and physiology of regular-interval noise: critical experiments for current pitch models and evidence for a 1st order, temporal pitch code," in *Physiological and psychophysical bases of auditory function*, A. J. M. Houtsma et al., eds., Proceedings of the 12th International Symposium of Hearing, Mierlo, The Netherlands, pp. 106-112.

Wier, C. C., Jesteadt, W., & Green, D. M. 1977, "Frequency discrimination as a function of frequency and sensation level", *Journal of the Acoustical Society of America*, vol. 61 no. 1, pp. 178-184.

Wilson, B., Finley, C., Lawson, D., Wolford, R., Eddington, B., & Rabinowitz, B. 1991, "Better speech recognition with cochlear implants", *Nature*, vol. 352, pp. 236-238.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. 1995, "An internal model for sensorimotor integration", *Science*, vol. 269, pp. 1880-1882.

Zhang, W., Salvi, R. J., & Saunders, S. S. 1990, "Neural correlates of gap detection in auditory nerve fibers of the chinchilla", *Hearing Research*, vol. 46, pp. 181-200.

Zimmerman, C. E., Burgess, B. J., & Nadol, J. B. 1995, "Patterns of degeneration in the human cochlear nerve", *Hearing Research* , vol. 90, pp. 192-201.

Zwicker, E. & Fastl, H. 1990, *Psychoacoustics. Facts and Models.* Springer-Verlag, Berlin.

Zwolan, T. A., Collins, L. M., & Wakefield, G. H. 1997, "Electrode discrimination and speech recognition in postlingually deafened adult cochlear implant subjects", *Journal of the Acoustical Society of America*, vol. 102 no. 6, pp. 3673-3685.

Zwolan, T. A., Kileny, P. R., Boerst, A. K., Collins, L. A., & Telian, S. A. 1997, "Electrode discrimination using monopolar and bipolar stimulation with the Nucleus multichannel cochlear implant, " in *Abstracts of the 1997 annual midwinter meeting, Association for Research in Otolaryngology*, Abstract #231.