Suyash Narendra Joshi

Modeling auditory nerve responses to electrical stimulation

Modeling auditory nerve responses to electrical stimulation

PhD thesis by

Suyash Narendra Joshi

Preliminary version: January 16, 2017

HEARING SYSTEMS

Technical University of Denmark

2017

Supervisors

Assoc. Prof. Bastian Epp
Hearing Systems Group,
Department of Electrical Engineering,
Technical University of Denmark,
Kgs. Lyngby, Denmark

Prof. Torsten Dau
Hearing Systems Group,
Department of Electrical Engineering,
Technical University of Denmark,
Kgs. Lyngby, Denmark

Abstract

Cochlear Implants (CIs) are auditory prostheses that are prescribed for treatment of severe hearing loss or deafness. CIs bypass the impaired cochlear processes and directly stimulate the auditory nerve fibers (ANFs) to restore the sense of hearing. In particular, CIs extract the slowly varying envelope of the acoustic signal and stimulate the ANFs with a train of charge-balanced biphasic pulses that are amplitude modulated by the envelope of the acoustic signal. Therefore, the success of CIs primarily depends on the fidelity of the envelope representation in the responses of the ANFs. Understanding the factors that affect robust encoding of the envelope-related cues in the electrically stimulated ANFs is thus necessary for developing advanced CI stimulation strategies. This thesis describes the development and evaluation of a model framework to study the stimulus related and neural factors that affect the representation of the envelope in the electrically stimulated ANF responses.

The main rationale underlying the development of the model was that any arbitrary pulse shape can be regarded as a combination of anodic and cathodic pulses. The responses and interaction of these pulse phases should then reflect the ANF responses to arbitrary electrical stimuli. Based on this, the model of ANF responses to electrical stimulation was initially developed to simulate the responses for stimulation with single monophasic anodic and cathodic pulses. Clinical CIs stimulate the ANFs however with trains of charge-balanced biphasic pulses, where only the cathodic phase is intended to depolarize the neural membrane and generate a spike while the anodic phase is intended neutralizes the electrical charge to avoid electrolysis in the cochlea. The current is however injected into the extracellular space rather than intracellularly which produces depolarizing- and hyperpolarizing regions throughout the cochlea. Hence, both the anodic and the cathodic phases of biphasic pulses to generate a spike, but at different sites along the ANF. Taking this into account, the model accounts for the data for stimulation with single pulses of different pulse shapes and dynamic stimuli such as pulse trains of different pulse rates. It was shown that a the responsiveness to both anodic and cathodic pulses could successfully be generalized to account for the responses to

stimulation with a broad variety of pulse shapes. It could also be shown that the site of uncertainty affects the representation of the stimulus envelope in the simulated ANF responses as evaluated in a modulation detection task. In addition, with the same approach could be used to account for the main aspects of behavioral data on detection of interaural time differences.

Overall, the results of this thesis suggest that the uncertainty in the site of spike generation affects the temporal coding in the electrically stimulated ANFs. Furthermore, the proposed model can account for the effect of different stimulation parameters on the behavioral measures of temporal encoding in the electrically stimulated ANFs. The model hence provides a useful framework to evaluate different CI stimulation strategies and their effect on the behavioral performance of the CI listeners.

Resumé

Cochlear Implantater (CI) er høreimplantater, der ordineres til behandling af svær hørenedsættelse eller døvhed. CI omgår de pathologier svækkede processer i cochlear og stimulerer direkte de auditive nervefibre (ANF) for at genoprette høresansen. Specifikt udtrækkes den langsomt fluktuerende indhyldningskurve fra det akustiske signal i CIs og ANFs stimuleres med en række ladnings-balancerede bifasiske pulser, der er amplitudemodulerede med indhyldningskurven af det akustiske signal. Derfor afhænger successen med CIs primært af nøjagtigheden af repræsentationen af indhyldningskurven i ANF signalet. En forståelse af de faktorer, der påvirker kodningen af de indhyldningskurve-relaterede information i de elektrisk stimulerede ANFs er således nødvendig for udviklingen af avancerede CI stimuleringsstrategier. Denne afhandling beskriver udviklingen og evalueringen af en model for at studere stimulus relaterede og neurale faktorer, der påvirker repræsentation af indhyldningskurven i de elektrisk stimulerede ANF signaler.

Det vigtigste grundlag for udvikling af modellen var, at enhver vilkårlig CI puls kan betragtes som en kombination af anodiske og katodiske pulser. Responset og interaktionen mellem puls faserne skal dernæst afspejle ANF signalerne til vilkårlige elektriske stimuli. Baseret på dette, blev modellen af ANF signaler til en elektrisk stimulering oprindeligt udviklet for at simulere signaler til stimulering med enkelt monofasisk anodiske og katodiske pulser. Kliniske CIs stimulerer dog ANFs med en række ladnings-balancerede bifasiske pulser, hvor kun den katodiske fase skal depolarisere nervemembranen og generere en impuls, mens den anodiske fase skal neutralisere den elektriske ladning for at undgå elektrolyse i cochlea. Midlertidigt sættes injiceres strømmen i det ekstracellulære rum snarere end det intracellulære rum, hvilket producerer depolariserende- og hyperpolariserende regioner i cochlea. Derfor kan både de anodiske og katodiske faser af bifasiske impulser generere en impuls forskellige steder langs ANF. Tages dette i betragning, kan modellen redegøre for data for stimulering med enkelte impulser af forskellige puls former og dynamiske stimuli, såsom rækker af pulser med forskellig rate. Det blev påvist, at modellens modtagelighed overfor både anodiske og katodiske pulser, kan generaliseres til at redegøre for responset til stimulering med et bredt udvalg af pulsformer. Det blev også påvist, at usikkerheden i stedet påvirker repræsenta-

tionen af stimulus indhylningskurven i det simulerede ANF signal som evalueret i en modulations-detektions-opgave. Desuden kunne det med den samme fremgangsmåde anvendes til at redegøre for de vigtigste aspekter af subjektive data til detektering af den interaurale tidsforskel.

Samlet set tyder resultaterne i denne afhandling på, at usikkerheden i stedet for hvor impulsen genereres påvirker den tidsmæssige kodning i de elektrisk stimulerede ANFs. Endvidere kan den foreslåede model redegøre for effekten af forskellige stimuleringsparametre i de subjektive mål af den tidsmæssig kodning i de elektrisk stimulerede ANFs. Dermed giver modellen mulighed for at vurdere forskellige CI stimulerings strategier og deres effekt på brugere af CIs.

Acknowledgements

I would like to begin with a thank you to many of my friends and family, whose enormous support I could not have done without during the past three years. First, my two my advisers, Bastian Epp and Torsten Dau.

Bastian, you have been my best champion for the last three years. All of the crazy ideas we went on discussing for the countless hours, those you agreed with but also those you didn't, you have supported me throughout. Thank you for pushing me to find my own limits, but also making sure I was having fun doing it. Thank you for believing in me, particularly in times when I did not believe in myself. I have always counted on your friendship and advice, without which these three years would have been much difficult.

Torsten, thank you so much for taking a chance on me, giving me the opportunity to move to Denmark and be part of this fantastic group. Your excitement for hearing research and enthusiasm for learning has been a great source of inspiration. Thank you for giving me the freedom to explore the topic and to choose the path ahead. I would especially like to thank you for your help, patience and advice during the writing process. Without your advice, the entropy in telling of our story would have been much higher. Thank you for your critical and always constructive advice, it has been a great learning experience.

I would like to express my sincere gratitude to Walt Jesteadt. My journey in hearing research began at the Boys Town National Research Hospital because of Walt. Thank you for giving me a chance to work in your laboratory, Walt. Thank you for giving me so many opportunities to explore various topics in psychoacoustics and for giving a great exposure to so many hearing researchers even though I was only an undergraduate student. I really appreciate your advice, which I have always

counted on. I would also like to thank Steve Neely. Thank you, Steve, for all the supportive discussions that encouraged me to applying to a technical university without any engineering background and for discussing and sharing with me your passion for auditory modelling.

I would like to thank Mike Heinz and all the members of his fantastic lab at Purdue. Thank you for hosting me and giving me the opportunity for exposure to AN recordings and physiological experiments. Thank you, Satya, Vijay, Dave and Mike for all your help in and outside the lab at Purdue.

I would like to thank Julie A. Bierer and all the members of her lab at UW-Seattle. Julie, thank you for hosting me, for our discussions and your advice. Thank you Steve, Julie and Elle for a great time discussing our modelling work, introducing me to great cafes around the campus and for the hikes and outings. I really appreciate your time.

I also want to thank an extraordinary group of colleagues and friends at the hearing systems group. I would particularly like to thank Caroline van Oosterhout. Caroline, your help in each and every matter has been so crucial! All my office mates in room 111 over past three years - Christoph, Thomas, Francois, Gerard, Golbarg, Helia, Søren, Antoine and Richard - many thanks for making it so much fun to be in the office every single day!

A very big thank you to my parents, my little sister and my grandparents for all the inspiration and encouragement. Thank you for believing in me and supporting me in every turn I took in my life, nothing is possible without you guys. Finally, Thank you, Søren Rytter for your friendship for the last three of years, I could not have been as happy without it.

Related publications

Journal papers

- Joshi, S. N., Wróblewski, M., Schmid, K.K., Jesteadt, W. (**2016**). "Effects of relative and absolute frequency in the spectral weighting of loudness." J. Acoust. Soc. Am. 139, 373-383. doi: 10.1121/1.4939893

- Joshi, S. N., Dau, T., Epp. B. (**2017**). "A model of electrically stimulated auditory-nerve fiber responses with peripheral and central sites of spike generation." J. Assoc. Res. Otolaryngol. 001–020. doi: 10.1007/s10162-016-0608-2

- Joshi, S. N., Dau, T., Epp. B. (**under review**). "Temporal coding in the electrically stimulated auditory nerve: The effect of uncertainty in the site of spike generation." J. Assoc. Res. Otolaryngol.

- Joshi, S. N., Dau, T., Epp. B. (**in prep**). "Cross-correlation model of interaural time difference coding in listeners with bilateral cochlear implants." J. Assoc. Res. Otolaryngol.

Conference papers

- Joshi, S. N., Dau, T., Epp. B. (**2014**). "Modelling auditory nerve responses to electrical stimulation." Proceedings of Forum Acusticum, *Krakow, Poland*

- Joshi, S. N., Dau, T., Epp. B. (**2015**). "A model of auditory nerve responses to electrical stimulation with single pulse." 38th Annual MidWinter Meeting of

the Association for Research in Otolaryngology, *Baltimore, MD, USA*

- Joshi, S. N., Dau, T., Epp. B. (**2015**). "A model of auditory nerve responses to electrical stimulation with pulse trains." Conference on Implantable Auditory Prostheses, *Lake Tahoe, CA, USA*

- Joshi, S. N., Dau, T., Epp. B. (**2016**). "What affects envelope coding in the electrically stimulated auditory nerve?" 39th Annual MidWinter Meeting of the Association for Research in Otolaryngology, *San Diego, CA, USA*

- Joshi, S. N., Dau, T., Epp. B. (**2016**). "Perspectives on temporal coding in CI listeners using a computational model of electrically stimulated auditory nerve." Die 42. Deutsche Jahrestagung für Akustik - DAGA 2016, *Aachen, Germany*

- Joshi, S. N., Dau, T., Epp. B. (**2016**). "Factors affecting temporal coding in electrically stimulated auditory nerve." 10th FENS Forum of Neuroscience 2016, *Copenhagen, Denmark*

- Joshi, S. N., Dau, T., Epp. B. (**2016**). "Modelling the effect of pulse-rate on coding of interaural time differences in listeners with cochlear implants." 5th Joint Meeting of the Acoustical Society of America and Acoustical Society of Japan, *Honolulu, HI, USA*

- Joshi, S. N., Dau, T., Epp. B. (**2017**). "Effect of Uncertainty in Site of Spike Generation on Envelope Coding in the Electrically Stimulated Auditory Nerve." 40th Annual MidWinter Meeting of the Association for Research in Otolaryngology, *Baltimore, MD, USA*

- Joshi, S. N., Dau, T., Epp. B. (**2017**). "Investigating the sensitivity to interaural time differences in bilateral CI listeners using a computational model of the electrically stimulated auditory nerve." Die 43. Deutsche Jahrestagung für Akustik - DAGA 2017, *Kiel, Germany*

Contents

1

Introduction

The auditory system performs extraordinary computations which convert sound pressure variations into a sensation of hearing. Acoustic stimulation leads to vibrations of the tympanic membrane in the outer ear, excites the ossicular chain in the middle ear that, in turn, produces vibrations of the basilar membrane in the fluid-filled inner ear, the cochlea. It is in the cochlea where the vibrations are converted into neural signals. The inner hair cells (IHC) are highly specialized cells in the cochlea which transduce the vibrations of the basilar membrane into a gradient of IHC receptor potentials. The IHC receptor potentials generate neurotransmitter releases in the synaptic cleft between the IHC and the auditory nerve (AN). The neurotransmitter then triggers action potentials, also referred to as spikes, at the level of the AN. The information related to the acoustic input is represented in terms of spike trains at all levels along the auditory pathway beyond the AN.

Any damage to the processes involved in the transmission of the vibration or the transduction processes causes a hearing loss (Moore, 2007). A conductive hearing loss is caused due to damage to the outer or the middle ear and refers to a dysfunctional transmission of the sound in the auditory periphery. A sensorineural hearing loss refers to dysfunctional transduction processes in the cochlea. The treatment strategies for conductive hearing losses are mainly corrective surgeries (e.g. transplantation of the tympanic membrane). The treatment strategies for sensorineural hearing losses have mainly focused on either the compensation for cochlear damage (e.g. through level-dependent amplification with hearing aids) or the substitution of the cochlear transduction process (implantation of a cochlear implant, CI). Especially in cases of severe hearing loss, amplification of the input signal via hearing aids often does not provide major benefits. In such cases, the listeners are commonly prescribed with CIs.

CIs are biomedical implants that bypass the cochlear transduction process and directly stimulate the AN fibers (ANF) with electrical current. CI technology has developed substantially over the past few decades and has attracted much attention for treatment of severe hearing loss or deafness. In the first systematic study conducted to evaluate the performance of CI listeners, results from 13 existing CI listeners were reported (Bilger et al., 1977). The number of CI listeners was estimated to be 12,000 in 1995, 120,000 in 2008 and 326,000 in 2012 (National Institute of Health, 2016), demonstrating the clinical success of CIs as a treatment of deaf or severely hearing-impaired listeners. CI listeners can achieve significant speech understanding in quiet. However, they still face great difficulties in tasks related to speech understanding in background noise, sound source localization and music perception. Furthermore, CI listeners often complain about the poor quality of the sound processed through their CIs.

CI signal processing strategies attempt to mimic mainly two key aspects of signal processing of the acoustic input in the normally functioning cochlea: (1) the frequency selective coding of sound in separate frequency channels, manifested as the frequency-place mapping in the cochlea and (2) the processing of the temporal envelope of the acoustic input in each frequency channel. Generally, the acoustic input received through the microphone of the CI is processed through an array of bandpass filters. Each of these bandpass filters corresponds to a separate electrode in the multiple-electrode array in the cochlea and aims to stimulate a selective population of AN fibers (ANFs) along the cochlea. The CI signal processors extract the slowly varying envelope of the signal at the output of each of the bandpass filters while the fast-varying fine structure (related to the carrier) is ignored. The envelope serves then as a modulator for a train of current-balanced, biphasic pulses. The amplitude modulated pulse trains are delivered to the corresponding electrode that injects the current into the cochlea. The perception of the CI listeners is therefore primarily based on the information carried in the envelope of the acoustic signals.

The ANFs form a bottleneck of information transfer between the CI and the central processing of the auditory system. A good understanding of the factors that affect the envelope coding in the electrically stimulated ANFs thus seems crucial. The aim of the current thesis was to analyze the effect of stimulus parameters, such

as carrier pulse rate and level, as well as the role of neural factors on temporal aspects of hearing. In this work, a phenomenological model of the electrically stimulated ANFs was developed and evaluated using physiological and behavioral data.

One factor that affects the temporal response statistics of the electrically stimulated ANFs is the spatial distribution of the site of spike generation along individual ANFs (van den Honert and Stypulkowski, 1984; Shepherd and Javel, 1999; Miller et al., 1999). In acoustic hearing, the ANFs are stimulated via synaptic input and a spike is considered to be generated in the ANF close to the synapse (Rutherford et al., 2012) at the distal end of the peripheral axon. The spike then propagates along the ANF reaching the cochlear nucleus in the brainstem. In contrast to the synaptic input, electrical current is injected via the CI electrodes into the cochlear fluid. This artificial current injection produces a large voltage gradient throughout the cochlea. These changes in the extracellular voltage trigger the voltage-gated ion channels in the ANF membrane. Depending on the magnitude of the extracellular voltage at various locations along the ANF, a spike can be generated at several locations along the ANF. The distance between the location at which the spike is generated and its destination in the cochlear nucleus affects the spike times and, thus, the temporal response statistics of the electrically stimulated ANF responses.

Another factor that influences the site of the spike generation besides the voltage magnitude is the polarity of the applied current. It has been assumed that a cathodic current depolarizes the AN membrane and generates a spike, while an anodic current hyperpolarizes the AN membrane. The stimulation with a monophasic current is, however, not applicable because it can lead to electrolytic reactions in the cochlea which would damage the neural membrane. To avoid any damage to the ANFs and the cochlea, all commercially available CIs use charge-balanced and mainly symmetric biphasic pulses for stimulation. It is assumed that the charge-balancing anodic current neutralizes the injected charge and therefore avoids electrolytic reactions. However, physiological recordings of the ANFs show that both anodic and cathodic charges can excite the ANFs and generate a spike (Miller et al., 1999; Shepherd and Javel, 1999). Moreover, spike latencies in response to a cathodic current are longer than those in response to an anodic current. This dif-

ference in spike latencies has been interpreted to reflect the different sites of spike generation. Detailed finite-element modeling of voltage distribution in the cochlea, together with a biophysical cable model of the ANFs, has shown that the cathodic phase depolarizes the peripheral axons of the ANF (closer to the IHCs, farther from the cochlear nucleus) and generates a spike while the anodic phase depolarizes the central axon of the ANF (farther from the IHCs, closer to the cochlear nucleus) and generates a spike. Hence, the simplified assumption of depolarizing anodic and hyperpolarizing cathodic charges does not apply to extracellular stimulation with CIs in the cochlea. This leads, in contrast to acoustic hearing, to an uncertainty in terms of the site of spike generation. This effect has not only been reported for monophasic stimulation but also for stimulation with symmetric biphasic pulses (van den Honert and Stypulkowski, 1984; Shepherd and Javel, 1999; Miller et al., 1999).

Even though reported in the literature, this aspect of electrical stimulation of ANF has not yet been considered in the analysis of existing nor in the design of novel stimulation strategies. In the current thesis, the effect of uncertainty in the site of spike generation on the envelope coding in the CI listeners was investigated in detail. The structure of the thesis is as follows:

Chapter 2 provides an analysis of two existing models of the ANF responses to electrical stimulation. The leaky integrate-and-fire type model of Horne et al., 2016 and the point process model of Goldwyn et al., 2012 are evaluated and compared to physiological data recorded from cat ANFs using stimulation with monophasic and biphasic pulses of different pulse shapes.

Chapter 3 describes a novel computational model of the ANF responses for electrical stimulation. The model is based on two assumptions that (1) a spike can be generated at different positions along the ANFs and that (2) both the anodic and cathodic currents can generate a spike. The model contains two exponential integrate-and-fire type neurons that represent the peripheral and the central axon of the ANF whereby the peripheral axon is assumed to be excited by the cathodic phase and inhibited by the anodic phase and the central axon is assumed to be excited by the anodic phase and inhibited by the cathodic phase. The model also includes subthreshold and suprathreshold feedback loops that modify the

spiking ability dependent on the stimulus history. The model is evaluated for stimuli consisting of single pulses with various pulse shapes and pulse trains with different pulse rate and stimulus levels. The model predictions are compared to physiological data from several studies.

Chapter 4 describes an analysis of stimulus related and neural factors that affect the representation of the stimulus envelope in the simulated ANF responses. Specifically, the effect of the carrier pulse rate on modulation detection thresholds is considered. The model from chapter 3 is used to simulate the ANF responses for various stimuli previously used in the literature and a neurometric analysis is used to predict modulation detection thresholds based on the simulated ANF responses. Model predictions are compared to behaviorally measured modulation detection thresholds in CI listeners. Furthermore, the effect of uncertainty in the site of spike generation on the modulation detection threshold is analyzed.

Chapter 5 describes a cross-correlation based model of interaural time difference (ITD) coding in listeners with CIs in both ears (bilateral CI). The ANF model from chapter 3 is used to simulate the monaural responses and to estimate the time delay between the two ears. The model is used to test the effects of various stimulus parameters, such as the pulse rate, envelope shape, modulation frequency and inter-pulse interval jitter, on stimulus envelope ITD coding. Model predictions are compared to behaviorally measured ITDs in bilateral CI listeners from various studies.

Finally, *Chapter 6* provides an overview of the results and perspectives in terms of the factors that influence the temporal coding of the input stimulus in the electrically stimulated ANFs. It is discussed how the findings described in this thesis may influence the development of new CI stimulation strategies.

In the *appendix*, a separate research project undertaken to study loudness perception of broadband sounds in normal hearing listeners is described. The established models of loudness perception assume that the overall loudness of broadband sounds is a linear sum of 'specific' loudness of its narrowband frequency components. In this project, the assumption regarding equal contribution of the frequency bands to the overall loudness of broadband sound was tested. Molecular psychophysical techniques along with loudness matching procedures was

used to study the 'spectral weighting' of loudness. The data is compared with the predictions of the established model of loudness perception.

2

Evaluation of state-of-the-art models of the electrically stimulated auditory nerve fiber responses[a]

Abstract

Cochlear implants (CI) directly stimulate the auditory nerve fibers (ANFs), bypassing the mechano-electrical transduction in the inner ear. Trains of biphasic, charge-balanced pulses (anodic and cathodic) are used as stimuli to avoid damage of the tissue. Physiological data shows that the pulses of either polarity are capable of generating spikes whereby the sites of initiation of the AP differ for the two polarities. A cathodic pulse generates a spike in the peripheral axon, whereas an anodic pulse generates a spike in the central axon. The latency difference between the spikes generated at the different sites is about 200 μs, which is large enough to affect the temporal coding of sounds and hence, potentially, the communication abilities of the CI listener. In the present study, two recently proposed models of electric stimulation of the ANFs (Hamacher, 2004; Goldwyn et al., 2012; Horne et al., 2016) were considered in terms of their efficacy to predict the response statistics for stimulation with both anodic and cathodic stimulation of the ANF of cat (Miller et al., 1999). The models' responses to the electrical pulses of various shapes (Shepherd and Javel, 1999) were also

analyzed. It was found that, while these models can account for the firing efficiency in response to symmetric biphasic pulses, they fail to correctly describe the timing of AP in response to monophasic pulses. Moreover, they also fail to correctly predict the effect of alternative pulse shapes such as the pseudomonophasic pulses or biphasic pulses with an inter-phase gap on the observed response statistics. Strategies for improving the model performance with respect to spike-response statistics are discussed. Eventually, a model that is able to account for the effect of different pulse shapes on response statistics in electric hearing will be useful for objective evaluation and improvement of CI stimulation strategies.

2.1   Introduction

Electrical stimulation of the auditory nerve fibers (ANF) with cochlear implants (CI) has attracted much attention from the scientific as well as the clinical communities in the last two decades. A prescription of CI to restore a sense of hearing in patients with severe to complete hearing losses has become a standard procedure. An estimated number of CI recipients worldwide in December 2012 was approximately 324,200 and is growing exponentially. The success of CIs has motivated the development of retinal and other sensory implants to overcome the sensory deficits. However, there is still substantial room for improvement regarding the performance of CI users in terms of speech perception in realistic listening environments, sound localization, pitch and music perception, and satisfaction with respect to sound quality (Moore, 2003; Wilson and Dorman, 2008). Several factors have been identified to affect the performance of listeners with CI. These include 'personal' factors like the duration of a person's deafness, the age at the time of implantation, the time period since implantation and cognitive abilities of the listener. The performance is also affected by 'physical' factors, such as CI electrode placement, electrode type, stimulation types, and status of the ANF. Various signal-processing and stimulation strategies have been proposed with the aim to

improve CI listener's speech perception. Behavioral evaluations of these, however, showed only modest improvements (e.g. Geurts and Wouters, 1999; Vandali et al., 2000). An ideal prostheses should provide CI listeners the cues that are available to a normally hearing listener. It is, therefore, important to have control over the neural responses evoked by electrical stimulation.

Some research has focused on developing objective physiological measures (e.g. eCAP, eABR) to evaluate the behavioral performance of CI listeners and corresponding ANF responses to electrical stimulation. While these measures provide a window into a neural representation of the signal, they are difficult to interpret in connection with the behavioral performance of the listener. Behavioral tests provide an ultimate evaluation of the CI stimulation strategies; however, they do not allow an analysis of how the stimulus and physiological factors affect the listener's performance. Physiologically inspired models may provide a useful framework to describe the data. Single-neuron recordings showed that the responses of the electrically stimulated ANFs are fundamentally different from responses of the normally functioning acoustically stimulated ANFs Javel, 1990. The reasons for these differences are unclear. Models of ANF responses developed for electrical stimulation, particularly those concerning single-neuron recordings, would be valuable to examine the processes involved in 'electric hearing' (Goldwyn et al., 2012; Fredelake and Hohmann, 2012; Finley et al., 1990).

Existing models describing responses of the electrically stimulated ANFs can be divided into two broad classes: biophysical and empirical models. Biophysical models are based on the Hodgkin-Huxley formalism and describe independent ion channels and their biophysical mechanisms that lead to the generation and propagation of action potentials. However, the large number of parameters used in these model make it difficult to describe the essential processes that cause a specific response, or to use them for predicting responses of a large population of neurons as well as behavioral data. In contrast, the empirical models are phenomenological in nature and aim at describing the main statistical properties of the neural responses. These models are based on the concept of the 'integrate-and-fire' neuron, defined by variations of a low-pass filter and a criterion threshold at which the model generates a spike by discharge. These "effective" models with
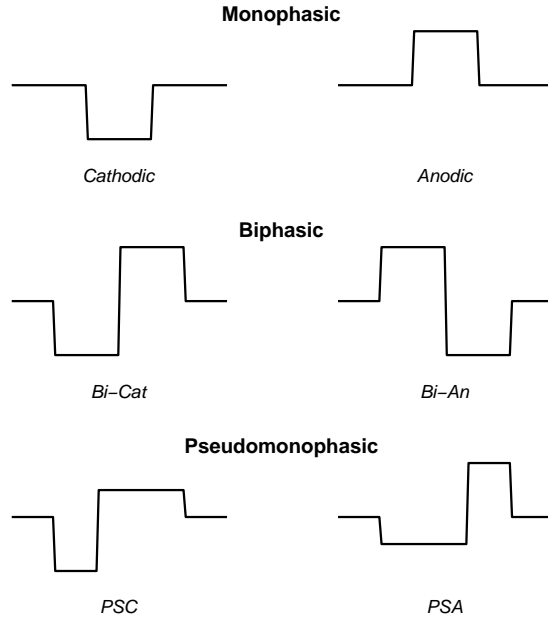
Figure 2.1: Various pulse shapes that have been used to investigate responsiveness EsAN response of single-neurons. (1) Monophasic pulses of either polarity (2) Symmetric biphasic pulss and (3) Assymetric biphasic pulses created to approximate the monophasic pulses and to balance the charge.

fewer parameters are easier to fit to the existing physiological data but do not provide precise information about each individual ion channel contributions to the responses. Ease of mathematical analysis, however, make the effective models suitable for predicting neural population response. Therefore, this study focuses on empirical models of electrically stimulated ANF responses.

A successful physiologically-inspired model of the ANF responses to electrical stimulation should be able to account for the physiological data in response to a variety of stimuli. However, existing models seem to be quite limited in their ability to describe such variety of data (Hamacher, 2004; Goldwyn et al., 2012; Horne et al., 2016; Fredelake and Hohmann, 2012; Bruce et al., 1999b; Bruce et al., 1999a; Bruce et al., 1999c; Campbell et al., 2012; Macherey et al., 2007; van Wieringen

et al., 2008). These models are fitted to electrically stimulated ANF responses to symmetric biphasic pulses since, symmetric biphasic pulses are used in the clinical application in order to balance the artificially introduced charge to protect the nervous tissue. However, both cathodic (negative) and anodic (positive) pulses can separately evoke a neural response (Miller et al., 1999). A clear understanding of the electrically stimulated ANF response behavior to various pulse shapes is required for an analysis of ANF responses to direct electrical stimulation (Shepherd and Javel, 1999). Since symmetric or asymmetric biphasic pulses are current steps with monophasic anodic and cathodic phase, a model should be able to reproduce the responses of the electrically stimulated ANFs to both monophasic as well as biphasic pulses.

2.2 Methods

This study presents an analysis of the model responses to electrical pulses shown in Fig. 2.1. Two models that were considered here are (i) the point-process model (Goldwyn et al., 2012) and (ii) a leaky integrate-and-fire neuron with stochastic delays (Horne et al., 2016). The evaluation was based on firing efficiency (FE) curves and spike time statistics. FE curves provide the probability of a spike as a function of the pulse level and have been used to describe a dynamic range of the neuron and its threshold (i.e. the level that evokes a spike with a probability of 0.5). The spike time statistics are described using the first-spike latency relative to stimulus onset. The temporal response pattern is strongly influenced by spike latencies, therefore, advanced models should also correctly represent them. The two models chosen here were considered because they describe both instantaneous firing rates and spike latencies, while most other models are focused on only the firing rates (Bruce et al., 1999b; Bruce et al., 1999a; Bruce et al., 1999c; Campbell et al., 2012; Macherey et al., 2007). Each model was stimulated 1000 times per stimulus level and for each pulse shape to achieve a reliable estimate of the neural responses. Each model was run with a sampling rate of 10^6 Hz.

2.3   Results

2.3.1   Monophasic pulses

The intra-cochlear electrodes create a complex spatial spread of electric charge that depends on the physical configuration of the electrode, the stimulation mode used and the polarity of the stimulating pulses (Finley et al., 1990; Frijns et al., 1996). It has been shown that both phases of biphasic pulses are capable of exciting the ANFs and generating spikes (Miller et al., 1999; Shepherd and Javel, 1999). Therefore, biphasic stimulation produces a complex neural response due to different neurons responding to either the anodic and cathodic phase.
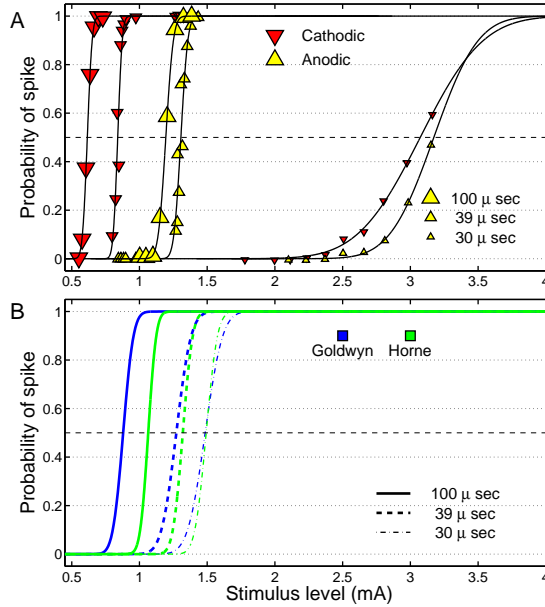
Figure 2.2: FE curves three pulse durations are shown in panel A for anodic (upward pointing triangle) and cathodic (downward pointing triangle). The simulation results from the Goldwyn model (blue) and Horne model (green) are shown in panel B. Neuron's threshold is defined as the level that elicits a spike probability of 0.5 on the FE curve (dotted line). Both models are only responsive to cathodic pulses and show smaller reduction in threshold with increasing duration than observed in the data.

Miller et al., 1999 investigated single-neuron responses to monophasic stimulation with both anodic and cathodic pulses with durations of 26 and 39 μs. They reported significant differences between the two phases with respect to stimulation threshold, spike times and jitter, but not regarding the dynamic range. Fig. 2.2 (A) shows the FE curves for the anodic and cathodic monophasic pulses for three durations. The data for 100 and 30 μs were taken from Shepherd and Javel, 1999 and the data for 39 μs were taken from Miller et al., 1999. As can be seen, the neuron's threshold for stimulation with cathodic pulse (red symbols) is lower than for stimulation with anodic pulses (yellow symbols). Notably, the differences in threshold for cathodic vs anodic pulses as a function of duration are different, highlighting differences in the value of signal integration properties of the nerve membranes.

Model simulations for the three pulse durations are shown in Fig. 2.2 (B). The GLM model of Goldwyn et al., 2012 (in blue) and leaky integrate-and-fire model (in green) of Horne et al., 2016 are only responsive to cathodic phase stimulation and process anodic phase stimulation as inhibitory input. Both models show smaller differences in thresholds with increasing phase duration than observed in the data. The change in thresholds for phase duration is related to the value of membrane chronaxie, that characterizes the integration properties of the membrane. Simulations show that the chronaxie values used in both the models are not suitable for monophasic stimulation and, therefore, may not be generalized.

The average spike latency (red and yellow symbols) and jitter (gray area) as a function of the dynamic range are shown in Fig. 2.3 (A). The data were digitized from Miller et al., 1999. Based on the spike latency differences between anodic (yellow) and cathodic (red) pulses, it is assumed that the two phases excite separate segments of the AN. Spike latencies suggest that the cathodic phase excites the peripheral axon, while the anodic phase excites the central axon (Miller et al., 1999). The spike latency decreases with increasing FE or the percent dynamic range. It is assumed that the place of spike initiation moves central for both phases of excitation.
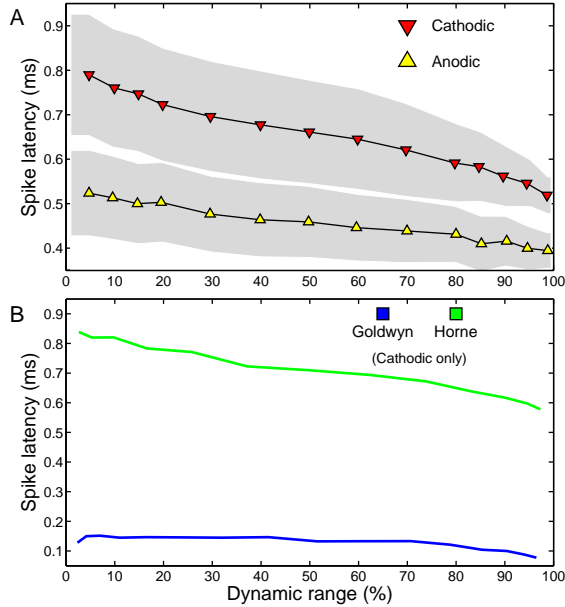
Figure 2.3: Panel A shows spike latency as a function of percent dynamic range for the neuron for anodic and cathodic pulses of 39 μs taken from Miller et al., 1999. Jitter level is also indicated as the standard deviations with gray area. Data shows average difference is greater than 200 μs between the spike initiated by the anodic and the cathodic phase. Panel B shows corresponding model simulation for cathodic pulses. The two models predict reduction in spike latency with increasing level.

2.3.2   Symmetric biphasic pulses

The influence of differences in thresholds and spike latencies between the anodic and cathodic pules is also seen in response to biphasic pulses. These differences are particularly pronounced when comparing the AN responses to symmetric biphasic pulses in which the first phase is either anodic (Bi-An) or cathodic (Bi-Cat). The Bi-An pulses produce two types of responses. As seen in Fig. 2.4 (A), few neurons show large differences in between the thresholds for Bi-An vs Bi-Cat stimuli. However, the spike latencies of Bi-An pulses are much longer (Fig. 2.5 (A)), in comparison to Bi-Cat pulses. These neurons become less responsive with increasing stimulus level (black filled upward pointing triangles in Fig. 2.4). These neurons have been
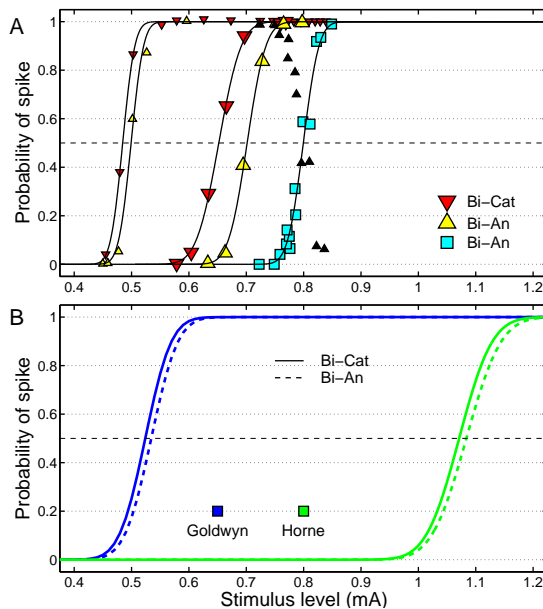
Figure 2.4: In Panel (A), FE curves for both types of symmetric biphasic pulses (cathodic first and anodic first) with phase duration of 100 μs each obtained from a same neuron. Smaller symbols are obtained from Fig. 15 of Miller et al., 1999 and large symbols from Fig. 5 of Shepherd and Javel, 1999. Some neurons show a small elevation in threshold for Bi-An pulse relative to Bi-Cat pulse (yellow triangle). These neurons become less responsive with increasing level (black triangles). Other neurons show a larger elevation in threshold relative to Bi-Cat responses (cyan squares). Panel B shows corresponding model simulations

hypothesized to be responding to the cathodic phase of the stimuli. The remaining neurons show higher thresholds for the anodic first stimuli (cyan filled squares). These neurons have been hypothesized to be responding to anodic phase. These hypotheses are supported by distinct spike latencies (Fig. 2.4) (B), consistent the differences in spike latencies produced by anodic and cathodic phases.

Consistent with the data, both models show a lower threshold for the Bi-Cat pulse than for the Bi-An pulse (Fig. 2.6 (A)). The predicted spike latencies follow the data with longer latencies for Bi-An pulses and with shorter latencies for Bi-Cat pulses, although the absolute latencies deviate from the data.
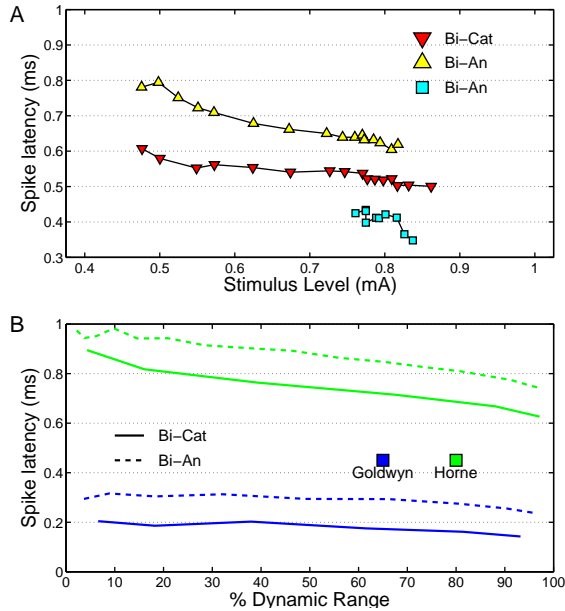
Figure 2.5: In Panel (A), spike latencies as function of stimulus level for Bi-An and Bi-Cat pulses obtained from Miller et al., 1999 is shown. Latency for first type of Bi-An pulses (yellow upward pointing triangles) are longer than for Bi-Cat pulses. Second type response to Bi-An pulses (cyan squares) show shorter latencies than Bi-Cat pulses. Panel (B) shows corresponding model simulations. Due to a large differences in the thresholds between the models, latencies have been presented as function of the dynamic range.

2.3.3   Asymmetric biphasic pulses

While biphasic pulses are a necessity in order to preserve the nervous tissue, various assymetric biphasic pulses have been shown to be better options for stimulating the AN (van Wieringen et al., 2008). An example for an asymmetric pulse shape with pseudomonophasic cathodic (PSC) and anodic (PSA) pulses is shown in Fig. 2.1. Fig. 2.6 shows differences in the FE curves produced by these pulse shapes Shepherd and Javel, 1999).

Both models are responsive to PSC and PSA pulses consistent with the data. Models also predict lower thresholds for PSC than for PSA consistent with the data. However, these differences are significantly larger than observed in the data. Since
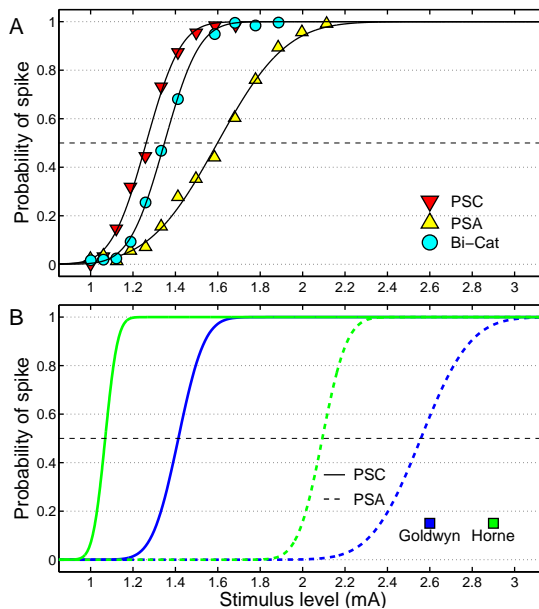
Figure 2.6: FE curves for asymmetric biphasic pulses and cathodic-first symmetric biphasic pulse for a single neuron. Data is obtained from Fig. 11 of Shepherd and Javel, 1999.

the spike latency information for these stimuli is not reported, predictions of the model could not be confirmed in terms of accuracy. The data show a larger dynamic range for the PSA in comparison to the PSC or Bi-Cat pulse. This difference in the dynamic range of a neuron is only accounted for by the Goldwyn model.

2.4   Conclusion

Single neuron recordings have shown that the AN membrane acts as an independent leaky integrator of charge from pulses of both polarities Shepherd and Javel, 1999, and these differences are reflected in differences in the thresholds-duration relationship between anodic and cathodic pulses. AN responses to monophasic stimulation provide the basis for studying EsAN responses. It was shown here that state-of-the-art models of EsAN responses fail to account for the correct threshold-

duration relationship and spike latencies observed for various pulse shapes. These models were developed for and fitted to EsAN responses to symmetric biphasic pulses. It is proposed here that models that correctly predict the responses to monophasic pulses will also be able to account for EsAN responses to various pulse shapes. Such a model, however, needs to include two independent leaky integrators to capture the response to anodic and cathodic stimuli, respectively, and to produce a spike to stimulation with either phase. Such a model may be generalized to assess EsAN responses to various pulse shapes.

Acknowledgements

3

A model of electrically stimulated auditory-nerve fiber responses with peripheral and central sites of spike generation [b]

Abstract

A computational model of cat auditory nerve fiber (ANF) responses to electrical stimulation is presented. The model assumes that (1) there exist at least two sites of spike generation along the ANF and (2) both an anodic (positive) and a cathodic (negative) charge in isolation can evoke a spike. A single ANF is modeled as a network of two exponential integrate-and-fire point-neuron models, referred to as peripheral and central axons of the ANF. The peripheral axon is excited by the cathodic charge, inhibited by the anodic charge, and exhibits longer spike latencies than the central axon; the central axon is excited by the anodic charge, inhibited by the cathodic charge, and exhibits shorter spike latencies than the peripheral axon. The model also includes subthreshold and suprathreshold adaptive feedback loops which continuously modify the membrane potential and can account for effects of facilitation, accommodation, refractoriness, and spike-rate adaptation in

ANF. Although the model is parameterized using data for either single or paired pulse stimulation with monophasic rectangular pulses, it correctly predicts effects of various stimulus pulse shapes, stimulation pulse rates, and level on the neural response statistics. The model may serve as a framework to explore the effects of different stimulus parameters on psychophysical performance measured in cochlear implant listeners.

3.1   Introduction

Patients with severe hearing loss or deafness are commonly prescribed with cochlear implants (CIs). The CIs bypass the impaired mechano-electrical transduction pathway through the cochlea and directly stimulate the auditory nerve fibers (ANFs) with electric pulses. CI signal processing strategies aim to mimic the cochlear processing of the acoustic inputs and to provide the CI listeners the "essential" cues for successful communication. Primarily, they extract the slowly-varying envelopes of the acoustic signals and stimulate the ANFs with a train of biphasic pulses modulated with the processed envelope (Wilson et al., 1991). Although most CI listeners can achieve some speech intelligibility in quiet with this strategy, they also face great difficulties in understanding speech in background noise and in other psychophysical tasks related to pitch and melody perception as well as sound localization (Wilson and Dorman, 2008). Deficits in the temporal coding in the electrically stimulated ANFs may contribute to these perceptual difficulties of the CI listeners. Despite significant efforts in the development of signal processing strategies for better and efficient processing of the acoustical cues, the improvements in the performance of the CI listeners have been minimal and have been related mainly to the more advanced cue extraction strategies (e.g. advanced combination encoding, continuous interleaved sampling

etc.) or to the current steering strategies that reduce the current spread in the cochlea (Bierer, 2010). For the CI stimulation strategies to be further beneficial for the listener, the ANFs must be able to encode the envelope cues delivered by the CI. A better understanding of the stimulus-response relationship of the ANF for electrical stimulation seems thus crucial for the development of novel and more efficient stimulation strategies. Quantitative models, particularly those concerned with temporal aspects of ANF responses, can be a useful tool to characterize such a relationship.

Clinical CI devices use trains of symmetric biphasic pulses for the stimulation due to safety regulations. Contemporary accounts of electrical stimulation of the ANFs assume that the cathodic phase of the biphasic pulse depolarizes the neural membrane and generates a spike while the anodic phase hyperpolarizes the membrane and to balance the charge in the cochlea. However, results from several studies have shown that an anodic pulse can also generate a spike and that responses measured at the nerve trunk to the anodic phase exhibit shorter spike latencies than responses to the cathodic phase (van den Honert and Stypulkowski, 1984; Miller et al., 1999; Shepherd and Javel, 1999). Biophysical models that consider detailed cochlear morphology and its effects on the charge conduction through the cochlea show that an anodic pulse depolarizes the neuron at a location more distant from the stimulating electrode than a cathodic pulse (Rattay et al., 2001; Rubinstein, 1991). These models suggest that the site of spike generation along the ANF differs for anodic and cathodic pulses. The site of spike generation is crucial since it determines the delay with which the spike arrives at further processing stages along the auditory pathway. A difference in spike latencies between the responses to anodic and the responses to cathodic pulses amounts to approximately 200 μs in cats (Miller et al. 1999) and possibly up to 400 μs in humans (Rattay et al., 2001; Rattay et al., 2013; Undurraga et al., 2013). The effect of multiple sites of spike generation on spike latencies has also been reported in ANF responses to stimulation with biphasic pulses (van den Honert and Stypulkowski, 1984; Miller et al., 1999; Shepherd and Javel, 1999). Since robust temporal coding depends on the precision in spike timing, the site of spike initiation as well as the uncertainty related to it can affect the temporal coding in ANFs

stimulated with symmetric biphasic pulses.

The ANF responses in cats show significantly lower thresholds for cathodic pulses than for anodic pulses (Miller et al., 1999). Therefore, it has been assumed that only the cathodic phase of a biphasic pulse will generate a spike. Based on this assumption, state-of-the-art quantitative models of ANF responses have mainly focused on the responsiveness of the ANF to the depolarizing cathodic phase (Bruce et al., 1999b; Hamacher, 2004; Fredelake and Hohmann, 2012; Goldwyn et al., 2012) or on inhibitory properties of the hyperpolarizing anodic phase (Rubinstein et al., 2001; Horne et al., 2016). Any charge-balanced pulse can be decomposed into anodic and cathodic charges, and responses to various pulse shapes are a consequence of the sensitivity to the single pulse phases and the interaction between these. Hence, the comparison of responses to different pulse shapes can be used to explore the response behavior of the electrically stimulated ANF. Current state-of-the-art models cannot account for the response statistics observed in the available data obtained with various pulse shapes (Joshi et al., 2014) and the effects of stimulation rate and level of pulse-train stimuli on the ANF responses (Bruce et al., 1999a; Goldwyn et al., 2012; Goldwyn et al., 2010a). Hence, these models cannot easily be generalized to assess the limitations and possible benefits of different CI stimulation strategies that aim to convey temporal information.

The current study presents and evaluates a computational model of the ANF responses for electrical stimulation. The model is based on the idea that there exist at least *two sites of excitation* along the ANF and that these sites differ in their *sensitivity to either cathodic or anodic charges*. Each unmyelinated node along the ANF can generate spikes in response to an extracellular voltage. However, for simplicity, the model presented in the current study considers all nodes along the peripheral axon of the ANF as one site of excitation and those along the central axon of the ANF as the other site of excitation. The model incorporates dynamic feedback loops that enable the prediction of response properties such as facilitation, accommodation, refractoriness, and spike-frequency adaptation. The model is evaluated for stimulus conditions with various pulse shapes and pulse trains of different stimulation rate and level.

3.2   The model

3.2.1   Structure

The structure of the model is shown in Fig. 3.1. A single ANF is divided into two parts: a peripheral axon and a central axon. The peripheral and central axons are each described by an exponential integrate-and-fire point neurons with two adaptive currents (Fourcaud-Trocmé et al., 2003; Brette and Gerstner, 2005), which have been shown to account for the spike-time statistics of ANF responses (Rutherford et al., 2012). Based on the data from electrically stimulated ANFs in cats, the peripheral axon is assumed to be excited by cathodic charge and inhibited by anodic charge. The central axon is assumed to be excited by anodic charge and inhibited by cathodic charge (Miller et al., 1999). Subthreshold and suprathreshold adaptation are included in the model via two feedback loops which continuously modify the membrane potential. Finally, each axon includes an independent noise source to model the stochastic response properties of the ANF. In the model presented here, the two axons of the ANF are modeled in parallel. The differences in spike times between the spikes generated at the peripheral and the central axons are achieved by the differences in membrane characteristics.

Both model axons simultaneously calculate the membrane voltage V in response to the stimulus input I_{Stim}:

$$C\frac{dV}{dt} = h(V) - I_{sub} - I_{supra} + I_{Noise} + I_{Stim} \qquad (3.1)$$

where C is the membrane capacitance, I_{sub} and I_{supra} are subthreshold and suprathreshold adaptation currents, and I_{Noise} is the noise current. h(V) describes the passive filtering of the stimulus along with the exponential upswing of the membrane potential during spike generation in the axon. The noise term I_{Noise} introduces stochasticity in the membrane voltage and is modeled as a random process with a Gaussian amplitude distribution and a power spectrum proportional to $1/f^{\alpha}$, inspired by the recordings of stochasticity in neural membrane voltages (Verveen and Derksen, 1965). The exponent α determines the spectrum of the noise such that if α equals 0, the noise has a flat spectrum. If α is either 1 or 2, the

Stimulus   **Model of Auditory Nerve Fiber**   **Spike train**
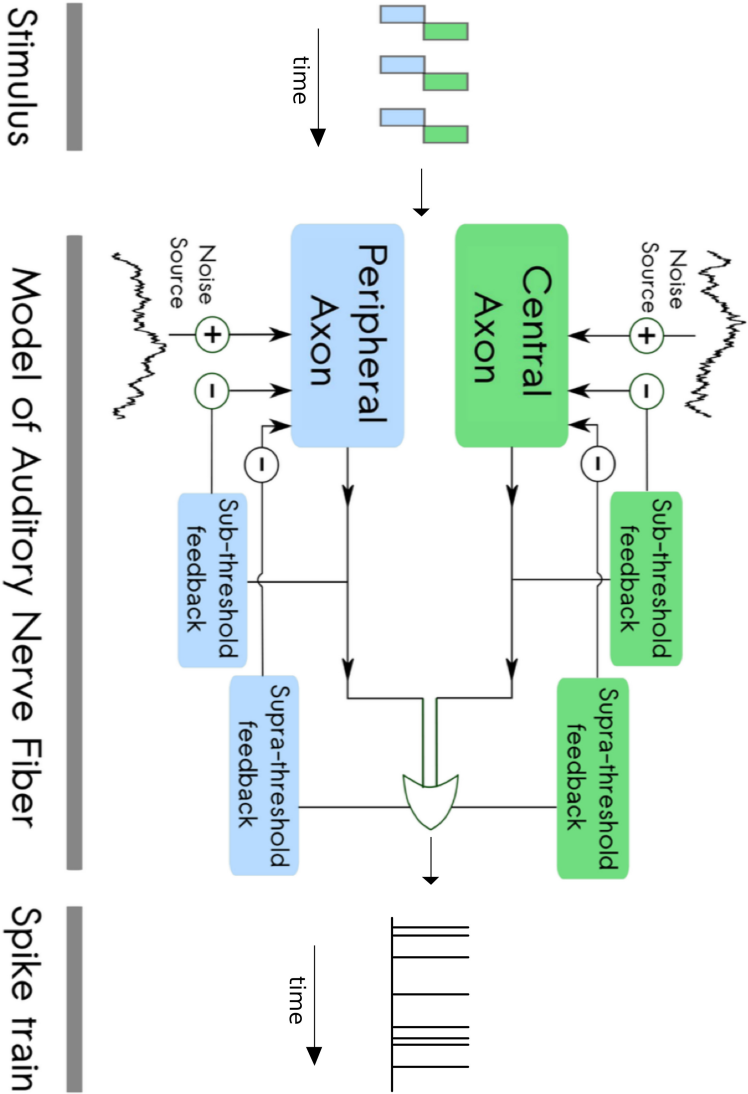
Figure 3.1: Structure of the proposed model. The ANF is modeled by a network of two point neurons describing the peripheral and the central axon of the ANF. The peripheral axon is excited by cathodic (negative) current and inhibited by anodic (positive) current. The central axon is excited by anodic current and inhibited by cathodic current. Independent noise inputs to each of the axons introduce stochasticity in the membrane potential. The two adaptive feedback currents account for subthreshold and suprathreshold adaptive properties observed in the ANF responses. The interaction between the two axons is modeled with an "OR" logic gate, selecting the first spike generated by either of the axons. The model receives the stimulus waveform as an input and provides spike times as the output.

noise spectrum decays with 3 or 6 dB per octave, respectively.

The passive filtering of the stimulus in each axon, is described as the sum of a linear and an exponential function:

$$h(V) = -g_L(V - E_L) + g_L \Delta T e^{\frac{V - v_{threshold}}{\Delta T}} \qquad (3.2)$$

where g_L is the membrane conductance, E_L is the resting membrane potential, ΔT represents the slope factor of the exponential function, and $v_{threshold}$ is the threshold potential. In common integrate-and-fire models, a spike is indicated when the membrane potential crosses $v_{threshold}$. In the exponential integrate-and-fire model described in Eq. 3.2, the membrane potential continues to increase after it has crossed $v_{threshold}$. The increase of the membrane potential beyond $v_{threshold}$ is exponential and the rate of the increase is determined by the slope factor ΔT. To indicate a spike discharge in this model, the peak voltage v_{peak} is introduced which indicates the membrane voltage at which the spike is generated. The corresponding time of spiking is denoted as t_{spike}. After spiking, the membrane potential is reset to v_{reset}.

The subthreshold and suprathreshold adaptive currents I_{sub} and I_{supra} are described in Eqs. 3.3 and 3.4, respectively:

$$\tau_{sub} \frac{dI_{sub}}{dt} = a_{sub}(V - E_L) - I_{sub} \qquad (3.3)$$

$$\tau_{supra} \frac{dI_{supra}}{dt} = a_{supra}(V - E_L) - I_{supra} \qquad (3.4)$$

with the conductances a_{sub} and a_{supra} and the time constants of the subthreshold and suprathreshold adaptation τ_{sub} and τ_{supra}, respectively. The magnitude of the adaptation currents depends on the membrane potential, V.

The interaction between the two axons is modeled with an "OR" logic gate that selects one output from the two inputs. This allows both axons to generate spikes independently, and the OR gate selects the axon that spikes first, indicated by t_{spike}. The spike triggers an adaptation process that modifies the suprathreshold adaptation current, described in Eq. 4, by an offset b, which accounts for the spike-

rate adaptation (Brette and Gerstner, 2005). Irrespective of which axon generated a spike, the spike-triggered adaptation is applied to both the peripheral and the central axon. After a spike has occurred, the neuron is set into an absolute refractory period (ARP; Miller et al., 2001a). During this period, no spike can be generated irrespective of the level of the stimulus current. Unlike traditional integrate-and-fire models that describe the ARP as the *dead time*, here both axons continue to integrate the membrane potential during the ARP. However, during this period, the membrane only receives the input from the subthreshold and suprathreshold adaptation currents as well as the membrane noise, but no input from the stimulus.

The input to the model is a temporal waveform of the stimulus, $I_{Stim}(t)$, in which $I^+(t)$ is the anodic (positive) charge and $I^-(t)$ is the cathodic (negative) charge. To allow the peripheral axon to be excited by a cathodic charge and inhibited by an anodic charge while using the same underlying equation, the input to the peripheral axon is the inverted temporal waveform of $I_{Stim}(t)$. An additional parameter, β, is included in the model to vary the effect of the inhibitory phase on spike generation. When β equals 0, the inhibitory phase of the stimulus is fully removed and when β equals to 1, the inhibitory phase remains unchanged. As a result, the total input stimulus waveform to the two axons can be expressed as:

$$I_{Stim\,(central)}(t) = \beta I^-(t) + I^+(t) \qquad (3.5)$$

$$I_{Stim\,(peripheral)}(t) = -(I^-(t) + \beta I^+(t)) \qquad (3.6)$$

where $I_{Stim\,(peripheral)}$ represents the stimulus input entering the peripheral axon, $I_{Stim\,(central)}$ represents the stimulus input entering the central axon, I^- is the cathodic charge, I^+ is the anodic charge and β is the scaling factor for the inhibitory charge. Both axons simultaneously receive the stimulus input and independently calculate the membrane potential at each location.

3.2.2   Parametrization

The model parameters that determine the charge integration properties of the neural membrane (g_L and C) and stochasticity of neural responses (I_{Noise}) were estimated from ANF responses to electrical stimulation with monophasic anodic and cathodic pulses in cats. The method used to estimate these parameters is described in the following paragraphs. The other remaining parameters of the model were estimated to best represent the responses to monophasic stimulation. The complete set of obtained values is presented in Table 3.1.

The parameters related to the charge integration properties of the integrate-and-fire models have been determined based on strength-duration data from neurons (e.g. Goldwyn et al., 2012). The strength-duration function represents the neural firing response threshold as a function of the pulse duration and can be described using the two parameters rheobase and chronaxie. Although previous studies described the strength-duration data by an exponential function (van den Honert and Stypulkowski, 1984; Shepherd et al., 2001) it has been suggested that, for extracellular stimulation, a linear function describing the threshold as inversely related to the pulse duration provides a better description of the data (Nowak and Bullier, 1998). In this case, the positive y-intercept of a regression line fit to the threshold data corresponds to the rheobase. The slope of the regression line divided by the rheobase corresponds to the chronaxie. This approach was used in the present study to derive the values for the rheobase and the chronaxie for monophasic anodic and cathodic pulses based on the ANF responses reported in Miller et al., 1999. The parameters and C in the Eqs. 3.1 and 3.2 were derived using these rheobase and chronaxie values.

Properties of the stochasticity of the neural responses have been described with a firing efficiency (FE) function which reflects the probability of spiking as a function of the stimulus level (Verveen, 1961). The FE function generally has a sigmoidal shape and can be approximated by an integral of a Gaussian distribution with mean θ and standard deviation σ. θ specifies the stimulus level that evokes a neural response with a probability of 0.5 and is defined as the threshold for ANF spiking. σ is a measure of the stochasticity of the neuron's response to a stimulus

Table 3.1: Complete set of obtained values

	Peripheral axon	Central axon
Membrane conductance, g_L	1.1 mS	2.7 mS
Membrane capacitance, C	856.96 nF	1772.4 nF
Slope factor, ΔT	10 mV	4 mV
Resting potential, E_L	-80 mV	
Threshold potential, $v_{threshold}$	-70 mV	
Peak potential, v_{peak}	24 mV	
Reset potential, v_{reset}	-84 mV	
Noise shaping parameter, α	0.8	
Inhibitory compression, β	0.75	
Subthreshold adaptation time constant, τ_{sub}	250 μs	250 μs
Suprathreshold adaptation time constant, τ_{supra}	4500 μs	2500 μs
Subthreshold adaptation conductance, a_{sub}	2 mS	
Suprathreshold adaptation conductance, a_{supra}	3 mS	
Dead time	500 μs	

at a given level. A normalized measure that combines θ and σ, the relative spread (RS), is obtained by taking the ratio of the standard deviation to the threshold of the neuron spiking (σ/θ). RS values are often reported to describe the dynamic range of the ANFs (Miller et al., 1999). Here, the RS values reported by Miller et al., 1999 for anodic and cathodic pulses were used to determine the intensity of entering the neural membrane by setting the variance of the underlying normal distribution to σ^2.

3.3 Methods

The effects of commonly used stimulus parameters, such as the pulse phase duration (PPD), interphase gap (IPG), stimulation pulse rate, and pulse level on the model responses were evaluated and compared to measured ANF responses. The

response statistics from measured ANF responses were obtained by digitizing the individual data points from the corresponding studies. The stimulus conditions and the statistical measures to describe the neural responses were chosen to be identical with the respective studies in order to facilitate a comparison of the predictions with the data. The model was implemented in MATLAB (version 2016a, The MathWorks Inc., Natick, MA). The differential equations of the model were solved using the forward Euler method with an integration time step of 1 μs. All simulations were run with a constant set of parameters shown in Table 3.1.

3.3.1   Stimuli

Single pulses: The considered pulse configurations for single pulses are illustrated in Fig. 3.2A-D. Although monophasic pulses of either polarity (Fig. 3.2A) can excite the ANFs, charge balancing is required for CI stimulation to avoid tissue damage. Symmetric biphasic pulses (Fig. 3.2B) can be regarded as the simplest charge balanced pulses, created by a monophasic pulse of one polarity immediately followed by the corresponding pulse of the opposite polarity. A drawback of the symmetric biphasic pulses is the interaction between the phases of opposite polarity. Alternative pulse shapes have been used to reduce the interaction between the phases of opposite polarities in a symmetric biphasic pulse. One strategy is to introduce a silent gap between the phases of opposite polarities (Fig. 3.2C). If such an inter phase gap (IPG) is large enough, the effect of the second phase on the neural response is minimal. A second strategy to reduce the interaction between the two phases with opposite polarities is to use asymmetric pulse shapes which are characterized by different pulse phase durations (PPD) of the two phases (Fig. 3.2D). For any given pulse duration and level of the leading phase, the second phase is made longer in order to neutralize the charge, but with a lower amplitude than for symmetric, biphasic pulses. The reduced amplitude of the second phase decreases its influence on the membrane potential. Since the duration of the two phases of opposite polarity is no longer the same and this pulse shape approximates monophasic stimulation, they are referred to as pseudomonophasic pulses.
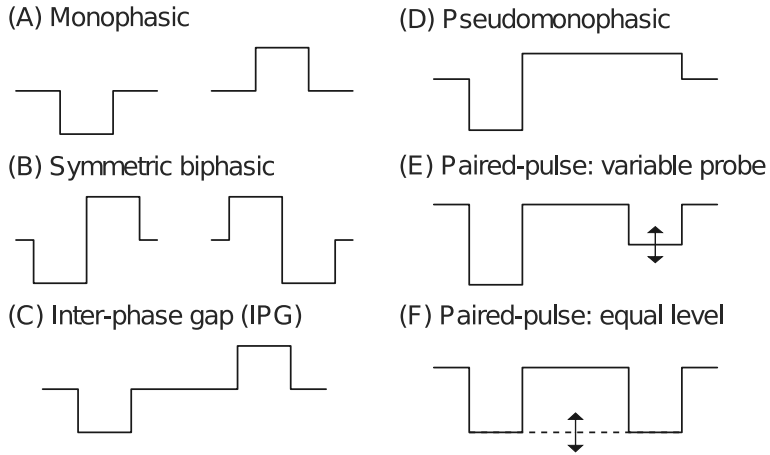
Figure 3.2: Stimulus conditions used to evaluate the model. **A** Monophasic cathodic and anodic pulses. **B** Symmetric biphasic pulses, with either cathodic or anodic leading polarity. **C** Symmetric biphasic pulse with an interphase gap (IPG). **D** Pseudomonophasic, charge-balanced pulse with different durations of the single pulse phases. **E** Paired pulse stimulus for the variable probe condition with constant conditioner (first) pulse and varied probe (second) pulse. **F** Paired pulse stimulus for the constant level condition, with conditioner and probe pulse at the same level.

In the present study, predictions were obtained for monophasic pulses with PPD of 25 to 500 μs of either anodic or cathodic polarity, and for biphasic pulses of either anodic or cathodic leading polarity with 25 μs to 10,000 μs PPD. The predicted response statistics for monophasic and biphasic pulses were compared with the data from Miller et al., 1999; Miller et al., 2001b; Rubinstein et al., 2001; Shepherd and Javel, 1999; Smith and Finley, 1997 and Bruce et al., 1999a. To test the effect of IPG in the model, thresholds were predicted for anodic leading and cathodic leading symmetric biphasic pulses with a PPD of 20 to 100 μs/phase and for IPGs from 2 to 200 μs. The predicted effects of IGP and PPD on the response statistics were compared to the data from Ramekers et al., 2014 and Shepherd and Javel, 1999. Furthermore, predictions were obtained for pseudomonophasic pulses with a leading short cathodic phase of 40 μs duration, followed by an anodic pulse of durations between 40 μs to 5 ms. The effect of the anodic phase duration on the response statistics was compared to data from Miller et al., 2001b.

Paired-pulses: Any stimulus, irrespective of whether it generates a spike or not, affects the neural response to a subsequent stimulus. Such interaction effects can be measured by stimulation with two subsequent pulses. In such paired-pulse stimuli, the first pulse is referred to as the 'conditioner' and the second pulse is referred to as the 'probe'. Two stimulus paradigms have been used to assess the effect of the conditioner on the probe. In one paradigm, the conditioner is presented at a fixed level and the probe threshold is measured for different delays of the probe (Fig. 3.2E). This condition is referred to as the *variable-probe condition*. In the second paradigm, the conditioner and the probe are presented at the same level and with different relative delays (Fig. 3.2F). This condition is referred to as the *constant-level condition*. By changing the level of the conditioner to be either below or above threshold, paired-pulse stimulation has been used to characterize the effects of facilitation, accommodation and refractoriness in the electrically stimulated ANFs.

The *subthreshold* conditioner refers to a pulse of low amplitude that does not generate a spike. In the variable-probe condition, the conditioner level was chosen to be either -0.9 or -2 dB below the threshold for a single pulse and probe thresholds were predicted for monophasic cathodic pulses of 100 μs PPD and inter-pulse delays ranging from 100 μs to 5000 μs. In the constant-level condition, the summation thresholds, i.e. the threshold for two pulses of equal level, were predicted for monophasic pulses of anodic or cathodic polarity for a PPD of 50 μs using inter-pulse delays in the range from 100 μs to 300 μs. A summation time constant was obtained by fitting the summation thresholds with an exponential function, as suggested in Cartee et al.Cartee et al., 2006. The effects of subthreshold paired-pulse stimulation in the model were compared to the data from Dynes, 1996 and Cartee et al., 2006.

The *suprathreshold* conditioner refers to a pulse of higher amplitude that always generates a spike. In the variable-probe condition, the conditioner level was fixed at either +1, +2, +4 or +6 dB above the threshold level for a single pulse and thresholds were predicted for 100-μs long monophasic cathodic probe pulses and inter-pulse delays ranging from 600 μs to 14,000 μs. In the constant-level condition, the probability of spiking for the second pulse was predicted for various inter-pulse

delays, using pseudomonophasic pulses. The pseudomonophasic pulses were composed of either anodic or cathodic leading phase (40 μs) and a long-duration second phase (160 μs) of opposite polarity. The conditioner and probe levels were kept constant at either +1 or +3 dB above the threshold for a single pulse and the inter-pulse delays were varied from 500 μs to 16,000 μs. The effects of the suprathreshold paired-pulse stimulation were compared to the data from Dynes, 1996 and Matsuoka et al., 2000.

*Pulse trains***:** Besides the pulse shape, pulse rate and stimulus level also affect the neural responses. Predictions were obtained for stimulation with pulse trains with rates from 200 to 10,000 pulses per second (PPS). The pulse trains were composed of symmetric biphasic pulses of either anodic or cathodic leading polarity. The amplitudes of the pulses in a pulse train were constant and varied across conditions between -10 to +10 dB above the threshold for a single pulse. The predictions were compared to the data from Javel, 1990; Bruce et al., 1999a; Zhang et al., 2007 and Miller et al., 2008.

3.3.2   Response statistics

For single and paired pulses, the response statistics used to characterize the neural response included threshold, dynamic range, spike latency and jitter. First, the spike probabilities were obtained by repeating the simulations 1000 times at multiple current levels. The probabilities were fit with an integrated Gaussian function described with mean θ and standard deviation σ to obtain the FE function. The threshold was defined as the value of θ of the FE function. The dynamic range of the neuron was estimated by calculating the RS as σ/θ. The spike latency distribution was calculated from the time delays of the generated spikes relative to the stimulus onset. The standard deviation of the spike latency distribution was then defined as the spike jitter.

For pulse trains, the response can be described by the spike rate and its variability, the dynamic range, the inter-spike interval (ISI) histogram and the vector strength (VS). The spike rate is defined as the observed number of spikes per second for any given stimulus at a particular stimulus level. The variability in spike rate can

be quantified using the Fano factor. The Fano factor is a normalized measure of the variance in spike rate and can be described as the ratio of the standard deviation of the spike rate to the mean spike rate, evaluated across repetitions of the stimulus. The dynamic range of the neuron was derived by calculating the rate-level function via simulations of the spike rate at multiple levels. The ISI histogram was considered to represent information related to the temporal dispersion of the spike times and was constructed with a bin-width of 1 ms. The VS was calculated to quantify the periodicity in the neural responses. It represents the phase locking of the spike times in response to sinusoidal stimulation:

$$VS = \frac{1}{N} \sqrt{\left[\sum_{i=1}^{N} cos\left(\frac{2\pi t_i}{T}\right)\right]^2 + \left[\sum_{i=1}^{N} sin\left(\frac{2\pi t_i}{T}\right)\right]^2} \qquad (3.7)$$

where N is the total number of spikes used in the analysis, t_i is an individual spike time and T is a period of the sinusoid to which the synchronization of spike times is being assessed. The value of the VS can maximally be 1, indicating perfect phase locking, and minimally 0, indicating no phase locking to the stimulus frequency. In order to avoid any effect of the onset responses to the stimuli, the spikes appearing in the first 50 ms of the stimulus were excluded from the calculation of the VS. The effect of the pulse rate on the VS was obtained for the pulse train stimuli presented at a level of 1 dB above the threshold for a single pulse. The effect of level on the model responses was quantified by selecting stimulus levels which result in spikes rates from 4 to 500 spikes/second.

Changes in neural responses across time were investigated by constructing the peri-stimulus time histograms (PSTH) and wide-bin adaptive peri-stimulus time histograms (aPSTH) in response to pulse trains. The PSTH were constructed with a bin-width of 1 ms from responses to the pulse train stimuli presented at a level of 1 dB above the threshold for a single pulse. The aPSTH refers to the PSTH constructed with increasing time windows across the stimulus duration. Following Zhang et al., 2007, the aPSTH were constructed for temporal windows of 0–4, 4–12, 12–24, 24–48, 48–100, 100–200, and 200–300 ms from the time onset in response to

a 300-ms-long pulse train.

3.4 Results

3.4.1 Responses to single-pulse stimulation

The simulated response statistics for *monophasic* anodic and cathodic pulses of 26- and 39-μs duration are shown in Fig. 3.3. The effects of current level on the FE, the spike latency, and the jitter for anodic (green upward pointing triangles) and cathodic (blue downward pointing triangles) pulses of 39-μs duration are shown in Fig. 3.3a–c, respectively. The responses to anodic and cathodic pulses are probabilistic, and the probability of spiking increases with an increase in stimulus level (Fig. 3.3a). The solid lines represent the FE functions fitted to the predicted probabilities with the 50 % points indicated as thresholds. The model predicts lower thresholds for cathodic than for anodic pulses. The spike latencies (Fig. 3.3b) are generally higher for cathodic than for anodic pulses and decrease with increasing pulse level. The jitter (Fig. 3.3c) decreases with increasing pulse level for both cathodic and anodic pulses.

The comparison of the predictions to data for 26-μs and 39-μs long anodic and cathodic monophasic pulses is shown in Fig. 3.3D-F. Thresholds (Fig. 3.3D) are higher for 26 μs duration than for 39 μs for both, anodic and cathodic pulses. However, the pulse duration has only a minor effect on the spike latency (Fig. 3.3E) or its jitter (Fig. 3.3F). The predictions (left panels) and the data (right panels) are in good agreement, except for the absolute values of the spike latencies which are about 200 μs lower than the corresponding values in the data. Fig. 3.3G shows the differences in spike latencies between anodic and cathodic pulses of 39 μs as a function of the probability of spiking. The predictions are indicated by the black line. For comparison, the data by Miller et al., 1999 are represented by the red circles. The difference between spike latencies for anodic vs cathodic pulses is about 200 μs for low spiking probabilities and reduces to about 150 μs for high spiking probability. The predictions are consistent with the data, i.e. the model can successfully account for the spike-latency differences between anodic and
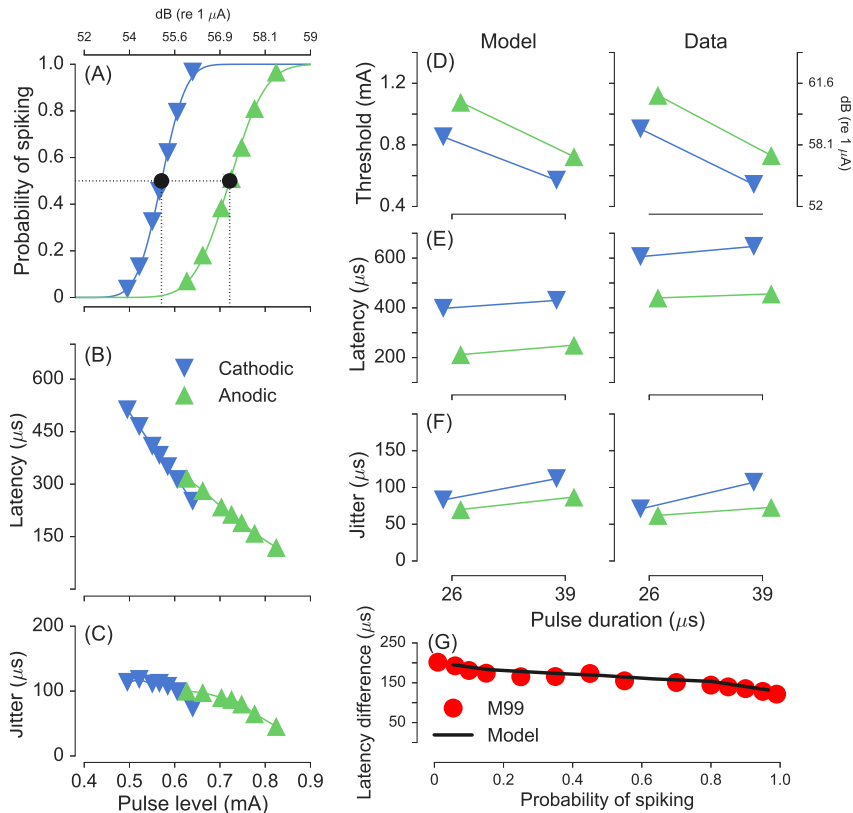
Figure 3.3: Responses to monophasic anodic (green downward pointing triangles) and cathodic (blue upward pointing triangles) pulses. **A** FE curves predicted by the model in response to 39-μs duration pulses together with a fit using an integrated Gaussian (lines). **B,C** Corresponding spike latency and jitter as a function of the pulse level. **D-F** Comparison of predicted thresholds, spike latencies, and jitter (left) with the corresponding data from Miller et al., 1999 (right) in response to monophasic pulses of 26- and 39-μs duration. **G** Predicted (black line) differences in spike latencies between cathodic and anodic monophasic pulses together with data from Miller et al., 1999 (red circles).

cathodic pulses.

Figure 3.4 shows the simulated response statistics for symmetric biphasic pulses. Figure 3.4A represents the thresholds for cathodic leading biphasic pulses (ordinate) in comparison to those for cathodic monophasic pulses (abscissa). Figure 3.4B represents the corresponding results for the spike latencies. The black squares

Figure 3.4: Responses to symmetric biphasic pulses. **A, B** Comparison of thresholds and spike latencies for a monophasic cathodic pulse and a cathodic leading symmetric biphasic pulse. Predictions are indicated by the black squares, and distributions of the data from Miller et al., 2001b are shown by the purple kernel density functions. **C** Predicted effect of PPD on th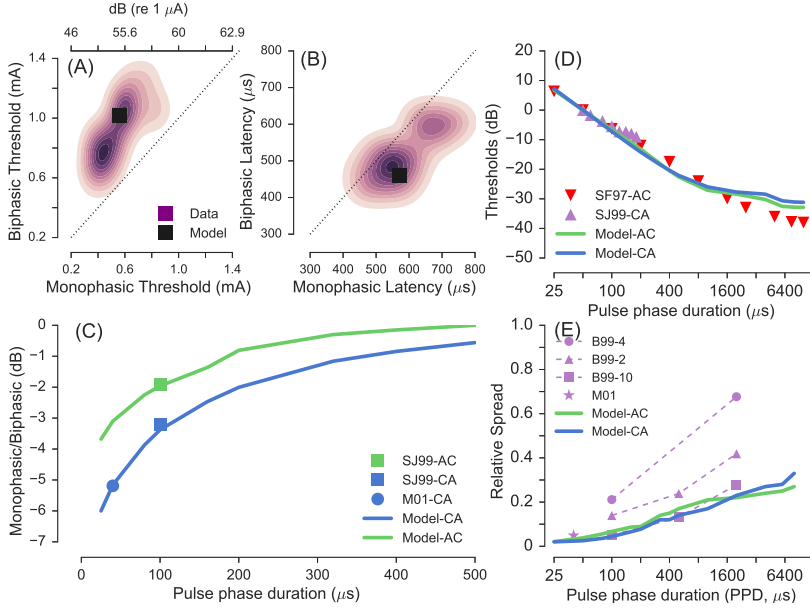e difference between monophasic and biphasic pulses of both polarities (lines) along with corresponding data from Miller et al., 2001b (M01-CA) and Shepherd and Javel, 1999 (SJ99AC, SJ99CA). **D** Predicted effect of PPD on the threshold for symmetric biphasic pulses of anodic (AC) or cathodic (CA) leading polarities (lines) along with the corresponding data from Shepherd and Javel, 1999 (SJ99-CA) and Smith and Finley, 1997 (SF97-AC). **E** Predicted effect of PPD on the RS (lines) along with the corresponding data from Bruce et al., 1999a (B99-4, B99-2, B99-10) and Miller et al., 2001b (M01).

in Fig. 3.4A and 3.4B represent the predictions. The shaded area indicates the kernel density function (estimated using function "KDEplot" of the seaborn toolbox in python) of the ANF responses in Miller et al., 2001b, whereby each shade represents a step of ten percentiles in the data. Both predictions and data show that the thresholds for biphasic pulses are higher than those for monophasic pulses (Fig. 3.4A) and that the spike latencies for biphasic pulses are lower than for monophasic pulses (Fig. 3.4B). Figure 3.4C shows the predicted differences in thresholds between monophasic and biphasic pulses with corresponding leading polarity (leading

cathodic (CA); leading anodic (AC); shown as lines) as a function of PPD. The predictions show that the difference between monophasic and biphasic pulses is largest for smaller PPD and reduces with increasing PPD. In general, the difference in threshold between monophasic and biphasic pulses is lower for monophasic anodic and AC pulses than for monophasic cathodic and CA pulses. The predictions are consistent with the available data obtained for individual values of PPD (Miller et al., 2001b, M01-CA with a circle; Shepherd and Javel, 1999, SJ99 with squares).

Figure 3.4D-E show the effect of PPD of symmetric biphasic pulses on threshold and RS. The predicted thresholds for CA pulses (indicated by the blue line) and AC pulses (indicated by the green line) are shown as a function of PPD. The thresholds have been normalized relative to the threshold for a biphasic pulse with 50 μs PPD. The thresholds are highest for short PPDs and decrease monotonically with increasing PPD, both for CA and AC pulses. This decrease is consistent with the data in Smith and Finley, 1997 and Shepherd and Javel, 1999, indicated by the downward (SF97) and upward pointing triangles (SJ99), respectively. The largest difference between the predictions and the data appears at longer PPDs beyond 2000 μs/phase, where the predictions are higher than the data observed by Smith and Finley, 1997. In Fig. 3.4E, the predicted RS values (solid lines) increase monotonically with increasing PPD, both for CA and AC pulses. This prediction is consistent with the trends in the data in Bruce et al., 1999a that also showed that the RS of an electrically stimulated ANFs in cat increased systematically with increasing PPD (B99-2, B99-4 and B99-10, green points in Fig. 3.4E). The absolute values of the predicted RS are consistent with the fiber labeled as B99-10 (indicated by purple square), while the remaining two fibers (B99-2, B99-4) show higher values of the RS.

Figure 3.5 shows simulated response statistics and corresponding data for stimulation with symmetric- and asymmetric pulse shapes. Figure 3.5A-C show the effect of IPG on thresholds for biphasic pulses with PPDs of 20, 50 and 100 μs/phase. The predictions for AC and CA pulse shapes are shown as green and blue lines, respectively. The data from Ramekers et al. (2014) for 20 μs/phase and 50 μs/phase are indicated by the red circles in Fig. 3.5A-B, and those from Shepherd and Javel (1999) for 100 μs/phase indicated by the red squares in Fig. 3.5C. The

thresholds decrease exponentially with increasing IPG both in the predictions
and the data. The model also predicts an effect of PPD on the threshold decrease
due to an IPG; the effect of IPG is stronger for smaller PPDs (Fig. 3.5A) than for
longer PPDs (Fig. 3.5C), and the effect of IPG is stronger for CA pulses than for AC
pulses. The predicted effect of PPD on the threshold reduction with increasing IPG
is consistent with the corresponding data.
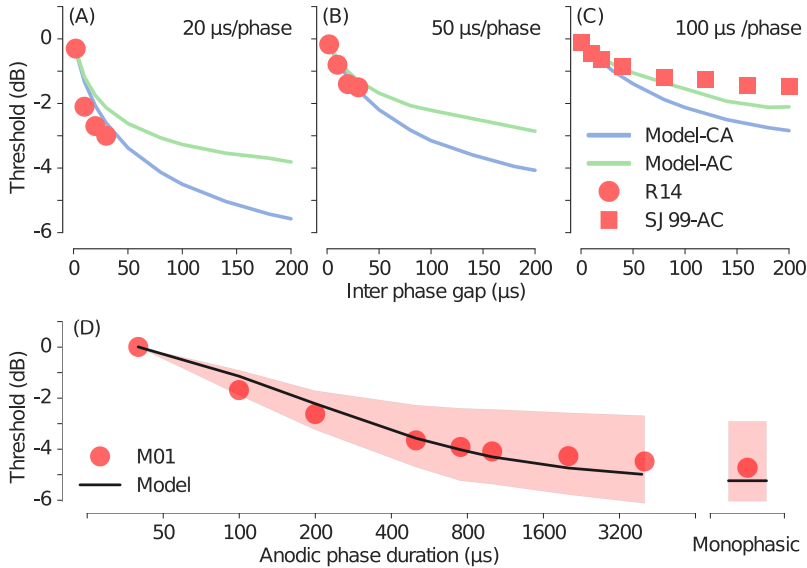
Figure 3.5: Effect of IPG on threshold for biphasic pulses of PPDs for **(A)** 20 μs/phase **(B)** 50 μs/phase,
and **(C)** 100 μs/phase. Predictions (lines) are compared to the corresponding data from Ramekers et al.,
2014 (R14) and Shepherd and Javel, 1999 (SJ99-AC). **D** Thresholds for pseudomonophasic pulses as a
function of the anodic phase durations along with the threshold for monophasic cathodic pulse for
comparison. Predictions (line) are compared to the corresponding data from Miller et al., 2001b (M01).
The shaded area represents the spread of the data.

Figure 3.5D shows the predicted thresholds for pseudomonophasic pulses as a
function of the duration of the second, anodic phase (black line). The thresholds
were normalized relative to the threshold for a symmetric biphasic pulse with 40
μs PPD. The threshold for a cathodic monophasic pulse of 40 μs is also shown on
the right for comparison. The corresponding mean data from Miller et al., 2001b

are indicated by the red circles, and standard deviation of the data is represented by the shaded area. Both predictions and data show that the thresholds decrease monotonically with increasing anodic phase duration. Thresholds continue to decrease beyond 1000 μs asymptotically approaching the threshold for a monophasic cathodic pulse. The predictions are in good agreement with the data.

3.4.2   Responses to Paired Pulse Stimulation

The simulated response statistics for paired pulses with a subthreshold conditioner are shown in Fig. 3.6. Figure 3.6A shows the predicted thresholds (solid and dotted black lines) for paired pulses in the variable-probe condition as a function of the interpulse delay. The thresholds were normalized relative to the threshold for a single pulse, which is indicated by a dashed horizontal line. The predictions show that for interpulse delays below about 800 μs, the thresholds for the paired pulses are notably lower than the thresholds for a single pulse. This process of temporal integration of the charge has been referred to as facilitation. For interpulse delays above 800 μs, thresholds are elevated relative to the threshold for a single pulse. This form of subthreshold masking has been referred to as accommodation. After about 2000 μs, the threshold approaches the threshold for a single pulse. The amount of facilitation observed at delays below 800 μs depends on the level of the conditioner pulse. In the predictions, the effect of facilitation is slightly stronger for a -0.9 dB (dashed black line) subthreshold conditioner than for a -2.0 dB (solid black line) subthreshold conditioner. The predictions are consistent with the data of Dynes, 1996 indicated by the light red circles (for the conditioner level of -2.0 dB) and in red circles (for the conditioner level of -0.9 dB).

In Figure 3.6B, predicted spike latencies (ordinate) and summation time constants (abscissa) are shown for paired-pulses stimulation in the equal-level condition. The black upward pointing triangle shows the results for anodic pulses, and the black downward pointing triangle shows the results for cathodic pulses. The summation time constant in the model is about 175 μs for the anodic paired pulses and 280 μs for the cathodic paired pulses. The summation latency amounts to 390 μs for the anodic pulses and 520 μs for the cathodic pulses. These predictions are

well within the range of the data from Cartee et al., 2006 which are indicated by the kernel densities of the distribution of ANF responses for the peripheral site of spike generation (blue area) and the central sites of spike generation (green area). The color shades of the kernel density functions represent a step of ten percentiles of the data.

Figure 3.6: Responses to subthreshold paired pulse stimulation. **A** Predicted effect of conditioner level on the probe threshold as a function of the interpulse delays (lines) along with the corresponding data from Dynes, 1996 (D96). **B** Comparison of the summation latencies obtained with the equal-level condition for the anodic and the cathodic pulses (abscissa) with the corresponding summation latencies (ordinate). Predictions (triangles) are compared with the corresponding distributions of the data of Cartee et al., 2006 (blue and green kernel density functions).

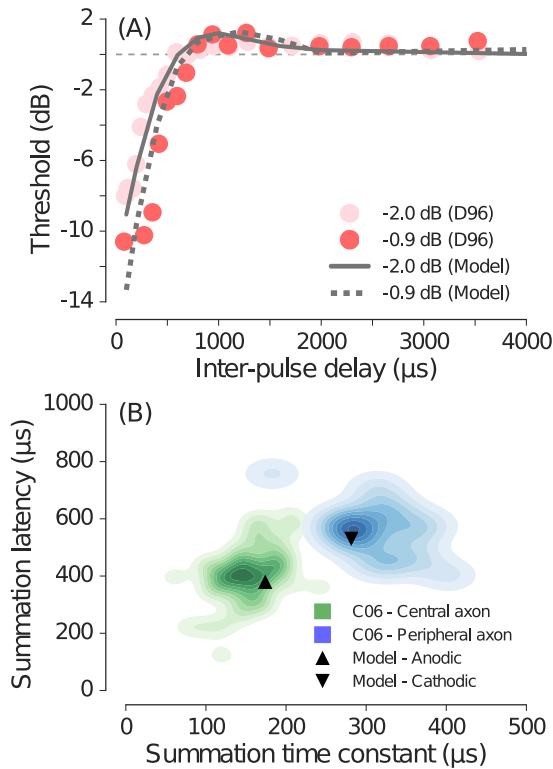The predictions for paired pulse stimulation with pseudomonophasic pulses
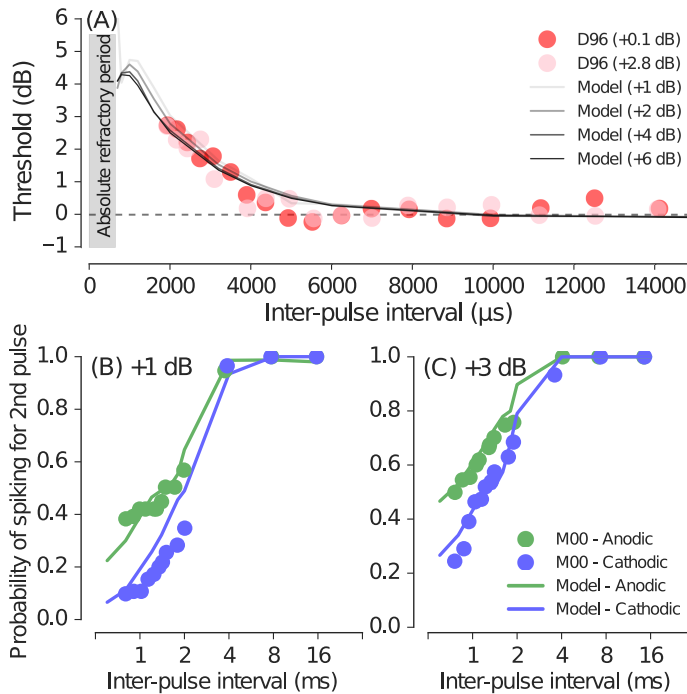
Figure 3.7: Responses to suprathreshold paired pulse stimulation. **A** Predicted effect of the suprathreshold conditioner level on the probe threshold as a function of the interpulse interval (lines) along with the corresponding data from Dynes, 1996 (D96). **B-C** Probability of spiking in response to the probe measured in the equal-level condition for two suprathreshold levels +1 dB **(B)** and +3 dB **(C)**. Predictions (lines) are compared to the corresponding data from Matsuoka et al., 2000 (M00). Note that single-pulse threshold for the cathodic pseudomonophasic pulse is lower (810 μA) than for the anodic pseudomonophasic pulse (885 μA).

with a suprathreshold conditioner are shown in Fig. 3.7. Figure 3.7A shows the predicted thresholds for the paired pulses in the variable-probe condition as a function of the interpulse interval. The different lines indicate the results for different conditioner levels. The thresholds have been normalized relative to the threshold for a single pulse, indicated by the dashed horizontal line. The predictions show that there is no spike generation in the interval immediately following a spike produced by the suprathreshold conditioner, as indicated by the ARP of about 600 μs (shaded gray area). Previous studies have reported mean ARP within a range of 300

to 700 μs (Dynes, 1996; Miller et al., 2001a; Imennov and Rubinstein, 2009). Beyond the ARP, the thresholds are increased relative to the thresholds for a single pulse for interpulse intervals of up to about 5000 μs. This period where the threshold is elevated represents the refractory period of the neuron. The predictions show only a negligible effect of the conditioner level on the refractory period, and the thresholds approach the baseline level of a single pulse threshold at about 5000 μs. The predictions are in good agreement with data from Dynes, 1996 that are indicated by the filled red and pink circles for the conditioner levels of +0.1 and +2.8 dB, respectively, as well as the estimates of Miller et al., 2001a of the refractory period.

Figure 3.7B shows the probability of spiking for the second pulse in the case of suprathreshold paired pulse stimulation in the equal-level condition for a level of +1 dB re single pulse threshold. The results for a level of +3 dB are shown in Fig. 3.7c. The blue and green lines indicate the results for cathodic and anodic pulses, respectively. The results show that the probability of spiking increases with increasing interpulse interval. The probability approaches 1 for interpulse intervals of about 5 ms, at both stimulus levels. For interpulse intervals below 5 ms, the probabilities are lower for cathodic than for anodic pulses. The difference between the probabilities for anodic and cathodic pulses decreases with increasing interpulse delays. The predictions are consistent with corresponding data from Matsuoka et al., 2000 represented as blue and green circles for cathodic and anodic pulses, respectively.

3.4.3   Responses to Pulse Trains

The predictions for pulse train stimulation are shown in Fig. 3.8. The predicted rate-level functions (Fig. 3.8A) are indicated as spikes/s for pulse rates in the range from 100 to 800 pps. The predictions show a monotonic increase in spike rate with increasing level. For the corresponding pulse rates, the spike rates saturate at levels between 60 and 65 dB to spike rates equal to the pulse rate. The dynamic range of the neural response for pulse train stimuli can be extracted as a difference between the neuron's threshold level and the level at which the spike rate saturates.
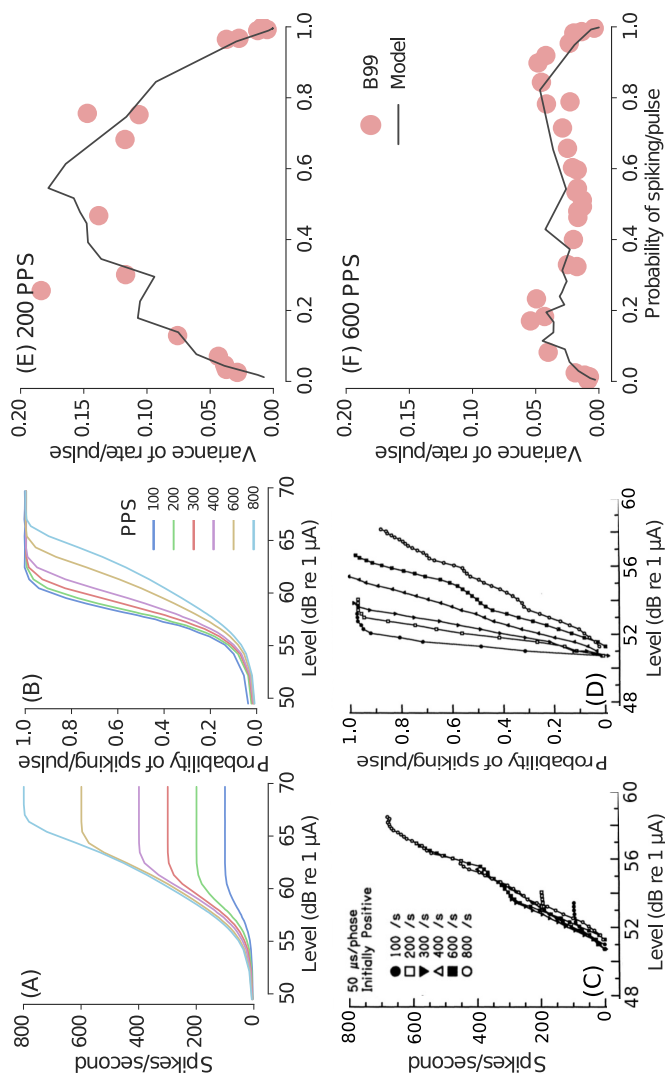
Figure 3.8: Predicted rate-level functions and spike rate variability in response to pulse train stimuli. **A** Predicted rate-level functions for stimulation rates ranging from 100 to 800 pps. **B** Predicted rate-level functions for stimulation rates ranging from 100 to 800 pps from Javel, 1990. **C** Corresponding measured rate-level functions normalized to the stimulus pulse rate. **D** Normalized rate-level functions from Javel, 1990. **E** Predicted variance in spike rate for stimulation with a pulse train of 200 pps (line) along with the data from Bruce et al., 1999a (B99, circles). **F** Predicted variance for stimulation with a pulse train of 600 pps (line) along with the corresponding data from Bruce et al., 1999a (B99, circles). The data in **C** and **D** have been reprinted from Javel, 1990 (Fig. 17.21) with kind permission of Springer Science + Business Media.

To measure the effect of the pulse rate on the dynamic range of the neuron, the rate-level functions can be normalized by dividing the spike rate by the stimulus pulse rate. Figure 3.8B shows the corresponding normalized rate-level functions representing the probability of spike/pulse for pulse rates in the range from 100 to 800 pps. It can be seen that the dynamic range of the neuron is larger for higher pulse rates than for lower pulse rates. Despite some differences in the absolute stimulus levels, the predicted rate-level functions for the different pulse rates (Fig. 3.8A, B) are consistent with corresponding data shown in Javel, 1990 (Fig. 3.8C, D).

Figure 3.8E shows the variability in spike rate across the dynamic range of the neuron for the pulse rate of 200 pps. The normalized dynamic range (abscissa) is indicated as the probability of a spike/pulse. The predictions, indicated by the black lines, show that the variability of the response for 200 pps is largest for probabilities around 0.5. For 600 pps (Fig. 3.8F), the variability is largest at about 0.25 and 0.75 and slightly lower at intermediate probabilities around 0.5. The absolute variance in spike rate is larger for 200 pps than for 600 pps. The data reported in Bruce et al., 1999a, indicated by the red circles, show a comparable effect of the simulation rate on the variability of the response.

Figure 3.9A shows the predicted variability of spike rate in terms of the Fano factor as a function of the mean spike rate for stimulus pulse rates of 250 (blue), 1000 (red), and 5000 pps (green). Figure 3.9b shows the data of Miller et al., 2008 indicated as the Fano factor as a function of the mean spike rate for stimulus pulse rates of 250 (circles), 1000 (squares), and 5000 pps (triangles). The predictions (Fig. 3.9A) as well as the data (Fig. 3.9B) indicate that the Fano factor is larger for low spike rates (corresponding to low stimulus levels) and decreases with increasing spike rates (corresponding to increasing stimulus levels). The stimulation pulse rate shows only a marginal effect on the Fano factor. The predictions are roughly consistent with the data in Fig. 3.9B. However, many data points, with each point corresponding to a different ANF, show notably higher values of the Fano factor (>1) than the predicted values, particularly at spike rates around 100 spikes/s, i.e., the high variability in the data across ANFs cannot be accounted for by the model.

Figures 3.10, 3.11, and 3.12 show the predicted response statistics related to the measures of temporal dispersion in the spike times, such as ISI histograms and
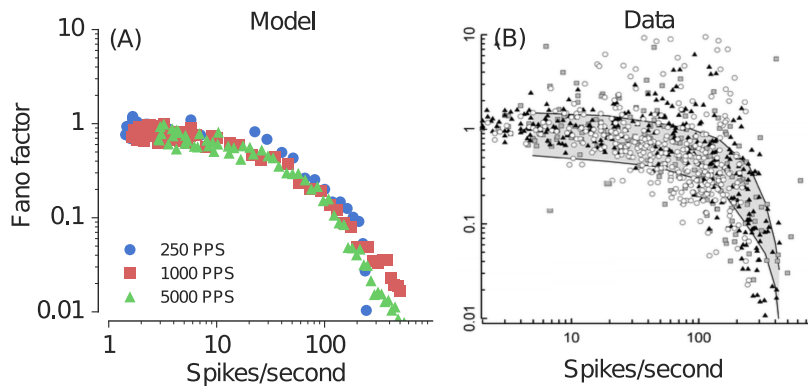
Figure 3.9: Fano factor derived from the predicted **(a)** and measured **(b)** ANF responses for stimulation with pulse trains of 250 pps (circles), 1000 pps (squares), and 5000 pps (triangles) at various stimulus levels as a function of the spike rate. The corresponding data **(b)** have been reprinted with kind permission of the Journal of the Association for Research in Otolaryngology, Springer Science + Business Media: Fig. 9 from Miller et al., 2008, © Association for Research in Otolaryngology 2007.

vector strength (VS). Figure 3.10 shows the simulated ISI histograms for stimulus pulse rates of 250, 1000, and 5000 pps. The data from Miller et al., 2008 are indicated in the insets. For pulse trains with a rate of 250 pps (Fig. 3.10A), the histogram shows the largest peak at about 4 ms, which corresponds to the pulse train period representing the (inverse of the stimulation pulse rates). The secondary peaks in the histogram appear at multiples of 4 ms. A similar trend was found for pulse rates of 1000 pps (Fig. 3.10B) where the histogram shows multiple peaks in intervals of 1 ms. However, the largest peak in the histogram appears at about 5 ms and not at the time corresponding to the inverse of 1000 pps as in the case for 250 pps. For 5000 pps (Fig. 3.10C), the histogram shows a single peak around 4 ms obtained with the temporal resolution used to construct the ISI histogram. The ISI histograms for 1000 and 5000 pps indicate that the neuron does not to generate a spike for each pulse in the pulse train. Hence, the largest peak in the distribution does not correspond to the period of the pulse train. Such transition between discrete ISI distributions at low pulse rates to more continuous distributions at higher pulse rates is consistent with data in Miller et al., 2008.

Figure 3.11 shows the VS of the neural response as a function of the stimulation
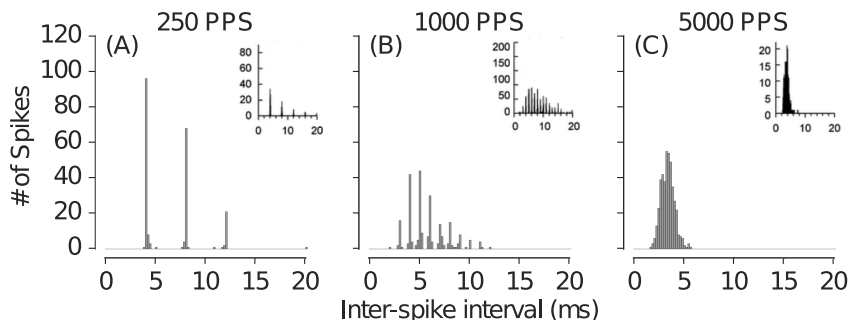
Figure 3.10: ISI histograms with a bin width of 1 ms derived from the predicted responses for stimulation with pulse trains of 250 pps **(A)**, 1000 pps **(B)**, and 5000 pps **(C)**, along with the corresponding data from Miller et al., 2008 in the insets. The data in the insets have been reprinted with kind permission of the Journal of the Association for Research in Otolaryngology, Springer Science + Business Media: Fig. 1 from Miller et al., 2008, © Association for Research in Otolaryngology 2007.

pulse rate. The predictions (Fig. 3.11A) show values close to 1 for low pulse rates, decreasing to about 0.4 for 10,000 pps. This prediction is consistent with the data from Hartmann and Klinke, 1990 (downward pointing triangle for HK90) and Dynes and Delgutte, 1992 (upward pointing triangle for DD92) shown in Fig. 11b. Thus, both data and predictions maintain a sustained synchronization to the stimulus frequencies even beyond 5000 pps.

Figure 3.12 shows the effect of the stimulus level on VS for pulse rates of 250, 1000, and 5000 pps. The dynamic range of the neural response depends on the stimulation pulse rate (as shown in Fig. 3.8). Therefore, the VS is shown as a function of the spikes/s instead of the stimulus level. For 250 pps (Fig. 3.12A) and 1000 pps (Fig. 3.12B), the VS approaches 1 with increasing spike rate. This indicates strong phase locking at higher levels for 250 and 1000 pps. For 5000 pps (Fig. 3.12C), the maximum value of the VS approached is about 0.7. The predictions represent the main trends observed in the data of Miller et al., 2008 shown in Fig. 3.12D-F. However, for 5000 pps, the predictions deviate from the data in that they show a nonmonotonic effect of spike rate on the VS, whereas the data indicates no clear change in VS with increasing level (Fig. 3.12F).

Figure 3.13 shows the changes in spike rate over time in terms of the PSTH and
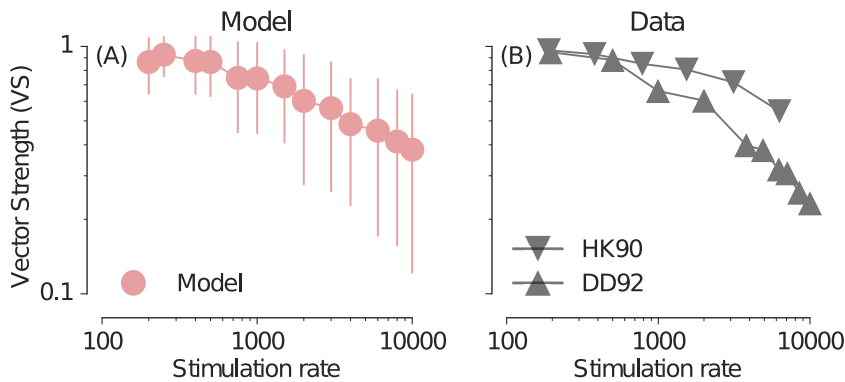
Figure 3.11: **A** Vector strength (VS) derived from the predicted responses as a function of the stimulation rate (pps). The error bars represent the standard deviation of the VS across 100 repetitions. **B** Corresponding data from Hartmann and Klinke, 1990 (HK90, downward pointing triangles) and Dynes and Delgutte, 1992 (DD92, upward pointing triangles).

aPSTH obtained from responses to stimulation rates of 250 pps (Fig. 3.13A), 1000 pps (Fig. 3.13B), 5000 pps (Fig. 3.13C), and 10,000 pps (Fig. 3.13D). The top row of Fig. 3.13 shows PSTH (gray bars) constructed from the model responses. For 250 pps (Fig. 3.13A), the PSTH shows spikes occurring in response to each pulse of the pulse train. For higher pulse rates, a strong onset response is followed by a decrease in the spike rate. The decrease of the spike rate relative to the onset is larger for 1000 pps (Fig. 3.13B) than for 5000 (Fig. 3.13C) and 10,000 pps (Fig. 3.13D). This effect of pulse rate on the changes in response over time can also be observed in aPSTH which is indicated by the red circles. The first point of the aPSTH, which corresponds to the spike rate within the time period of the 0 to 4 ms from the stimulus onset, is always larger than the remaining points. This represents the onset response. The aPSTH show that the onset response becomes stronger with increasing pulse rate (Fig. 3.13A-D). After the onset response, there is a substantial reduction of the response that is observed, as is reflected in the aPSTH. The predictions follow the trends observed in the data of Zhang et al., 2007 for the pulse rates of 250, 1000, and 5000 pps, indicated in the bottom row in Fig. 3.13E-G. However, the data for 10,000 pps (Fig. 3.13H) show that the response
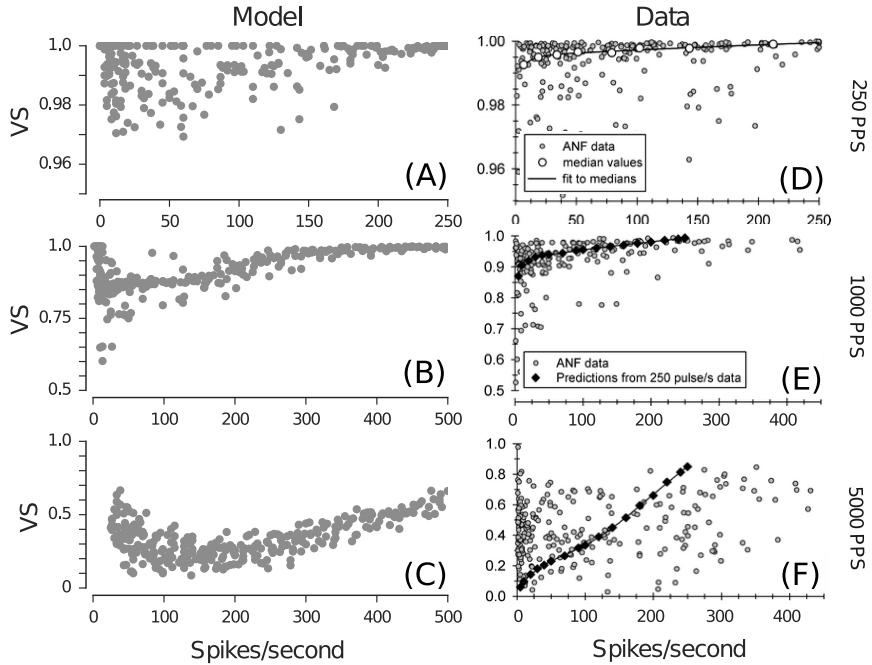
Figure 3.12: Predicted changes in VS as a function of the spike rate for stimulation with pulse trains of 250 pps **(A)**, 1000 pps **(B)**, and 5000 pps **(C)** along with the corresponding data from Miller et al., 2008 **(D-F)**. The data have been reprinted with kind permission of the Journal of the Association for Research in Otolaryngology, Springer Science + Business Media: Fig. 8 from Miller et al., 2008, © Association for Research in Otolaryngology 2007.

exhibits a large onset response followed by a strongly reduced spiking of the ANFs (Zhang et al., 2007). Such a "shutdown" of the neural activity following the strong onset response is not accounted for by the model (Fig. 3.13D).

3.5 Discussion

3.5.1 Summary of Main Findings

In this study, a computational model of the ANF responses to electrical stimulation was presented. The structure of the model was motivated by the experimental
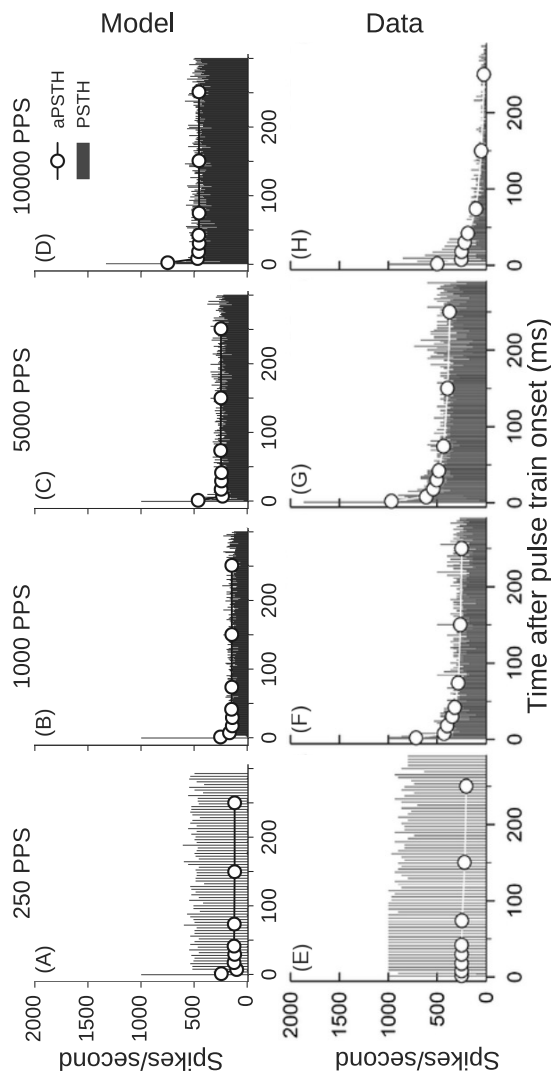
Figure 3.13: PSTH (gray bars) and aPSTH (red circles) constructed from the predicted responses for stimulation with pulse trains of 250 pps (**A**), 1000 pps (**B**), 5000 pps (**C**), and 10,000 pps (**D**) along with the corresponding data from Zhang et al., 2007 (**E–H**). The data have been reprinted with kind permission of the Journal of the Association for Research in Otolaryngology, Springer Science + Business Media: Fig. 2 from Zhang et al., 2007, © Association for Research in Otolaryngology 2007.

observation that both anodic and cathodic charges can generate a spike, whereby the site of spike generation differs for the two polarities. The model assumes two independent point neurons which mimic two sites of spike generation, namely the peripheral and the central axons of the ANF. The differences in sensitivity for anodic and cathodic charges was introduced in the model through the parameters derived from the rheobase and the chronaxie for monophasic pulses, using a procedure which describes the strength-duration function as a linear function instead of the traditionally considered exponential function. The model was evaluated in stimulus conditions with single pulses of various pulse shapes as well as with paired pulses and pulse trains. The results demonstrate that the charge integration properties of the neural membrane derived from responses to monophasic pulses can characterize the responsiveness of the neuron to electrical stimulation sufficiently well to account for the stimulation with biphasic pulses and pulses of alternative shapes.

The dynamic response properties of the ANF responses were modeled by two feedback current loops, describing the subthreshold and the suprathreshold currents which continuously modify the membrane voltage with respective time constants (τ_{sub} and τ_{supra}). The values of these adaptation time constants were adjusted to match the data obtained with paired pulse stimulation and were evaluated in conditions with pulse trains of various pulse rates and levels. The results demonstrated that this model can correctly predict the trends in the response statistics, such as the Fano factor and the VS. The parameters of the feedback current loops that were derived to match the data for paired pulse could, at least partly, be successfully applied to more dynamic stimuli such as pulse trains.

3.5.2 Relation to Existing Models

Multiple phenomenological models of the ANF responses to electrical stimulation have been proposed (Bruce et al., 1999b; Bruce et al., 1999a; Rubinstein et al., 2001; Litvak et al., 2003; Nourski et al., 2006; Macherey et al., 2007; Fredelake and Hohmann, 2012; Goldwyn et al., 2012; Morse et al., 2015; Horne et al., 2016). These models do not consider multiple sites of spike generation and their effect

on spike time statistics, and hence cannot be generalized to assess different CI stimulation strategies (Joshi et al., 2014). In contrast, the model presented in this study considers two sites of spike generation along the ANF as well as differences in their sensitivity to anodic and cathodic charges and is shown to account for the effect of different pulse shapes on the ANF response statistics.

The existing models also fail to predict the effects of subthreshold and suprathreshold stimuli on the responses to the following stimulation, which strongly affect the temporal responses of the ANF responses (Boulet et al., 2016). Only Goldwyn et al., 2012 used the summation threshold time constants to model facilitation, enabling their model to predict temporal integration effects across pulse trains. Nevertheless, their model did not include any effects of accommodation and therefore cannot account for accommodation observed in spike responses at different pulse rates. In the model proposed in the present study, the response properties of facilitation as well as the accommodation are accounted by the combination of the passive membrane filtering and the inclusion of a single subthreshold adaptive feedback loop.

The existing models simulate the stochasticity in the ANF responses using the stochastic threshold framework proposed by Bruce et al., 1999b; Bruce et al., 1999a. In that framework, a spiking threshold is assumed to be random with a Gaussian distribution identical to that underlying the FE function and is approximated by white noise. In contrast, in the present study, the stochasticity is simulated as fluctuations in the membrane voltage using spectrally shaped noise ($1/f^\alpha$) instead of the white noise. The use of $1/f^\alpha$ shaped noise produces physiologically realistic fluctuations of the membrane voltage (Verveen and Derksen, 1965) and has also implications for the temporal response properties of the model. The $1/f^\alpha$ shaped spectrum of the noise enhances the coding of the lower frequencies in the ANF responses and results in correlated activity due to low frequency oscillations in the membrane across the multiple fibers (Pozzorini et al., 2013).

Furthermore, the existing models simulate the difference in thresholds for monophasic and biphasic pulses using a parameter called the activation time (Rubinstein et al., 2001). The higher thresholds observed for biphasic pulses have been attributed to the inhibitory properties of the second phase (van den Honert and

Mortimer, 1979). It has been proposed that there exists a time delay, the activation time, between the point in time when the neural membrane voltage has crossed a threshold for spiking and when the actual spike occurs; if an inhibitory charge occurs during this delay, the initiated spike can be canceled. Rubinstein et al., 2001 showed that the inclusion of a constant activation time delay in a simple linear integrate-to-threshold model can account for the threshold differences between monophasic and biphasic pulses. Such a delay has also been considered in recent models, such as the leaky integrate-and-fire model proposed by Horne et al., 2016. In contrast to that model, the model presented here does not require such activation time to account for the differences in thresholds between the monophasic and biphasic pulses, since the model does not indicate a spike when it crosses the threshold voltage. Instead, the membrane voltage starts to grow exponentially after it has crossed the threshold voltage. If enough inhibitory stimulation occurs during the exponential upswing of the membrane voltage, the spike is canceled. This inherent feature in the exponential integrate-and-fire neuron provides a biophysically relevant alternative to the use of an activation time.

Finally, it has been suggested that spike latency differences between the peripheral and the central axons occur due to a somatic delay. Accordingly, several models have assumed that spikes generated on the peripheral axon must pass through the soma where they are delayed due to a high capacitance of the soma (Rattay et al., 2001; Rattay et al., 2013). However, the role of the soma in spike generation or spike conduction has not yet been confirmed experimentally. In fact, some biophysical models that do not include the soma and simulate the ANF as a uniform cable have also been successful in capturing various dynamics of the ANF responses (e.g., Imennov and Rubinstein, 2009). The model presented in this study was also able to account for the differences in spike latencies between spikes generated at the peripheral and the central axons without assuming the passive compartment of the soma to add a delay to the spikes generated at the peripheral axons. Instead, latency difference was achieved by employing different values of the term ΔT in Eq. (2) between the two axons. Biophysically, ΔT has been related to the sharpness of the voltage-gated Na+ channel activation in the neural membrane of the ANF (Rutherford et al., 2012). Hence, the difference in

ΔT between the two axons of the model may reflect the differences in dynamics of the voltage-gated Na+ ion channels in the peripheral and the central axons of the ANF. Indeed, significant differences in the distributions of voltage-gated ion channels along the ANFs have been reported (for review, see Davis and Crozier, 2016), but their roles in spike generation or conduction along the ANF are yet to be explored (Negm and Bruce, 2014; Boulet et al., 2016). Overall, the results of the present study suggest that while the two sites of spike generation are necessary to account for the differences in spike latencies, modeling the passive soma may not be necessary to capture the latency differences between the peripheral and the central axons.

3.5.3   Limitations of the Model

The proposed model captures the essential features of the spike generation without modeling the biophysical details regarding the mechanism of the action potential generation. While many phenomena are captured reasonably well, it is challenging to attribute the model parameters to biophysical elements, such as distributions of various voltage-gated ion channels. The model components are physiologically inspired. Nevertheless, a direct biophysical interpretation of the model parameters to the detailed ion channel dynamics is not possible. The insights gained from this approach will, however, be useful for more detailed modeling approaches to identify the corresponding biophysical elements.

The model was shown to correctly capture the spike latency differences between the two sites of spike generation as they are excited by different polarities. Although the model predicts correct trends of spike latency differences, the absolute spike latencies reported in Miller et al. (1999) were about 200 μs longer than those predicted by the model (Fig. 3e). This might be due to the recording site in the experiment: in the physiological experiments, the ANF spikes were recorded at the root of the AN, where the ANFs make innervations with the cochlear nucleus. Consequently, the spike latencies measured at that location also include the duration it takes for the spike to travel along the ANF. Although the model presented here aims to mimic the responsiveness of both the peripheral and central nodes

of the ANFs, it does not explicitly model the entire ANF through which a spike must travel. Since the offset is constant, the addition of a constant time delay of approximately 200 μs could be assumed to account for a complete travel time measured at the root of the ANF.

Although the model considers two sites of excitation, these point neurons are considered to be located in parallel rather than in series as they would be in the ANF. Since the two axons of the model independently integrate the stimulus charge, there is no interaction of the membrane voltage between the two axons. Instead, the interaction between the two axons is solely represented through the OR gate. Because of this architecture, the model only predicts the orthodromic spikes, i.e., the spikes traveling in afferent direction along the ANF. However, spikes generated along more central sites along the ANF can travel in efferent direction, known as the antidromic potential. Such an effect of the antidromic potential can affect the ECAP recordings (Miller et al. 2004) but has not been considered in the current framework. A development of alternative strategies to describe the interaction between the axons of the ANF is required to model an effect of antidromic potential on the ECAP responses.

While the model is mostly successful in predicting the effect of pulse rate on neural responses, the predictions of the temporal response properties deviate from the data at very high pulse rates, e.g., 10,000 pps. Measured ANF responses to high pulse rate stimulation have been characterized as dynamically unstable pattern, showing a strong onset followed by almost a shutdown of the neural responses (O'Gorman et al., 2010; O'Gorman et al., 2009). The model, in its current form, is unable to capture such effect.

The parameters of the model were chosen to account for single fiber responses and to describe the trends in the data. It does not account for the observed variability in the data. However, due to its computationally efficient integrate-and-fire-type structure, the model can easily be extended to represent a population of neurons by considering repeated simulations of the model for a desired number of neurons. This is an advantage over complex biophysical models, which are computationally expensive and hence limit the simulations to fewer neurons. In combination with a current spread model, the presented model can also be used to investigate individ-

ual factors leading to the variability in the predictions. Such an approach would be useful to study the effect of various parameters such as the spiking threshold (e.g., van den Honert and Stypulkowski, 1984; Miller et al., 1999) and refractoriness (e.g., Miller et al., 2001a) on population responses to dynamic stimuli, such as those reported by Matsuoka et al., 2000.

Finally, the model was developed based on responses recorded mainly from the ANFs of acutely deafened cats implanted with CI. It has been suggested that differences in sensitivity to anodic and cathodic currents may be dependent on the species under investigation. For example, recordings of the ANFs from guinea pigs show no significant differences between the thresholds for monophasic anodic and cathodic pulses. In contrast, in cats, the thresholds for anodic pulses are higher than for cathodic pulses (Miller et al., 1999; Miller et al., 1998) while in humans, the ANFs appear to be more sensitive to anodic than to cathodic pulses (Macherey et al., 2008). Despite the differences in sensitivity to anodic and cathodic currents, the differences in spike times between the two currents are consistent across species. Although the model presented here has been parameterized using the data for cats, it does not make explicit assumptions about differences in sensitivity to pulses of opposite currents. Hence, the model sensitivity to the current pulses will need to be modified by changing the membrane characteristics of the axons (and not the model structure), in order to account for the species specific differences in sensitivity.

3.6   Conclusion

A computational model of ANF responses to electrical stimulation was presented. The model was inspired by the observation that charges of either polarity can generate a spike, but with different spike latencies. Based on the spike latency data for electrical stimulation with monophasic pulses of cat ANF, it was assumed here that the site of spike generation for monophasic anodic pulses is closer to the brainstem (central axon) than the site of spike generation for a monophasic, cathodic pulse (peripheral axon). The model was parametrized using limited data

for monophasic stimulation, and model responses were compared to available data, mainly from electrically stimulated ANFs in cats. The model was shown to correctly account for stimulation with alternative pulse shapes as well as with dynamic stimuli such as pulse trains of different pulse rates and stimulus levels. Overall, the proposed model framework can serve as a useful tool to explore the temporal coding in the electrically stimulated ANF in response to various CI stimuli and to quantify potential benefits to the CI listeners.

Acknowledgments

4

Temporal coding in the electrically stimulated auditory nerve: The effect of uncertainty in the site of spike generation [c]

Abstract

Cochlear implants (CI) extract the slowly varying envelope of the input sound and stimulate the auditory nerve fibers (ANF) with amplitude-modulated trains of biphasic current pulses. Since envelope cues are the primary cues available to CI listeners, enhancing the fidelity of the envelope representation in the ANF responses has been a primary goal of various stimulation strategies. This study presents a model-based analysis of stimulus- and neural factors that affect the envelope representation in electrically stimulated ANFs. The model of Joshi et al., 2017 was used to predict modulation detection thresholds (MDTs) for various stimulus levels and carrier pulse rates in the range from 250 to 4000 pulses per second. The predicted MDTs were compared to behavioral data from CI listeners. It is shown that the model accounts for the data due to the assumption of a distributed spatial spike generation at peripheral and central axons of the ANF. With this configuration, the

---

reduced fidelity of the envelope representation in the ANF responses to high carrier rate stimulation results from the degraded temporal coding at the central axon. The model can be useful for evaluating the effects of various stimulation strategies on the envelope representation in the AN and for modulation perception.

Keywords

4.1 Introduction

Listeners with severe hearing loss or deafness are commonly prescribed with cochlear implants (CI). CIs bypass the impaired cochlear transduction process and electrically stimulate the auditory nerve fibers (ANFs) to restore the sense of hearing. Nevertheless, CI listeners typically experience challenges, in comparison to their normal hearing peers, in terms of speech understanding in background noise, pitch perception and sound source localization (Wilson and Dorman, 2008). CI processors primarily extract the slowly varying envelope of the acoustic signal and stimulate the ANF with a current pulse train, the carrier, which is amplitude modulated with the signal envelope. Normal-hearing listeners can achieve substantial speech understanding in quiet (Shannon et al., 1995) and in background noise, primarily using envelope cues (Jørgensen and Dau, 2011; Swaminathan and Heinz, 2012). Hence, the degraded perception of the CI listeners mainly demonstrates deficits in the encoding of envelope related information in the electrically stimulated ANFs. A clear understanding of the characteristics of envelope encoding in electrically stimulated ANFs is thus an essential prerequisite for the development of advanced CI stimulation strategies.

A characteristic feature of ANF responses to electrical stimulation is a lack of stochasticity in the spike times that results in a strongly phase-locked activity across the ANFs (Hartmann et al., 1984; Javel, 1990; Javel and Viemeister, 2000). In contrast, acoustically evoked ANF responses show a considerable amount of

stochasticity (Lowen and Teich, 1992). It has been argued that the reduced, or absent, stochasticity impairs the envelope representation in the electrically stimulated ANF (Kiang and Moxon, 1972; Morse and Evans, 1996; Rubinstein et al., 1999; Chatterjee and Robert, 2001). Various strategies have been proposed to introduce stochasticity in the ANF responses to electrical stimulation and, thus, to improve the fidelity of the envelope representation. In one strategy, it was proposed to use high carrier pulse rates, because spike-time distributions in response to pulse trains with a high pulse rate show higher stochasticity than those observed at low pulse rates (Rubinstein et al., 1999). Hence, high pulse-rate stimulation may also enhance the envelope representation in the electrically stimulated ANF. However, psychophysical modulation detection experiments (which are considered to measure the fidelity of the envelope information available to the listeners) showed that amplitude modulation detection thresholds (MDTs) increased (worsened) with increasing carrier pulse rate (Galvin and Fu, 2005; Galvin and Fu, 2009). Thus, this approach does not seem to improve the performance of the CI listeners in a basic temporal resolution task.

Another factor that can be expected to influence the representation of the envelope in electrically stimulated ANF responses is the spatial distribution of the sites of action potential (spike) generation. The current injected by the CI electrodes generates depolarizing and hyperpolarizing regions throughout the cochlea, enabling spike generation at peripheral as well as central axons of the ANF (Rubinstein, 1991). The differences in the site of spike generation result in differences in spike times that amount to about 200 μs in cats and possibly about 400 μs in humans (van den Honert and Stypulkowski, 1984; Miller et al., 1999; Shepherd and Javel, 1999; Rattay et al., 2001; Rattay et al., 2013; Undurraga et al., 2013). Responses to pulse train stimulation also exhibit bimodal spike-time distributions, which have been suggested to represent the uncertainty in the spike generation sites along the ANF (van den Honert and Stypulkowski, 1984). Furthermore, some experimental data indicated that the amount of uncertainty in terms of spike generation sites depends on the stimulation pulse rate (Javel and Shepherd, 2000). However, to what extent such uncertainty affects the envelope representation in the electrically stimulated ANF has not yet been examined.

The purpose of this study was twofold: (1) to quantify the effect of low and high carrier pulse rates on a) absolute detection threshold, b) dynamic range and c) envelope coding in electrically stimulated ANF responses and (2) to explicitly examine the effect of uncertainty in the site of spike generation on envelope encoding for low and high carrier pulse rates. Previous modelling studies have been unable to account for the observed effects of carrier pulse rate and stimulus level on psychophysical MDTs, since they do not account for the temporal response properties of the ANF responses to electrical stimulation (Goldwyn et al., 2010a; Goldwyn et al., 2012). In the present study, a recently proposed computational model of ANF responses to electrical stimulation was considered (Joshi et al., 2017) to predict MDTs using a neurometric analysis. This model has been shown to account for the physiological response statistics reported in the literature in various stimulation conditions, including stochasticity in spike times. The structure of the model also allows analyzing the effect of the uncertainty related to the site of spike generation on the predicted MDTs. Predicted neural MDTs were compared to available psychophysical data.

4.2   Methods

4.2.1   Stimulus Conditions

The carrier pulse trains consisted of cathodic leading, symmetric biphasic pulses with a duration of 40 μs/phase and an inter-phase gap of 8 μs. The stimulation pulse rates considered were 250, 500, 1000, 2000 or 4000 pulses per second (pps). The carrier pulse train was modulated using a sinusoidal amplitude modulator following Eq. 4.1:

$$I_{stim}(t) = \text{⊥⊥}\,[1 + m \cdot cos(2\pi f t)] \tag{4.1}$$

where I_{stim} represents the amplitude modulated pulse train, ⊥⊥ is the waveform of the unmodulated pulse train, m the modulation depth and f the modulation fre-

quency. Modulation depths of $m = 0$ (unmodulated), 0.007, 0.0125, 0.025, 0.05, 0.1, 0.2, 0.4, 0.8 and 1 (fully modulated) were used. The modulation frequencies were 16, 32, 64, 128, 256, 512 and 1024 Hz. For any given pulse train, only modulation frequencies below half the corresponding pulse rate of the carrier pulse train were used to avoid under-sampling of the modulation frequency.

The stimulus levels were chosen to facilitate a comparison of the predicted neural MDTs with behavioral MDTs in human CI listeners. Psychophysical studies have shown that MDTs are largely influenced by the overall loudness of the pulse trains and that MDTs obtained at different pulse rates should be compared at stimulus levels that produce the same loudness (McKay and Henshall, 2010). Due to the large variability of stimulus detectability (i.e., absolute detection threshold) and dynamic range of audible levels (i.e. difference in level at detection threshold vs maximum comfortable loudness) across human listeners, the MDTs were obtained at "normalized" stimulus levels in terms of percent dynamic range. These levels thus refer to stimulus levels above the detection threshold of a given individual listener. Since the neural code for loudness is unknown, previous modelling studies have assumed a spike-rate based criterion as a correlate of detection threshold and equal loudness across conditions (Bruce et al., 1999c). In the current study, stimulus levels were used that produced equal spike rates (of 1, 4, 8, 16, 32, 64, 96, 128, 160, 192 and 224 spikes/s) in the model across the different stimulation pulse rates (between 250 and 4000 pps) for the unmodulated pulse trains. The resulting stimulation levels across the pulse rates were accordingly represented as equal-spike rate contours. The detection threshold of a stimulus was defined in the model as the level that produced a spike rate of 1 spike/s. The dynamic range was defined as the difference between the threshold and the level required to produce a spike rate of 224 spikes/s. The detection threshold and the dynamic range were predicted as a function of the stimulation pulse rate and compared to measured behavioral detection thresholds and dynamic range of the CI listeners in the study of Kreft et al. (2004) and Zhou et al. (2015).

4.2.2   Auditory-Nerve Model

The computational model proposed by Joshi et al., 2017 was used to simulate the ANF responses. Briefly, the model includes two exponential integrate-and-fire neurons to simulate the responses of the peripheral and the central axons of the ANF. The peripheral axon of the model is depolarized by a cathodic charge and exhibits longer spike latencies than the central axon for an anodic charge. The central axon is depolarized by an anodic charge and exhibits shorter spike latencies than the spikes evoked on the peripheral axon with a cathodic charge. Both axons receive independent noise inputs to account for the stochastic nature of the ANF responses. Each of the two axons receives a continuous feedback current via two feedback loops (subthreshold and suprathreshold adaptation loops) which account for the phenomena of facilitation, accommodation, refractoriness and spike-rate adaptation. An interaction between the two axons is implemented using an 'OR' logic gate in the model. The axon that generates a spike first is fed through the OR gate and the complete model is set into an absolute refractory period during which no spike can occur. The model was parameterized using responses to monophasic stimulation and has been evaluated in stimulus conditions with various pulse shapes and dynamic pulse train stimuli. Besides providing spike times, the model also provides information if the spike was generated at the peripheral or the central axon of the simulated ANF. A detailed description of the model and its evaluation can be found in Joshi et al., 2017.

4.2.3   Analysis of Neural Responses

A neurometric threshold for the detection of amplitude modulations in the stimulus was determined using an analysis like that considered in Middlebrooks, 2008. The amount of synchronization of the neural response to the stimulus envelope was measured using the phase-projected vector strength (VS_{pp}). The VS_{pp} represents a normalized version of the vector strength (VS), and is calculated as the difference between the mean phase angle of spikes in each single trial and the mean phase angle across all trials for that condition. This normalization is particularly useful in cases where the number of spikes used to calculate the synchronization in each

trial is small, which can result in unrealistically high values of VS (Yin et al., 2011). VS_{pp} was calculated using Eq. 4.2:

$$VS_{pp} = cos(\varphi_t - \varphi_c)\sqrt{\frac{\left[\sum_{i=1}^{N} cos\left(\frac{2\pi t_i}{T}\right)\right]^2 + \left[\sum_{i=1}^{N} sin\left(\frac{2\pi t_i}{T}\right)\right]^2}{N}} \qquad (4.2)$$

where t_i represents the spike latency of the ith spike, T the period of the modulation frequency and N represents the total number of spikes in a trial $\{t_i\}$. φ_t represents the mean phase angle, in radians, for all the spikes in trial $\{t_i\}$. φ_c represents the mean phase angle, in radians, across all trials. The phase angles, φ_t and φ_c, of the spikes were calculated using Eq. 4.3:

$$\varphi = arctan2 \frac{\sum_{i=1} sin\left(\frac{2\pi t_i}{T}\right)}{\sum_{i=1} cos\left(\frac{2\pi t_i}{T}\right)} \qquad (4.3)$$

where t_i represents the time of the ith spike and T is the period of the modulation at a given modulation frequency. VS_{pp} is limited to values between -1 (when all the spikes in a trial occur 180 degrees out of phase from the global mean phase) and 1 (when all the spikes in a trial occur perfectly in phase with the global mean phase).

For each modulation depth, a receiver operating characteristic (ROC) was constructed determining the discriminability between the spike trains in response to unmodulated pulse trains (m = 0) and modulated pulse trains (m > 0) using the distribution of VS_{pp} obtained from multiple repetitions of the stimulus (Green and Swets, 1966). The area under the curve of the ROC then represents the percent correct discrimination performance for an unbiased observer (at a given modulation depth m). The collected values for all modulation depths were then used to construct the neurometric function. Neurometric functions were derived for all stimulus levels and fitted (using fmincon, MATLAB) with a logistic function described by:

$$y = a + \frac{b}{1 + e^{\frac{-(x-\mu)}{s}}} \tag{4.4}$$

with percent correct y, offset a, height b, and the mean of the modulation depths (μ) on the abscissa. The modulation depth corresponding to 79.7 percent correct discrimination was then considered as the discrimination threshold of the ANF.

The uncertainty related to the site of spike generation (between the two axons of the model) was quantified using entropy (e.g., Borst and Theunissen, 1999):

$$H(S) = -\sum_{i=1}^{2} p(s_i) \, log_2 \, p(s_i) \tag{4.5}$$

where $H(S)$ is the entropy (bits) related to the site of spike generation and $p(s_i)$ is the probability of a spike generated on axon i. Since the possible sites of spike generation are two (i = 1 or 2), the maximum entropy can be 1 bit.

4.3   Results

4.3.1   Effect of pulse rate on absolute threshold and dynamic range

Figure 4.1A shows predicted equal-rate contours, i.e. the stimulus levels that produce equal spiking rates (from 1 spike/s at the bottom to 224 spikes/s on the top) as a function of the stimulation pulse rate. In general, the stimulus level required to obtain a given spike rate decreases with increasing stimulation pulse rate. The difference in level required to produce a constant spike rate as a function of pulse rate is substantially larger (12 dB) for low spike rates (e.g. 1 spike/s) than for higher spike rates (e.g. 224 spikes/s; 1 dB). The stimulus levels that produced a spike rate of 1 spike/s were considered as the absolute detection threshold for the unmodulated pulse train stimuli. Figure 4.1B shows the effect of stimulation pulse rate on the absolute detection thresholds (normalized to the threshold for 250 pps).
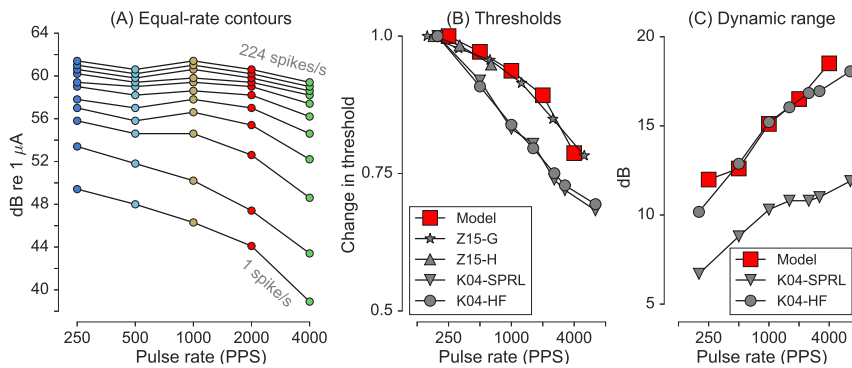
Figure 4.1: Effect of carrier pulse rate and stimulus level on the spike rate. A: Predicted equal-rate contours representing the stimulus levels (shown in dB re 1 μA) required to produce equal spike rate across carrier rates. The contours are shown for the spike rate from 1 spike/s (lowest contour) to 224 spikes/s (highest contour). B: Predicted effect of carrier pulse rate on detection threshold along with the corresponding behavioral data from Kreft et al., 2004 and Zhou et al., 2015. C: Predicted effect of carrier pulse rate on dynamic range along with the corresponding behavioral data from Kreft et al., 2004.

The predicted absolute thresholds are shown as red squares. The corresponding measured data reported in Zhou et al., 2015 and Kreft et al., 2004 are indicated by gray symbols. The star symbols (Z15-G) represent behavioral detection thresholds obtained in guinea pigs implanted with CIs. Upward triangles (Z15-H), circles (K04-HF) and downward triangles (K04-SPRL) represent behavioral detection thresholds obtained in three groups of human CI listeners. The data were normalized with respect to the lowest pulse rates used in those studies. Predicted and measured data show that the threshold decreases monotonically with increasing pulse rate. Fig. 4.1C shows the predicted dynamic range of the ANF (red squares), calculated as the difference in level between the bottom and the top equal-rate contours in Fig. 4.1A. This corresponds to a difference in levels at which the spike rate is 1 spike/s and 224 spikes/s, respectively. The corresponding data for the human CI listeners described in Kreft et al., 2004 are indicated by the gray symbols. Despite some differences between the predictions and the data, the model can capture the main finding that the dynamic level range increases with increasing stimulation pulse rate. Therefore, the spike-rate criterion used in the current study provides

a good approximation to study the effect of stimulation pulse rate on envelope encoding in the electrically stimulated ANFs.

4.3.2   Effect of pulse rate on MDTs
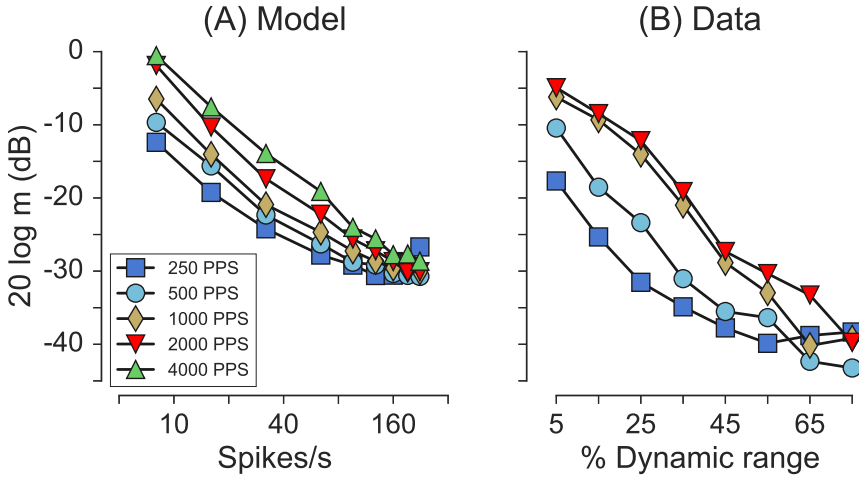
Figure 4.2: Effect of stimulus level and the carrier pulse rate on MDTs. A: The predicted MDTs are indicated as a function of spike rate. B: The corresponding behavioral data from Galvin and Fu, 2009 indicated as a function of percent dynamic range of the CI listeners.

Figure 4.2A shows predicted MDTs for a modulation frequency of 16 Hz as a function of the stimulation level as represented by the spike rate. Results are shown for carrier pulse rates of 250 pps (squares), 500 pps (circles), 1000 pps (diamonds), 2000 pps (downward triangles) and 4000 pps (upward triangles). For all pulse rates, the MDTs decrease substantially with increasing spike rate (by 20-30 dB between 4 and 224 spikes/s). MDTs are generally higher at high pulse rates than at low pulse rates. For example, at the lowest stimulus level (corresponding to 4 spikes/s), the predicted MDT is about 10 dB lower for 250 than for 2000 pps. Fig. 4.2B shows measured MDTs in human CI listeners (Galvin and Fu, 2009) for a modulation frequency of 20 Hz. Here, the stimulus level is represented as the percentage of the dynamic range above the absolute detection threshold (see methods). Similar

trends as in the predictions (Fig. 4.2A) can be observed in the psychophysical data, i.e. the measured MDTs decrease with increasing stimulus level and are lower at low carrier rates than at high carrier rate.

4.3.3   Effect of modulation frequency on MDTs

Figure 4.3 shows predicted temporal modulation transfer functions (TMTF), i.e. MDTs as a function of modulation frequency, for carrier pulse rates of 250 pps (Fig. 4.3A), 500 pps (Fig. 4.3B), 1000 pps (Fig. 4.3C), 2000 pps (Fig. 4.3D) and 4000 pps (Fig. 4.3E). In each panel, the TMTFs are indicated for four stimulus levels (corresponding to spike rates of 8 spikes/s indicated with squares; 32 spikes/s indicated with circles; 96 spikes/s indicated with diamonds; and 192 spikes/s indicated with downward triangles). In general, the predicted MDTs were almost independent of the modulation frequency for the carrier pulse rates from 250 to 2000 pps (Fig. 4.3A-D). For 4000 pps (Fig. 4.3E), the MDTs obtained for the highest modulation frequency (1024 Hz) were notably higher than those obtained for modulation frequencies below 500 Hz. For all carrier rates, the MDTs were typically highest for the lowest stimulus levels and decreased monotonically with increasing level. In these conditions, no corresponding data were available for comparison.

4.4   Discussion

4.4.1   The effect of multiple sites of spike generation on temporal coding

The essential features of the model used in the current study are the presence of two sites of spike generation and the ability to generate spikes in response to either (i.e. anodic or cathodic) phase of biphasic pulses. The analysis of the uncertainty in the site of spike generation showed that it affects the predicted MDTs. Figure 4.4 shows the relationship between the predicted MDTs and the average entropy (averaged across modulation depths between m = 0.007 and 1) for the carrier rates of 250 pps
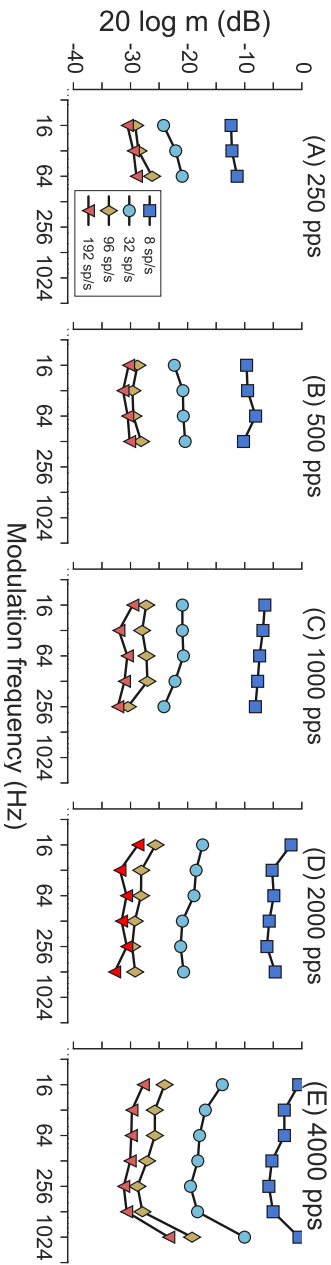
Figure 4.3: Effect of modulation frequency and stimulus level on predicted MDTs. The predicted MDTs are shown as a function of modulation frequencies for the carrier pulse rates from 250 pps (A) to 4000 pps (E) for four stimulus levels corresponding to 8, 32, 96 and 192 spikes/s.

(Fig. 4.4A), 500 pps (Fig. 4.4B), 1000 pps (Fig. 4.4C), 2000 pps (Fig. 4.4D) and 4000 pps (Fig. 4.4E). For this analysis, all predicted MDTs obtained across stimulus levels (4 - 224 spikes/s) and modulation frequencies were considered. In each panel, each dot represents the predicted MDT (on the abscissa) and the average entropy in that condition (on the ordinate). The different symbols represent the MDTs obtained for different modulation frequencies. The line through the data represents a least-squared regression while the shaded area shows 95 % confidence intervals. A significant correlation between MDT and entropy was found for all stimulation pulse rates, confirming that the predicted MDT co-varies with the amount of uncertainty in the site of spike generation across all modulation frequencies and stimulus levels. The correlation was, however, lower for pulse rates of 2000 and 4000 pps than for lower pulse rates.

4.4.2 Effect of pulse rate on MDTs

To further investigate the relation between the uncertainty in the site of spike generation and the predicted MDT, the response behavior of the model was analyzed in terms of peri-stimulus time histograms (PSTHs; 1 ms bin-width), whereby the spikes generated at the peripheral and the central axon were isolated in the model. Figure 4.5 shows normalized PSTHs (such that the maximum rate equals to 1), generated using the model responses for the carrier pulse rates of 250 pps (Fig. 4.5A), 1000 pps (Fig. 4.5B) and 4000 pps (Fig. 4.5C), modulated at a rate of 16 Hz. In each panel, the top row shows the results for the unmodulated (m = 0) stimuli and the bottom row shows the results for the fully modulated (m = 1) stimuli. The blue functions represent the PSTHs for the spikes generated at the peripheral axon whereas the green functions show the PSTHs for the spikes generated at the central axon. The results show that, as the modulation depth increases the responses of the peripheral axon (blue) follow the envelope modulation of the stimulus for all carrier rates. In contrast, the PSTHs of the spikes generated by the central axon (green) show more random fluctuations for the carrier pulse rates of 250 and 1000 pps at all modulation depths, whereas for the 4000-pps carrier rate the PSTHs of both the central and the peripheral axon follow the modulations only at large mod-
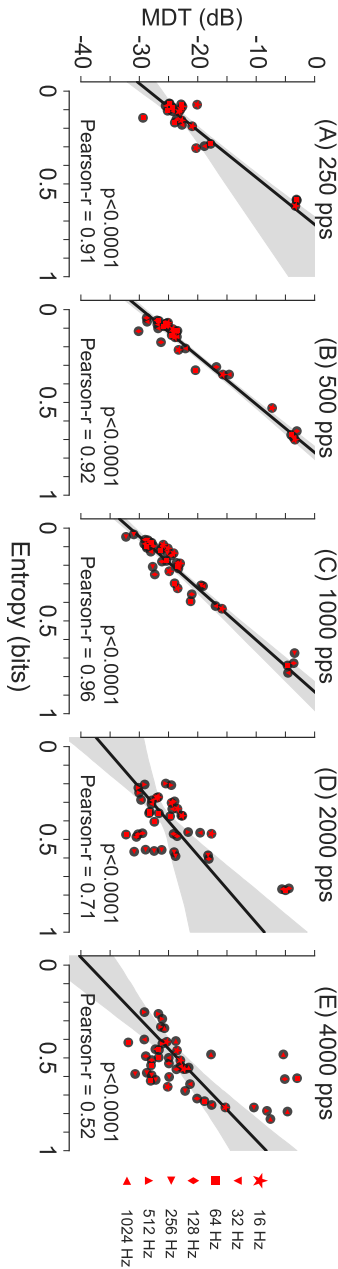
Figure 4.4: Predicted MDTs correlate with the entropy in terms of the uncertainty in the site of spike generation for the carrier pulse rates from 250 pps (A) to 4000 pps (E). The modulation frequencies are indicated with different symbols and the significance of the regression fits (in terms of p-value), Pearson correlation and the 95 % confidence intervals are also indicated.

ulation depths (above m = 0.4). Thus, the simulations suggest that the uncertainty in the site of spike generation adds random fluctuations in the temporal spike pattern, degrading the accuracy of the temporal envelope encoding in the ANF responses.

Additional simulations were run considering only the peripheral or only the central axon in the model. This operation thus removed any effect of uncertainty between the two possible sites of spike generation. Using these 'independent' axons, MDTs were predicted for the modulation frequency of 16 Hz using the same procedure as considered for the 'full' model. Figure 4.6 shows the predicted MDTs for the full model (Fig. 4.6A, replot from Fig. 4.2A), the peripheral axon only (Fig. 4.6B) and the central axon only (Fig. 4.6C). In each panel, the MDTs are shown for the carrier rate of 250 pps (squares), 500 pps (circles), 1000 pps (diamonds), 2000 pps (downward triangles) and 4000 pps (upward triangles). Compared to the predictions with the full model, the results obtained with the model considering only the peripheral axon (Fig. 4.6B) show a similar effect of spike rate but almost no effect of the carrier pulse rate. The predictions obtained when considering only the central axon of the model (Fig. 4.6C) showed the lowest MDTs for the highest carrier rate (4 kHz) whereas the full model and the data generally showed the lowest MDTs for the smallest carrier rate (250 Hz). Furthermore, the difference in MDTs obtained for the highest and lowest carrier rates increased with increasing spike rate, in contrast to the full model predictions and the data where the MDT difference decreased with increasing spike rate.

These simulations suggest that the neural representation of the stimulus envelope would be more accurate if the spikes were generated only at one axon. It may be possible to enhance the envelope coding in electrically stimulated ANFs by developing alternative pulse shapes that generate spikes at only peripheral or central axon of the ANF. Previously, alternative pulse shapes in CI stimulation have been demonstrated to affect thresholds, dynamic range, loudness growth and pitch perception in CI listeners (see van Wieringen et al., 2008; Chua et al., 2011; Macherey et al., 2011; Carlyon et al., 2013; Carlyon et al., 2015). However, their effect on the envelope representation in terms of MDTs has not been tested. The presented model may provide a framework to help develop alternative pulse
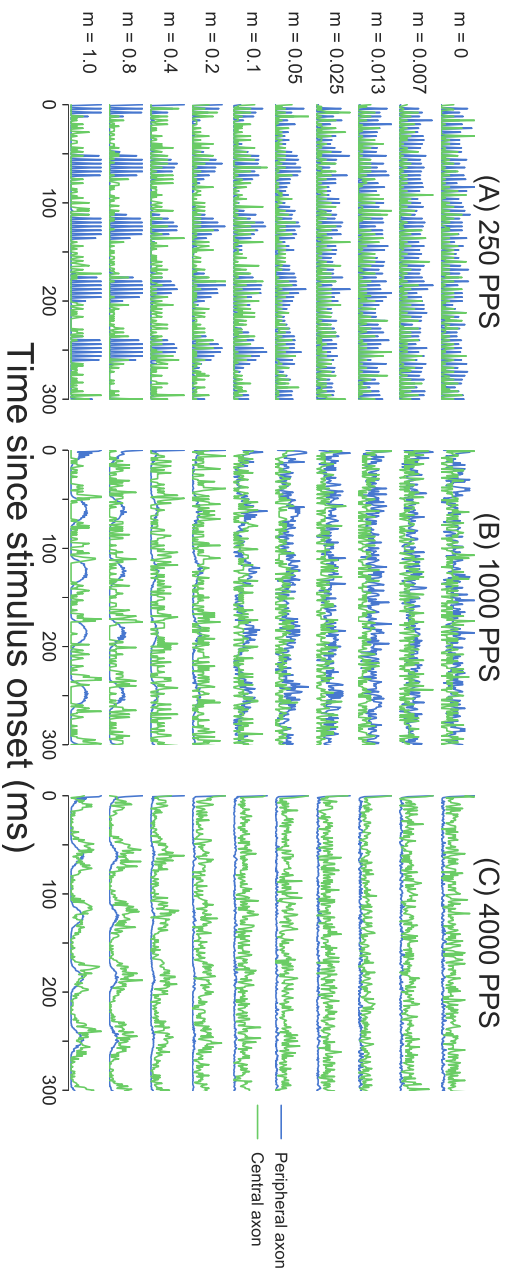
Figure 4.5: Effect of carrier pulse rate and modulation depth observed in the PSTHs. A normalized PSTHs constructed using only spikes generated by the peripheral axon (blue functions) or the central axon (green functions) are shown for the carrier pulse rates of 250 pps (A), 1000 pps (B) and 4000 pps (C).
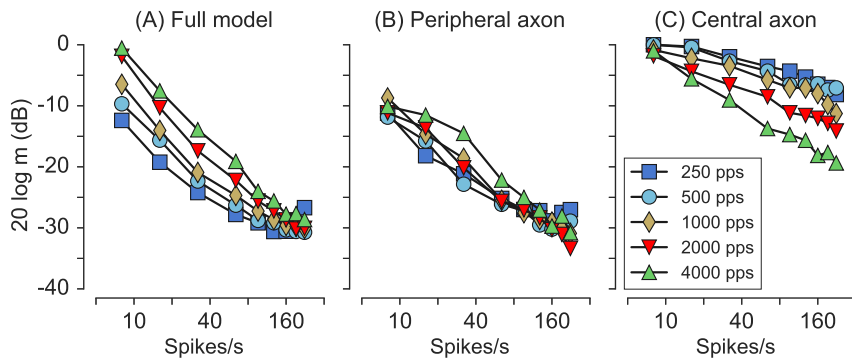
Figure 4.6: Effect of modulation frequency and stimulus level (in terms of spike rate) on predicted MDTs for three versions of the model. A: The predicted MDTs from the full model are replotted from Fig. 2A. B: The predicted MDTs considering only the peripheral axon of the model. C: The predicted MDTs considering only the central axon of the model.

shapes that reduce the uncertainty in terms of site of spike generation and hence enhance the representation of the stimulus envelope in the CI listeners.

4.4.3   Shape of the TMTF

Acoustically stimulated ANFs demonstrate a characteristic 'low-pass' shape of the TMTFs, irrespective of the stimulus carrier frequency (Joris and Yin, 1992). At low carrier frequencies (< 2000 Hz), the cut-off frequency of the TMTFs corresponds to about one third of the carrier frequency. At higher carrier frequencies (> 2000 Hz), the cut-off frequency has been found to be about 1000 Hz, irrespective of the carrier frequency, reflecting the inability of the ANF to follow envelope fluctuations above 1000 Hz. In contrast, the predicted TMTFs in the present study did not show any effect of carrier pulse rate on the MDTs.

Psychophysical studies have shown that CI listeners are unable to follow envelope modulations beyond frequencies of about 300 Hz (Shannon, 1992) and, although the absolute MDTs change with carrier pulse rate, the cut-off frequency of the TMTFs does not differ across carrier rates (Fraser and McKay, 2012). Similar cut-off frequencies have been observed in normal-hearing listeners (Bacon and

Viemeister, 1985; Ewert and Dau, 2000) and have been attributed to the limits of neural processing of envelope fluctuations in the brainstem (Joris et al., 2004). Such higher-level processes of envelope coding have not been considered in the framework of the present model that is focused on studying the limitations caused by the neural factors related to ANF processing. If the simulations are accurate, it suggests that the 'low-pass' characteristic observed in the TMTFs indeed stem from the limitations in processing beyond the ANF (e.g. cochlear nucleus or inferior colliculus). To clarify this point, an analysis of ANF recordings analogue to the analysis performed on the simulation results would be necessary.

4.4.4   Implications for current-steering strategies

The reduced spectral resolution at the electrode-neuron interface has been identified as a factor that reduces the ability of the CI listeners to utilize speech-related envelope cues (Oxenham and Kreft, 2014). Therefore, various studies have focused on developing current-steering methods to reduce the current spread in the cochlea (for review, see Bierer, 2010). Current steering methods provide currents of opposite sign to neighboring electrodes. In these methods, it has been assumed that only polarity of the stimulation pulse is excitatory and produces spikes. However, since both polarities can generate spikes, such current-steering strategy may introduce an additional uncertainty in terms of the spatial distribution in the site of spike generation that, in turn, may degrade the envelope encoding. It seems thus crucial to explore the effects of various current-steering strategies in relation to the spatial distribution of spike generation and to design current steering strategies that enhance the envelope representation of the stimuli in the electrically stimulate AN.

4.5   Summary and conclusion

A computational model of ANF responses to electrical stimulation (Joshi et al., 2017) was used to examine the effect of carrier pulse rate on the fidelity of the envelope representation of the stimuli in the ANF responses. The envelope representations

were quantified using a neurometric analysis to predict the MDTs for various carrier pulse trains. The predicted MDTs were compared to behavioral CI data from the literature. It was shown that the model successfully accounts for the effect of the carrier pulse rate on the MDTs because of the uncertainty in terms of the site of spike generation assumed in the model. Thus, in the framework of the model, the reduced fidelity of the envelope representation in the ANF responses was shown to result from the degraded temporal coding in the central axon. The model provides a useful framework for developing stimulation strategies that enhance the fidelity of the envelope representation in the electrically stimulated auditory nerve.

Acknowledgement

5

Cross-correlation model of interaural time difference coding in listeners with bilateral cochlear implants [d]

Abstract

A phenomenological model of interaural time difference (ITD) coding in bilateral cochlear implant listeners is presented. The model performs a coincidence detection analysis between the spike-train inputs from the two ears and based on this estimates the lateralization of the binaural input signal relative to the midline. It consists of three parts: A computational model of the auditory-nerve fiber (ANF) responses to electrical stimulation (Joshi et al., 2017); a binaural processing stage performing a neural cross-correlation analysis (Joris, 2003); and the decision stage. The model was used to predict the effects of stimulation level, stimulation pulse rate, modulation frequency and jitter of inter-pulse interval on ITD discrimination thresholds in bilateral cochlear implant listeners. The predictions were compared to the psychophysical data from various studies. This simple scheme based on cross-correlation analysis can successfully account for the effect of various stimulus parameters on the ITD thresholds observed in various studies. This model can serve as a useful tool to study the effect of

---

various stimulation strategies on ITD sensitivity in bilateral cochlear implant listeners.

5.1   Introduction

The ability to localize sound sources is crucial for successful communication and navigation in realistic acoustic environments. The auditory system exploits the differences in time at which the sound arrives at the two ears (interaural time difference, ITD) along with the differences in sound level between the two ears (interaural level difference, ILD) to identify the location of the sound source. Hence, sound source localization primarily depends on (1) access to binaural input and (2) the ability of both ears to accurately encode the spatial information contained in the ITDs and ILDs. Listeners with a hearing loss in one or both ears often show impaired localization performance, hence a hearing loss may impair the ability to encode ITDs and ILDs. Listeners with severe hearing loss or deafness in both ears are commonly prescribed with bilateral cochlear implants (BiCI). However, despite restoring the access to binaural input with BiCIs, listeners typically show a severely impaired performance in tasks related to sound source localization, which has been mainly attributed to their inability to process ITDs (Litovsky et al., 2006; Litovsky, 2016; Litovsky et al., 2010; Seeber and Fastl, 2008; van Hoesel, 2011; van Hoesel, 2012; Laback et al., 2015).

Cochlear Implants (CIs) bypass the impaired mechano-electrical transduction pathway through the cochlea and stimulate the auditory-nerve fibers (ANF) with electrical current pulses. Primarily, CI signal processing strategies extract the slowly varying envelope of the acoustic signal and stimulate the ANF with a carrier pulse train that is modulated with the extracted envelope. In normal hearing (NH) listeners, it has been shown that envelope cues play a crucial role for speech intelligibility, whereas the rapidly varying fine structure delivers ITD cues for the localization of a sound source (Shannon et al., 1995; Smith et al., 2002).

The impaired sound localization performance in BiCI listeners may be caused by the failure of the carrier pulse trains to deliver ITD cues. This idea is supported by psychophysical experiments which have demonstrated differences in just notice-able difference of ITDs (ITD-JND) between NH and BiCI listeners in conditions when the ITD was applied to the carrier. For example, NH listeners typically show lowest ITD-JNDs at about 900 Hz and increased ITD-JNDs for frequencies below and above 900 Hz (Brughera et al., 2013). In contrast, ITD-JNDs of BiCI listeners are lowest and constant for carrier pulse rates up to 400 pulses per second (pps) and increase substantially with increasing carrier rate Kan and Litovsky, 2015; Laback et al., 2015). However, when ITDs were applied to the envelope, BiCI listeners show ITD-JNDs that are similar to those of the NH listeners (Noel and Eddington, 2013). These results suggest that the carrier pulse trains used in CI stimulation do not provide the ITD cues required for robust source localization.

Physiologically inspired models of ITD coding can provide a useful framework for studying the limiting factors underlying the perceptual deficits in BiCI listeners. Previously, the modelling efforts related to the encoding of ITDs in BiCI listeners have mainly focused on biophysical properties of neurons in the medial superior olive (MSO; Colburn et al., 2009; Chung et al., 2015; for review, see Dietz, 2016). The MSO neurons receive monaural inputs from both ears and "calculate" the ITD by coincidence detection of the input spikes (Joris and Yin, 2007; Grothe et al., 2010; Franken et al., 2015). Although the exact biophysical mechanism that enables the MSO neurons to calculate the coincidences between the inputs from the two ears is yet unknown, it has been shown that the ITD sensitivity of the MSO neurons can be accounted for using a cross-correlation analysis when accurate descriptions of the monaural ANF responses are available (van der Heijden et al., 2011; Plauška et al., 2016).

The goal of the present study was to present a physiologically inspired phe-nomenological model of ITD coding in BiCI listeners. For this purpose, a cross-correlation based phenomenological model of ITD coding in BiCI listeners was developed. It combines a recently proposed model of the electrically stimulated ANF (Joshi et al., 2017) with a cross-correlation model of the binaural processor (Colburn, 1973; Stern et al., 1988; Shackleton et al., 1992). The model was evaluated

in conditions with varying stimulation level, carrier pulse rate, envelope shape, modulation frequency and IPI jitter, whereby the predictions were compared to psychophysical data from several studies.
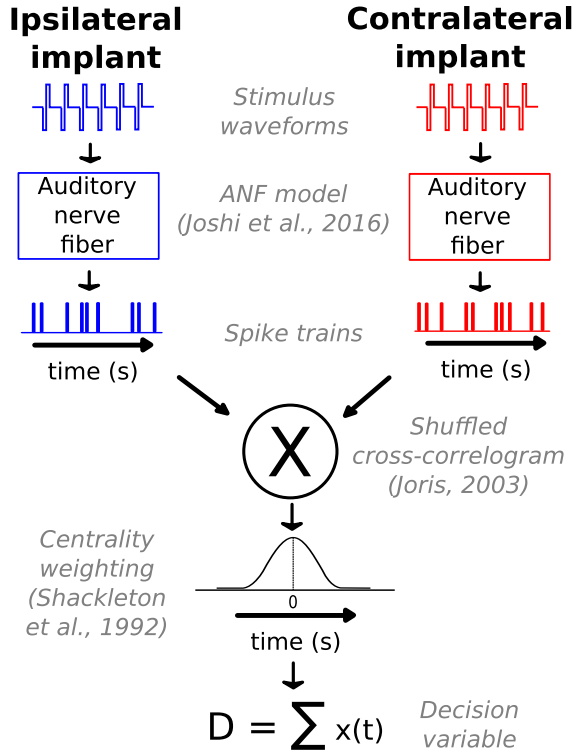
Figure 5.1: The structure of the cross-correlation model of ITD coding is shown. The model consists of a model of electrically stimulated auditory nerve fiber responses, followed by cross-correlation stage that is realized with a neural shuffled cross-correlogram analysis. The shuffled cross-correlogram is then passed through a centrality weighting function described with a Gaussian function with mean zero and a standard deviation of 600 μs. The decision variable, D is calculated by integrating the weighted shuffled cross-correlogram.

5.2   The Model

The structure of the model is shown in Fig. 5.1. It is a combination of a model of
the electrically stimulated ANF (Joshi et al., 2017) and a binaural stage inspired
by previously proposed cross-correlation models of ITD coding (Colburn, 1973;
Colburn, 1977; Colburn and Latimer, 1978). The ANF responses from both ears for
pulse train stimulation serves as an input into the binaural stage, which calculates
a neural cross-correlation between the two spike trains. The output of the binaural
stage is then passed on to a decision stage where the neural cross-correlation is
converted into a decision variable, D, for neurometric analysis.

The monaural ANF responses were simulated using a recently proposed com-
putational model of ANF responses to electrical stimulation (Joshi et al., 2017).
The model was developed with two main assumptions: (1) the ANF can gener-
ate spikes in response to currents of both phases (i.e. anodic and cathodic) of
biphasic pulses and (2) a cathodic charge generates a spike along the peripheral
axon while an anodic charge generates a spike along the central axon of the ANF.
The differences in the site of spike generation affect the spike timing, such that
the spikes generated at the central axon exhibit shorter spike latencies than those
generated at the peripheral axon. The model can account for ANF responses to
single pulses of various pulse shapes as well as paired pulses and pulse train, and
therefore can be generalized to study the effect of various stimulation strategies on
the ANF responses and ITD.

The coincidence detection analysis was implemented using the shuffled cross-
correlogram (Joris, 2003; Louage et al., 2004; Joris et al., 2006; Heinz and Swami-
nathan, 2009), similar to the procedure used in van der Heijden et al., 2011. The
resulting SCC functions were multiplied with a centrality weighting function rep-
resented as a Gaussian distribution with mean zero and a standard deviation of
600 μs, as in Shackleton et al., 1992. A multiplication of the SCC with the centrality
weighting emphasized the correlogram at the delays close to zero, enhancing the
importance of ITDs closer to the midline. Finally, the decision variable, D, was
calculated by integrating the weighted SCC function.

5.3   Methods

5.3.1   Stimuli
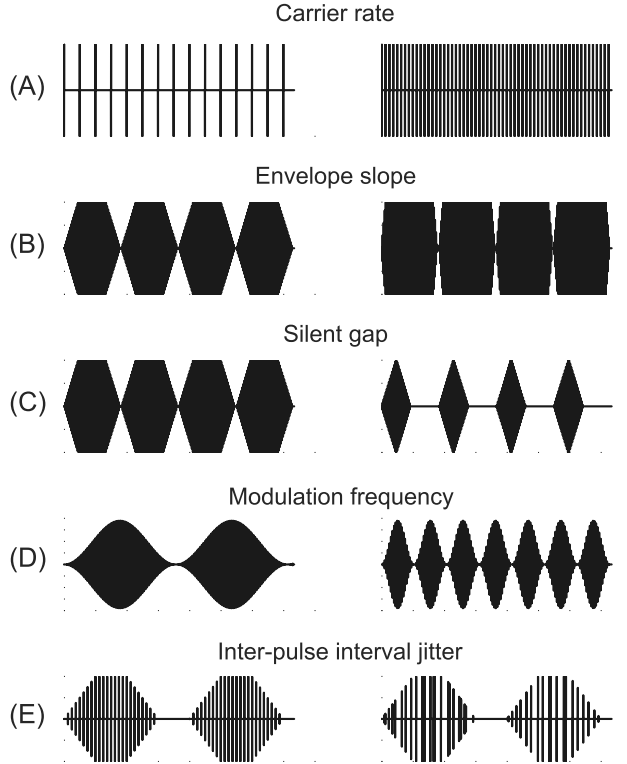
Figure 5.2: Illustration of stimuli used in the current study, showing the variation in carrier rate **(A)**, variation in envelope defined using envelope slope **(B)** or duration of the silent gap between the envelope bursts **(C)**, variation in sinusoidal modulation frequency **(D)** and jitter of the inter-pulse-intervals **(E)**

The considered stimulus conditions are illustrated in Fig. 5.2. Effects of carrier pulse rate (Fig. 5.2A), envelope slope (Fig. 5.2B), silent gaps in the envelope (Fig. 5.2C), sinusoidal modulation frequency (Fig. 5.2D) and inter-pulse interval (IPI) jitter (Fig. 5.2E) on the predicted ITD-JNDs were examined. In all conditions, the

carrier pulse trains consisted of cathodic leading, symmetric biphasic pulses with 40 μs/phase and 8 μs inter-phase gap. The total stimulus duration of each pulse train was 1 s. The remaining stimulus parameters were chosen to be consistent with psychophysical studies to facilitate a comparison across the model predictions and the psychophysical data.

In condition 1, the effect of carrier pulse rate was investigated for rates between 100 pps and 4000 pps. In condition 2, the effect of the envelope shape on ITD-JNDs was investigated by varying the envelope slope or the silent gaps between the modulation periods. A carrier pulse train with a rate of 1500 pps was modulated using a trapezoidal modulator, described by the two parameters 'ramp-time' and 'off-time' (Laback et al., 2011). The ramp-time refers to the duration in which the stimulus level increases from 0 to the peak level (or vice versa); the off-time refers to the period at which there is a silent gap between the two modulations (stimulus amplitude $= 0$). The duration of the ramp-time was varied between 4.5 and 9 ms/ramp. The duration of the off-time was varied between 1 ms and 21 ms. The modulation period of the trapezoidal modulator was constant at 36.7 ms and was defined as the sum of the duration of the onset- and offset ramps, the off time and the period during which the stimulus level was highest.

In condition 3, the effect of a sinusoidal amplitude modulation on ITD was predicted for modulation frequencies between 4 Hz and 500 Hz. The modulations were applied using a cosine-raised modulator wave with a modulation depth of 1 onto a carrier pulse train of rate of 1000 pps and 4000 pps.

In condition 4, the effect of IPI jittering was evaluated. The stimulus conditions were identical to those described in Laback and Majdak, 2008. In this condition, the carrier pulse train was modulated with a trapezoidal envelope with a ramp time of 20 ms and an off time of 20 ms. The IPI of the carrier pulse train was jittered from trial-to-trial following a uniform distribution in the range from IPI(1-k) to IPI(1+k), with a time resolution of 2 μs. Six levels of IPI jitter (k = 0, k = 0.125, k = 0.25, k = 0.5, k = 0.75 and k = 0.9) were used in the simulations.

The stimulation levels were adjusted across carrier rates to produce equal spike rate (16, 48 and 96 spikes/s) in the model of the ANF to avoid effects of stimulus level and loudness of the carrier pulse rate on ITD perception (Kreft et al., 2004;

Egger et al., 2016). Such a strategy is consistent with physiological studies where a criterion spike rate was used to equalize the stimulation levels across carrier rates (Bruce et al., 1999c).

5.3.2   Shuffled cross-correlogram

The shuffled cross-correlograms (SCCs) were constructed by calculating all-order inter-spike interval histograms between the spike trains from the ipsilateral and the contralateral ANFs. Briefly, N spike trains were simulated by a repeated presentation of the same stimulus. These spike trains are referred to as the ipsilateral spike trains (h_{ipsi-i}) where subscript notation 'i' refers to i^{th} spike train. The contralateral spike trains ($h_{contra-i\Delta}$) were created by adding a delay (ΔT, ranged from 0 to 2500 μs) to h_{ipsi-i}. The ISIs were calculated for forward and backward intervals between all combinations of ipsilateral and contralateral spike trains. To avoid a correlation between the spike train and its own delayed version, the original spike train and its delayed version were excluded from constructing the SCC. A normalized histogram was constructed using a bin-width of 50-μs. The number of spike trains used for the construction of the SCCs were chosen such that a minimum of 2000 spikes were considered for both ipsilateral and contralateral spike trains (Heinz and Swaminathan, 2009). Figure 5.3A shows some examples of the SCC functions for a pulse train of 1000 pps with an imposed ITDs of 0 μs (blue line), 160 μs (green line) and 480 μs (red line). Corresponding centrality-weighted correlograms are shown in Fig. 5.3B. The resulting distributions of decision variables for the three ITDs based on 100 repetitions of the model simulations are shown in Fig. 5.3C for each of the three ITD values. The distributions of D show that the is highest for ITD = 0 μs and decreases with increasing the ITD.

5.3.3   Neurometric analysis

In each stimulus condition, the model simulations were conducted for ITD values ranging from 0 to 2500 μs (0, 10, 20, 40, 80, 160, 240, 320, 480, 640, 960, 1280, 1600, 2000 and 2500 μs). For each ITD value, the simulations were repeated 100 times to obtain a distribution of the decision variable D corresponding to each
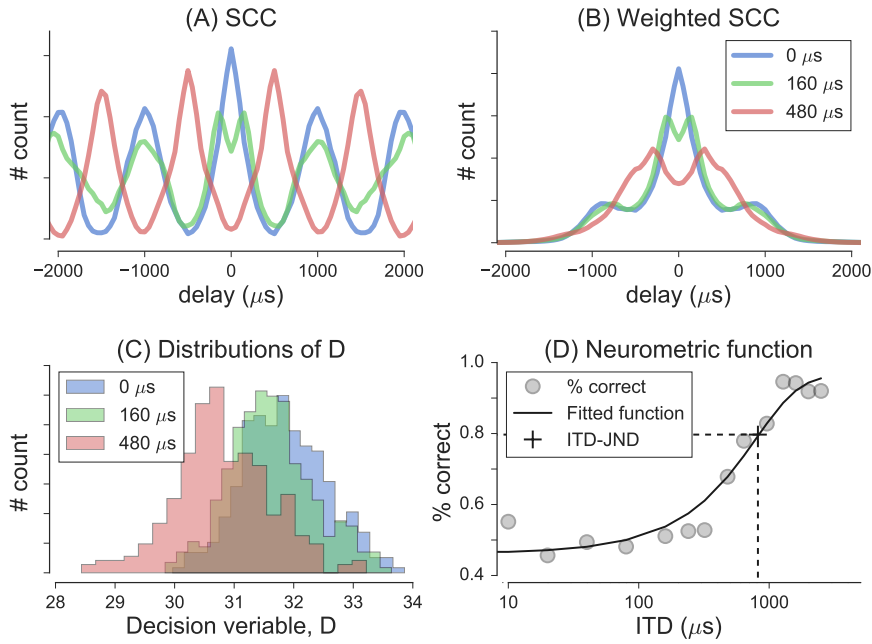
Figure 5.3: Examples of the SCC functions for a pulse train of 1000 pps with an imposed ITDs of 0 μs (blue line), 160 μs (green line) and 480 μs (red line) are shown in **A**. Corresponding centrality-weighted correlograms are shown in **B**. The resulting distributions of decision variables for the three ITDs based on 100 repetitions of the model simulations are shown in **C** for each of the three ITD values. The distributions of D show that the is highest for ITD = 0 μs and decreases with increasing the ITD. An example of the neurometric function fitted to the percent discrimination as a function of the value of imposed ITDs is shown in **D**.

ITD. A receiver operating characteristic (ROC) was constructed determining the discriminability between the distribution of D for ITD = 0 μs and ITD above zero μs (Green and Swets, 1966). The area under the curve (AUC) of the ROC curve then represents the percent correct discrimination performance for an unbiased observer. The collected AUC values for all ITDs were then used to construct the neurometric function, showing the discrimination performance as a function of ITD. The neurometric discrimination performance was fitted (fmincon, MATLAB) with a logistic function described by:

$$y = a + \frac{b}{1 + e^{\frac{-(x-\mu)}{s}}} \tag{5.1}$$

where y refers to the percent correct value, a represents an offset, b is the height, and μ is the mean on the abscissa (in this case, the ITD). The ITD corresponding to 79.7 percent correct discrimination was then considered to represent the predicted ITD-JND. An example of such a neurometric function fitted to the percent discrimination as a function of the ITDs is shown in Fig. 5.3D.
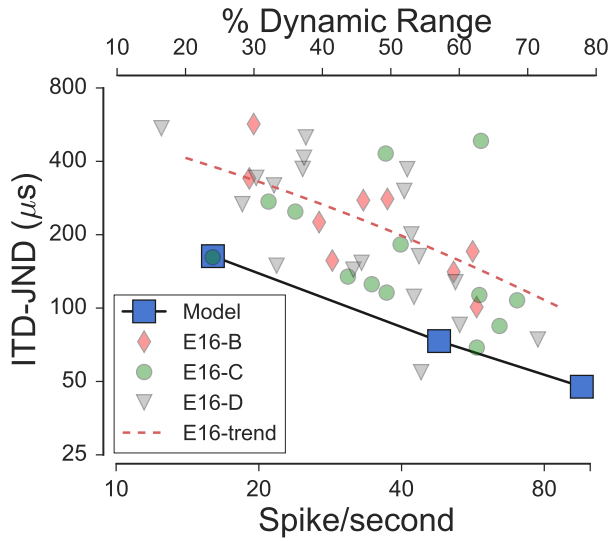
5.4   Results

Figure 5.4: Predicted ITD-JNDs (squares) as a function of spike-rate for 100 pps pulse train. The corresponding behavioral data from Egger et al., 2016 are indicated with different symbols as a function of percent dynamic range indicated on the top axis. The dashed line indicates a trend-line through the data.

5.4.1   Effect of stimulus level

Figure 5.4 shows the effect of stimulus level on the ITD-JND. The predicted thresholds (blue squares) are shown as a function of the spike rate for a carrier pulse train of 100 pps for three stimulus levels corresponding to spike rates of 16, 48 and 96 spikes/s, respectively. For comparison, psychophysical ITD-JNDs of BiCI listeners reported by Egger et al., 2016 are shown as a function of the dynamic range (in %). The behavioral thresholds are shown for three different electrode pairs denoted as E16-B (diamond), E16-C (circles) and E16-D (downward triangles). A least-squared regression line fitted to all the behavioral data is indicated by a dashed red line. The predictions and the data show that the ITD-JNDs decrease with increasing stimulus level. The experimental data shows a large variability, but the slope of the regression line fitted to the data and that of the predictions show a similar trend of the stimulus level on ITD-JNDs.

5.4.2   Effect of pulse rate

Figure 5.5 shows the effect of carrier pulse rate on the ITD-JND. The blue squares and the error bars show the mean and standard deviation of the predicted thresholds for the three stimulus levels considered (corresponding to 16, 48 and 96 spikes/s). For comparison, the psychophysical data from Laback et al., 2007(red circles), van Hoesel, 2007 (green downward triangles) and van Hoesel et al., 2009 (gray upward triangles) are shown. The predictions show a clear trend of increasing thresholds with increasing carrier rate. The threshold for the carrier pulse rates between 100 and 400 pps is about 100 μs and is increased substantially to about 500 μs for the carrier pulse rates above 1000 pps. The predicted thresholds did not change notably for carrier pulse rates above 1000 pps due to a spike-rate saturation effect at those pulse rates.

5.4.3   Effect of envelope shape

Figure 5.6 shows the effect of the envelope shape on the ITD-JND as a function of the off time. The results are shown for the ramp-times of 9 ms (blue squares),
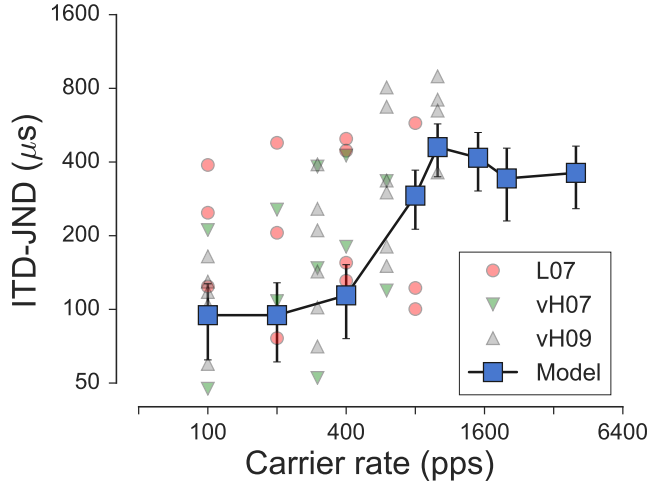
Figure 5.5: The predicted effect of carrier pulse rate on ITD-JNDs (blue squares), along with correspond-
ing behavioral data from Laback et al., 2007 and van Hoesel, 2007; van Hoesel et al., 2009.

6.8 ms (green circles), 5.4 ms (yellow downward triangles) and 4.5 ms (red upward
triangle). The mean of the predicted ITD-JNDs across the three stimulus levels
(corresponding to 16, 48 and 96 spikes/s) are shown in Fig. 5.6A. The corresponding
psychophysical data from Laback et al., 2011 are shown in Fig. 5.6B. Predictions
and data show that ITD-JNDs are highest for the stimuli without silent gap between
the modulator periods (off time = 0 ms) and decrease with increasing off time.
Similar trends are observed across the ramp-times. The largest reduction of the
ITD-JNDs (from about 800 to 300 μs) is observed between 0 and 15 ms long silent
gaps, whereas no further change occurs for gaps above 15 ms. In the data, the
ITD-JND is about 1400 μs for the smallest silent gap and reduced to the value up
to 500 μs for the silent gap durations above 15 ms.
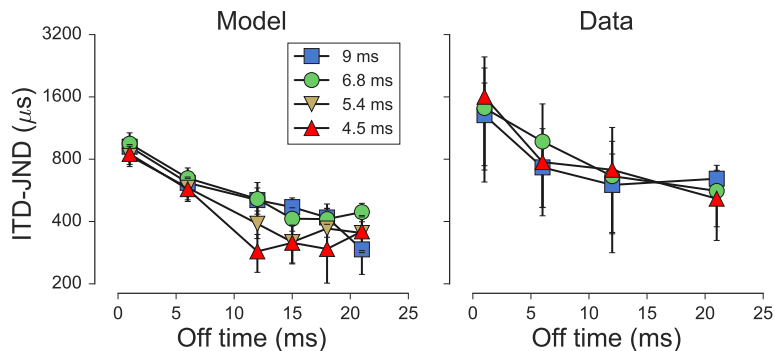
Figure 5.6: The predicted effect of envelope shape on ITD-JNDs. The ITD-JNDs for the envelope slope of 9 ms (squares), 6.8 ms (circles), 5.4 ms (downward triangles) and 4.5 ms (upward triangles) are indicated as a function of the silent gap duration. The corresponding behavioral data from Laback et al., 2011 is shown for a comparison.
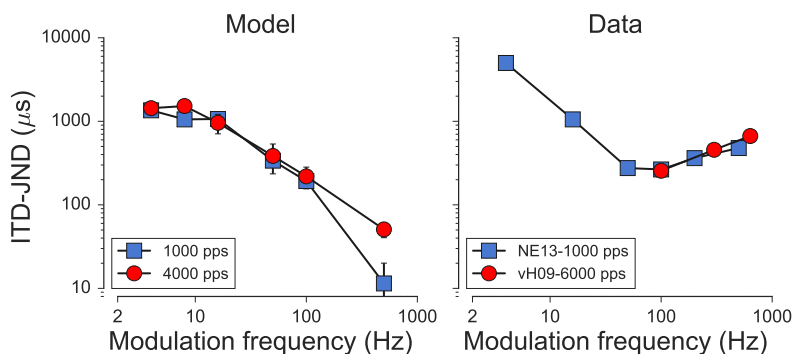
Figure 5.7: The predicted effect of modulation frequency on ITD-JNDs for carrier rates of 1000 pps (squares) and 4000 pps (circles). The corresponding behavioral data from Noel and Eddington, 2013 and van Hoesel et al., 2009 are shown for a comparison. The model can predict reasonably accurate trends for the modulation frequencies up to 100 Hz.

5.4.4 Effect of modulation frequency

Figure 5.7 shows the effect of the sinusoidal modulation frequency on the ITD-JNDs. The predicted thresholds for a carrier rate of 1000 pps (blue squares) and 4000 pps (red circles) are shown as a function of the modulation frequency (Fig.

5.7A). Figure 5.7B shows corresponding psychophysical ITD-JNDs from Noel and Eddington, 2013 for a carrier pulse rate of 1000 pps (blue squares) and van Hoesel et al., 2009 for a carrier pulse rate of 6000 pps (red circles). The predictions show that ITD-JNDs are highest for the low modulation frequencies and decrease with increasing modulation frequency. The psychophysical data show a similar trend for modulation frequencies up to 100 Hz. However, above 100 Hz, the behavioral thresholds tend to increase with increasing modulation frequency. A decrease in sensitivity towards higher envelope frequencies, as observed in the data likely results from the processing in the cochlear nucleus and trapezoid body, which has been shown to enhance the slow fluctuations while reducing faster envelope fluctuations (Joris et al., 2004).

5.4.5   Effect of inter-pulse interval jitter

Figure 5.8 shows the effect of IPI jitter on the ITD-JND as a function of the carrier pulse rate. Figure 5.8A shows the thresholds for k = 0 (unjittered stimuli; blue squares) as well as the averaged thresholds for k = 0.125, 0.25 and 0.5 (red squares) and k=0.75 and 0.9 (green triangles). The corresponding psychophysical data (calculated from the percent correct responses reported by Laback and Majdak, 2008) are shown in Fig. 5.8B. Predictions and data show that, for the unjittered pulse trains, the thresholds increases with increasing carrier rate. The introduction of IPI jitter gradually decreases the thresholds at high pulse rates. The reduction of the ITD-JNDs is largest for the largest values of k (0.75 and 0.9). At low carrier rates up to 400 pps, there is no effect of the IPI jitter on the ITD-JNDs. Beyond the range of the experimental data (i.e. above 1515 pps), the model predicts no further increase of the ITD-JND with increasing the carrier pulse rate.
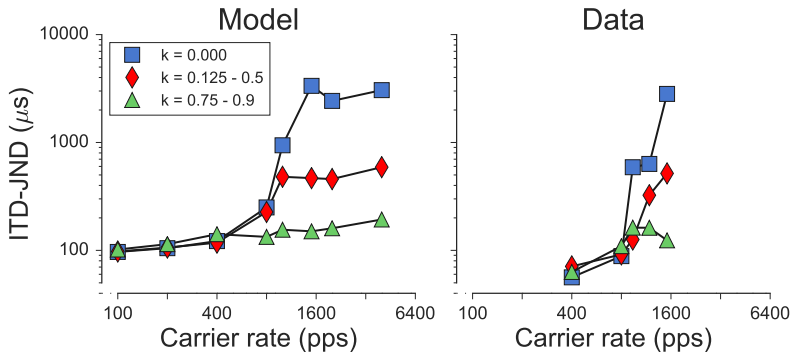
Figure 5.8: The predicted effect inter-pulse-interval jitter on the ITD-JNDs indicated as a function of the carrier rate along with corresponding data from Laback and Majdak, 2008.

5.5 Discussion

5.5.1 Effect of degraded temporal coding in the electrically stimulated ANFs on ITDs

The differences in ITD sensitivity between NH and BiCI listeners has been hypothesized to arise from two limitations. First, a degraded temporal coding in the electrically stimulated ANFs does not encode the information at all, or at least not in a manner suitable for the binaural neurons to calculate the ITDs. A direct comparison of the ANF responses to electrical and acoustical stimulation indeed shows some striking differences (Hartmann et al., 1984; Javel, 1990; Javel and Viemeister, 2000). Some recent modelling studies have begun to investigate the effects of such differences in ANF responses on the responses of the models of brainstem neurons. These studies have employed statistical descriptions of the ANF responses related to high values of phase locking (Colburn et al., 2009) or existing phenomenological models of the electrically stimulated ANFs (Chung et al., 2015) as input to a detailed biophysical model of the binaural neurons in the MSO. The model of the MSO neurons was developed using detailed biophysical recordings from animals with normal hearing and it was assumed that the response properties of the brainstem neurons remain unchanged in case of the electrically stimulated ANFs.

The results of these models predicted similar sensitivity to ITDs for the electrically as well as acoustically stimulated ANF and therefore were unable to account for the differences observed in ITD sensitivity between NH and BiCI listeners. However, the models of ANF responses used in those studies are unable to account for temporal response properties observed in physiological and psychophysical data (Goldwyn et al., 2010a; Joshi et al., 2014). In order to encode ITDs, a precision in spike-time statistics is considered crucial in order to explain the trends observed in the psychophysical data (Brand et al., 2002).

Alternatively, it has been hypothesized that a mismatch between the stimulation rate and the place of stimulation in the cochlea limits the ability of the binaural neurons to process that information (Chung et al., 2015). It is assumed that the binaural neurons are tuned to process frequency-specific information they receive from the cochlea due to frequency-dependent differences in their biophysical properties. Indeed, changes in parameters such as the synaptic input strength and membrane conductance of the biophysical model of MSO neurons has been shown to affect their sensitivity to ITD (Chung et al., 2015) and such changes were required to account for the effect of stimulation pulse rate on sensitivity to ITD. However, such a systematic tonotopic gradient of the biophysical membrane characteristics is observed mainly in the lateral superior olive but not in the MSO (Scott et al., 2005; Remme et al., 2014) and therefore this hypothesis is mainly speculative. In addition, the psychophysical data does not support the hypothesis, since the data show no differences in ITD-JNDs as a function of the stimulation place in the cochlea (for review, see Laback et al., 2015).

The model presented in the current study supports the first hypothesis that the differences in sensitivity to ITDs between the NH and BiCI listeners arise from a degraded temporal coding in the electrically stimulated ANFs. The model of ANF responses used in the current study has been shown to account for the key temporal aspects of the ANF responses observed in the data and therefore plays a crucial role in successfully predicting the effect of stimulation parameters on sensitivity to ITD. This model required no change in the parameters to account for psychophysical data from several studies. Hence, suggesting that the binaural processor does not require frequency dependent changes in the biophysical properties and therefore,

the binaural processor for the BiCI listeners can be modelled with parameters independent of the stimulation place in the cochlea.

5.5.2 The centrality weighting

In the current study, the binaural processor was modelled using a cross-correlation analysis between the spike trains arriving from two ears followed by the centrality weighting function that emphasized the importance of the ITDs close to midline. Specifically, the peak of the function used in the current study is at the delay of 0 and decays with increasing delays (Shackleton et al., 1992). The ITD sensitive neurons in the MSO of mammals show increased spiking activity for the binaural input with ITDs close to zero. In case of the NH animals, most MSO neurons respond maximally when the nonzero ITDs (McAlpine et al., 2001; Franken et al., 2015; van der Heijden et al., 2011). Therefore, previously proposed physiological models, which aimed to predict the responses of the MSO neurons in response to various ITDs for the acoustic input, have used weighting functions that emphasized the nonzero ITDs (Hancock and Delgutte, 2004; van der Heijden et al., 2011; Plauška et al., 2016). However, in contrast to the NH animals, the MSO neurons that receive an input from the electrically stimulated ANFs show maximal activity when the bilateral stimulus has an ITD = 0 μs and decreases with increasing the ITD (Smith and Delgutte, 2007). Therefore, the centrality weighting function used in the current study can successfully account for various stimulus parameters on the ITD-JNDs.

5.6 Conclusion

- A physiologically inspired phenomenological model of ITD coding in bilateral CI listeners is presented and the model predictions were compared with behavioral data from several studies.

- The model can successfully account for the effect of various stimulus parameters such as the stimulus level, carrier pulse rate, envelope shape, modulation

frequency and the inter-pulse-interval jitter on the ITD-JNDs observed in the data.

- In case of the amplitude modulated pulse train, the model can account for the effect of modulation frequencies up to 100 Hz. Beyond 100 Hz, the model predicts lowering in the ITD-JND whereas the data shows a slight increase.

- The model can be used to study the effect of various stimulus and auditory nerve fiber related parameters on the coding of neural coding and perception of ITDs.

Acknowledgement

6

Summary and Perspectives

This thesis describes the development and evaluation of a computational model of auditory nerve fiber (ANF) responses to electrical stimulation with cochlear implants (CI). The overall goal of this project was to develop a computational framework for an objective evaluation of CI stimulation strategies to identify the stimulus related and neural factors that affect the performance of CI listeners in various behavioral tasks. The development of the model was focused on predicting the temporal response statistics of the electrically stimulated ANFs. The model concepts were based on previously published measured single-neuron responses of cat ANFs. The simulated responses were evaluated in conditions with different pulse shapes, paired-pulses and pulse trains and compared to the data from several published studies. Furthermore, the model was used to predict the effect of various stimulation parameters on behavioral measures, such as amplitude modulation detection thresholds and thresholds for the detection of interaural time differences.

6.1   Summary of main results

In *chapter 2*, two state-of-the-art models of electrically stimulated ANFs (Goldwyn et al., 2012; Horne et al., 2016) were evaluated in terms of their predicted response statistics for stimulation with various pulse shapes. It was postulated that a reasonably accurate description of the effect of both anodic and cathodic charges on spike generation is required. The two models assume that a cathodic charge depolarizes the neural membrane and generates a spike. The point process model of Goldwyn et al., 2012 does not consider any effect of the anodic charge on spike generation whereas the leaky integrate-and-fire model of Horne et al., 2016

considers the anodic charge as inhibitory input. The predicted responses of the two models to single pulses of various pulse shapes were compared to physiological data from the literature. It was shown that the two models could accurately predict the ANF response statistics for stimulation with monophasic cathodic and cathodic-leading symmetric biphasic pulses. However, both models failed to account for the responses to different types of pulse shapes. Specifically, while the point process model was shown to account for the effect of the inter-phase gap duration, it failed to describe the effect of the duration of the anodic phase in pseudomonophasic pulses. The leaky integrate-and-fire model could account for the effect of duration of the anodic phase in pseudomonophasic pulses, but failed to account for the effect of the inter-phase gap duration. These limitations were attributed to the missing responsiveness of these models to the anodic charges. The evaluation results supported the hypothesis that considering the effect of both anodic and cathodic charges is necessary to account for the ANF's responses to electrical stimulation.

In *chapter 3*, a model of electrically stimulated ANF responses was presented. Following the limitations of the existing models as highlighted in chapter 2, the new model was based on two observations: (1) ANF can respond to both phases of the biphasic pulses; and (2) the spike latencies are longer for spikes generated in response to the cathodic pulses than for those generated in response to the anodic pulse. The model consisted of two exponential integrate-and-fire type neurons representing the peripheral and the central axons of the ANF. The peripheral axon was assumed to be excited by the cathodic phase and inhibited by the anodic phase. The central axon was assumed to be excited by the anodic phase and inhibited by the cathodic phase. The model parameters were set such that the spike latencies for the spikes generated at the central axon were shorter than those generated at the peripheral axon. Moreover, the membrane potential at each of the axon was assumed to be modified by two adaptive current feedback loops that describe subthreshold and suprathreshold adaptation processes. The interaction between the two axons of the model was implemented using an 'OR' logic gate that receives input from both the axons and delivers a single output spike. The membrane characteristics and other parameters in the models were chosen

to fit the physiological data recorded in cat's ANF for monophasic stimulation. The model was evaluated in stimulus conditions with single pulses of various pulse shapes, paired pulses and pulse train stimuli. The model predictions were compared with physiological data from several studies. It was shown that the model can provide a reasonably accurate description of the data in all considered conditions.

Chapter 4 presented an evaluation of the factors affecting the envelope representation in electrically stimulated ANF's responses. Specifically, the effect of stimulus pulse rate and uncertainty in terms of the site of spike generation on the representation of the envelope in ANF responses was examined. The model proposed in chapter 3 was used to simulate ANF responses to pulse trains of different pulse rates. Modulation detection thresholds (MDTs) were predicted using a neurometric analysis to quantify the discriminability between responses to unmodulated and sinusoidally amplitude modulated pulse trains. The predicted MDTs were compared to available psychophysical data from the literature. The proposed model could successfully account for the effect of stimulation pulse rate observed in the behavioral data. The uncertainty in terms of the site of spike generation was found to be correlated with the predicted MDTs in all considered conditions. Furthermore, when the uncertainty was "switched-off" in the model by considering only the peripheral or only the central axon, the model was unable to account for the effect of the carrier pulse rate on MDTs. A further analysis of the model responses showed that the representation of the envelope in the ANF responses was limited due to the degraded temporal coding in the spikes generated in the central axon due to the "raised noise floor" resulting from the spatial distribution of the sites where the spike generated.

In *chapter 5*, a phenomenological model of interaural time difference coding in listeners with bilateral CI (BiCI) was presented. This model consisted of three main components: The model of ANF responses to electrical stimulation presented in chapter 3, the binaural process performing a neural cross-correlation analysis and a decision device. A neurometric analysis was used to estimate the lateralized position of the stimulus in terms of its deviation from the midline. The model was then used to predict the effects of level, pulse rate, modulation frequency, envelope

shape and inter-pulse-interval jitter on interaural time difference (ITD) coding. The predictions were compared to behavioral data in the BiCI listeners from the literature. It was shown that the model could predict the major trends observed in the behavioral data. The results suggested that the limited abilities of BiCI listeners to utilize the ITD information for lateralization may be mainly due to the degraded temporal coding at the level of the ANF.

6.2   Perspectives

The new model proposed in chapter 3 was developed based on the observations of cat ANF responses to electrical stimulation. Specifically, the model parameters were extracted from the response statistics of cat ANFs to monophasic stimulation. However, there are notable morphological differences between cat and human ANFs. Specifically, the peripheral axon of human ANF is notably longer than that of cat ANF. In addition, the soma of the human ANF is unmyelinated and thus has a higher capacitance than the soma of the cat ANFs (Ota and Kimura, 1980; Liberman and Oliver, 1984; Spoendlin and Schrott, 1989; Rattay et al., 2013). These differences should affect some of the response properties of the ANFs (Rattay et al., 2001; Rattay et al., 2013). For example, in cats, the spiking threshold for stimulation with monophasic cathodic pulses is lower than for stimulation with monophasic anodic pulses. However, in humans, the spiking thresholds for stimulation with monophasic anodic pulses can be assumed to be lower than for cathodic pulses (Rattay et al., 2001; Rattay et al., 2013). Although monophasic stimulation has not been applied in human CI listeners due to the charge-balancing requirement in the clinical CI devices, some investigators have used alternative pulse shapes in their experiments in the CI listeners. The results of such experiments have indeed suggested that the human ANFs are more sensitive to anodic charge than to cathodic charge (Macherey et al., 2008). Furthermore, in human ANFs, the difference in spike latencies for the spikes generated at the central and the peripheral axons have been estimated to about 400 µs, which is twice as much as for cats (Rattay et al., 2001; Rattay et al., 2013; Undurraga et al., 2013). Despite these differences between

cat and human ANFs, the proposed model was applied to predict the measures of temporal coding (e.g. modulation detection thresholds and ITD-discrimination thresholds) in humans. A comparison of the model predictions with behavioral data from the literature shows that the model was able to describe the effect of various stimulus parameters on those measures, such as pulse rate and stimulus level, supporting the hypothesis that the uncertainty in terms of sites of spike generation strongly contributes to the temporal response patterns in human CI listeners. Future work will need to further evaluate the model by investigating methods to 'humanize' the model parameters.

The proposed model was designed based on physiological findings. Nevertheless, it bypassed detailed ion-channel dynamics by modelling only the "effective" changes affecting membrane potentials and their effects on the ANF spiking activity. The details of various ion channel and their dynamical responses are considered in Hodgkin and Huxley type models (e.g. Imennov and Rubinstein, 2009; Negm and Bruce, 2014; Boulet et al., 2016; Boulet and Bruce, 2017). Recently, efforts are made to identify the distributions of ion channels that are present at various locations along the ANFs in mammals (for review see Davis and Crozier, 2016). In parallel to this experimental work, an inclusion of those ion channels dynamics in Hodgkin and Huxley type models have provided insights into the effects of these ion channels in the ANF spiking activity (Boulet and Bruce, 2017). However, the substantial computational requirements of these models limit their usability for the testing of new stimulation strategies. The phenomenological nature of the model proposed in this thesis may provide a physiologically plausible and efficient alternative to the detailed biophysical models.

Another important aspect in relation to CI stimulation is related to frequency specific processing. Acoustic stimulation of the cochlea generates frequency-selective vibrations of the basilar membrane, often referred to as the frequency-place mapping. CIs emulate such frequency-place mapping using multiple electrodes that stimulate the cochlea at different places. Each electrode aims at selectively stimulating a small number of ANFs within a small region in the cochlea. However, the stimulation with any single electrode results in a large current spread due to the conductive properties of cochlea, reducing the number of 'independent

channels'. Various current-steering strategies have been proposed to counteract the conductive "smearing" and increase the number of independent frequency channels (for review, see Bierer, 2010). These strategies use simultaneous stimulation with two or more adjacent electrodes to produce focused current fields in the cochlea (e.g. bipolar, tripolar or multipolar stimulation) by using charges of opposite polarity on adjacent electrodes. However, the fact that both charges can stimulate the ANF and generate spikes has not been considered in the development of such current-steering strategies. The use of opposite polarity charges may increase the uncertainty in terms of the site of spike generation along the ANF, further degrading temporal coding in the ANF responses.

The model framework presented here therefore needs to include a description of the charge conduction in the cochlea to quantify the effect of various current-steering strategies on temporal responses of the ANF. In its current version, the model presented in chapter 3 bypasses the stage of charge conduction in the cochlea by assuming that the current injected by the CI electrode is equivalent to the "intracellular" stimulation of the neurons. However, the neuron model can be considered in combination with a charge conduction model, such as the one proposed by Goldwyn et al., 2010b, which includes a description of the conductance of the fluid filled scala tympani and the temporal bone. This model has been shown to describe the current field in the cochlea in various current-steering conditions. Considering such a front-end of the neuron model may provide a phenomenological framework to study the effects of various current steering strategies and to study their benefit on CI listener's performance in behavioral tasks.

Furthermore, a benefit of such a combination of the neuron model with charge conduction model would be the possibility to provide a framework for simulating realistic population responses. Currently, population responses can be generated by repeating the simulations of the model. However, an inclusion of the charge conduction model will allow a consideration of responses of the ANFs along the entire length of the cochlea. Future work will focus on the combination of the charge conduction model with the neuron model proposed in the current study. Such a model framework may be useful for the evaluation of the effects of different CI stimulation strategies on auditory perception.

Bibliography

Bacon, S. P. and N. F. Viemeister (1985). "Temporal Modulation Transfer Functions in Normal-Hearing and Hearing-Impaired Listeners". In: *International Journal of Audiology* 24.2, pp. 117–134.

Bierer, J. A. (2010). "Probing the Electrode – Neuron Interface With Focused Cochlear Implant Stimulation". In: *Trends in Amplification* 14.2, pp. 84–95.

Bilger, R. C., F. O. Black, and N. Hopkinson (1977). "Research plan for evaluating subjects presently fitted with implanted auditory prostheses." In: *The Annals of Otology, Rhinology & Laryngology Supplement* 86.3 Pt 2 Suppl 38, p. 21.

Borst, A. and F. E. Theunissen (1999). "Information theory and neural coding". In: *Nature neuroscience* 2.11, pp. 947–957.

Boulet, J. and I. C. Bruce (2017). "Predictions of the Contribution of HCN Half-Maximal Activation Potential Heterogeneity to Variability in Intrinsic Adaptation of Spiral Ganglion Neurons". In: *Journal of the Association for Research in Otolaryngology (JARO)*, pp. 1–22.

Boulet, J., M. White, and I. C. Bruce (2016). *Temporal Considerations for Stimulating Spiral Ganglion Neurons with Cochlear Implants.*

Brand, A, O Behrend, T Marquardt, and D McAlpine (2002). "Precise inhibition is essential for microsecond interaural time difference coding". In: *Nature* 417.6888, p. 543.

Brette, R. and W. Gerstner (2005). "Adaptive exponential integrate-and-fire model as an effective description of neuronal activity". In: *Journal of Neurophysiology* 94.5, pp. 3637–3642.

Bruce, I. C. et al. (1999a). "A stochastic model of the electrically stimulated auditory nerve: pulse-train response". In: *IEEE Transactions on Biomedical Engineering* 46.6, pp. 630–637.

Bruce, I. C. et al. (1999b). "A stochastic model of the electrically stimulated auditory nerve: single-pulse response". In: *IEEE Transactions on Biomedical Engineering* 46.6, pp. 617–629.

Bruce, I. C., M. W. White, L. S. Irlicht, S. J. O'Leary, and G. M. Clark (1999c). "The effects of stochastic neural activity in a model predicting intensity perception with cochlear implants: low-rate stimulation". In: *IEEE Transactions on Biomedical Engineering* 46.12, pp. 1393–1404.

Brughera, A., L. Dunai, and W. M. Hartmann (2013). "Human interaural time difference thresholds for sine tones: the high-frequency limit". In: *The Journal of the Acoustical Society of America* 133.5, pp. 2839–2855.

Campbell, L. J., D. J. Sly, and S. J. O'Leary (2012). "Prediction and control of neural responses to pulsatile electrical stimulation". In: *Journal of neural engineering* 9.2, p. 026023.

Carlyon, R. P., J. M. Deeks, and O. Macherey (2013). "Polarity effects on place pitch and loudness for three cochlear-implant designs and at different cochlear sites." In: *The Journal of the Acoustical Society of America (JASA)* 134.1, pp. 503–9.

Carlyon, R. P., J. M. Deeks, and C. M. McKay (2015). "Effect of Pulse Rate and Polarity on the Sensitivity of Auditory Brainstem and Cochlear Implant Users to Electrical Stimulation". In: *Journal of the Association for Research in Otolaryngology (JARO)* 16.5, pp. 653–668.

Cartee, L., C. Miller, and C. van den Honert (2006). "Spiral ganglion cell site of excitation I: Comparison of scala tympani and intrameatal electrode responses". In: *Hearing Research* 215.1-2, pp. 10–21.

Chatterjee, M. and M. E. Robert (2001). "Noise enhances modulation sensitivity in cochlear implant listeners: Stochastic resonance in a prosthetic sensory system?" In: *Journal of the Association for Research in Otolaryngology (JARO)* 2.2, pp. 159–171.

Chua, T., M Bachman, and F. Zeng (2011). "Intensity coding in electric hearing: effects of electrode configurations and stimulation waveforms". In: *Ear "&" Hearing* 32.6, pp. 679–689.

Chung, Y., B. Delgutte, H. S. Colburn, C. Y., D. B., and C. H.S. (2015). "Modeling binaural responses in the auditory brainstem to electric stimulation of the auditory nerve." In: *Journal of the Association for Research in Otolaryngology (JARO)* 16.1, pp. 135–58.

Colburn, H. S. and J. S. Latimer (1978). "Theory of binaural interaction based on auditory nerve data. III. Joint dependence on interaural time and amplitude differences in discrimination and detection". In: *The Journal of the Acoustical Society of America (JASA)* 64.1, pp. 95–106.

Colburn, H. S. (1973). "Theory of binaural interaction based on auditory-nerve data. I. General strategy and preliminary results on interaural discrimination." In: *The Journal of the Acoustical Society of America (JASA)* 54.6, pp. 1458–70.

Colburn, H. S. (1977). "Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise." In: *The Journal of the Acoustical Society of America (JASA)* 61.2, pp. 525–33.

Colburn, H. S., Y. Chung, Y. Zhou, and A. Brughera (2009). "Models of brainstem responses to bilateral electrical stimulation." In: *Journal of the Association for Research in Otolaryngology (JARO)* 10.1, pp. 91–110.

Davis, R. L. and R. A. Crozier (2016). "The Electrophysiological Signature of Spiral Ganglion Neurons". In: *The Primary Auditory Neurons of the Mammalian Cochlea*. Springer, pp. 85–116.

Dietz, M. (2016). "Models of the electrically stimulated binaural system: a review". In: *Network: Computation in Neural Systems* 27.2-3, pp. 186–211.

Dynes, S. (1996). "Discharge characteristics of auditory nerve fibers for pulsatile electrical stimuli". PhD thesis. Harvard University.

Dynes, S. B. C. and B. Delgutte (1992). "Phase-locking of auditory-nerve discharges to sinusoidal electric stimulation of the cochlea". In: *Hearing Research* 58.1, pp. 79–90.

Egger, K., P. Majdak, and B. Laback (2016). "Channel Interaction and Current Level Affect Across-Electrode Integration of Interaural Time Differences in Bilateral

Cochlear-Implant Listeners". In: *Journal of the Association for Research in Oto-laryngology (JARO)* 17.1, pp. 55–67.

Ewert, S. D. and T. Dau (2000). "Characterizing frequency selectivity for envelope fluctuations." In: *The Journal of the Acoustical Society of America (JASA)* 108.3 Pt 1, pp. 1181–96.

Finley, C. C., B. S. Wilson, and M. W. White (1990). "Models of Neural Responsiveness to Electrical Stimulation". In: *Cochlear Implant.* Pp. 55–96.

Fourcaud-Trocmé, N., D. Hansel, C. van Vreeswijk, and N. Brunel (2003). "How spike generation mechanisms determine the neuronal response to fluctuating inputs". In: *The Journal of Neuroscience* 23.37, pp. 11628–11640.

Franken, T. P., M. T. Roberts, L. Wei, N. L. Golding, and P. X. Joris (2015). "In vivo coincidence detection in mammalian sound localization generates phase delays". In: *Nature Neuroscience* 18.3, pp. 444–52.

Fraser, M. and C. M. McKay (2012). "Temporal modulation transfer functions in cochlear implantees using a method that limits overall loudness cues". In: *Hearing Research* 283.1-2, pp. 59–69.

Fredelake, S. and V. Hohmann (2012). "Factors affecting predicted speech intelligibility with cochlear implants in an auditory model for electrical stimulation". In: *Hearing Research* 287.1-2, pp. 76–90.

Frijns, J. H. M., S. L. De Snoo, and J. H. Ten Kate (1996). "Spatial selectivity in a rotationally symmetric model of the electrically stimulated cochlea". In: *Hearing Research* 95.1-2, pp. 33–48.

Galvin, J. J. and Q.-J. Fu (2005). "Effects of stimulation rate, mode and level on modulation detection by cochlear implant users". In: *Journal of the Association for Research in Otolaryngology (JARO)* 6.3, pp. 269–279.

Galvin, J. J. and Q. J. Fu (2009). "Influence of stimulation rate and loudness growth on modulation detection and intensity discrimination in cochlear implant users". In: *Hearing Research* 250.1-2, pp. 46–54.

Geurts, L. and J. Wouters (1999). "Enhancing the speech envelope of continuous interleaved sampling processors for cochlear implants". In: *The Journal of the Acoustical Society of America (JASA)* 105.4, pp. 2476–2484.

Goldwyn, J. H., E. Shea-Brown, and J. T. Rubinstein (2010a). "Encoding and decoding amplitude-modulated cochlear implant stimuli—a point process analysis". In: *Journal of Computational Neuroscience* 28.3, pp. 405–424.

Goldwyn, J. H., S. M. Bierer, and J. A. Bierer (2010b). "Modeling the electrode-neuron interface of cochlear implants: effects of neural survival, electrode placement, and the partial tripolar configuration". In: *Hearing research* 268.1, pp. 93–104.

Goldwyn, J. H., J. T. Rubinstein, and E. Shea-Brown (2012). "A point process framework for modeling electrical stimulation of the auditory nerve". In: *Journal of Neurophysiology* 108, pp. 1430–1452.

Green, D. W. and J. A. Swets (1966). *Signal detection theory and psychophysics.* Peninsular, Los Altos, CA.

Grothe, B., M. Pecka, and D. McAlpine (2010). "Mechanisms of sound localization in mammals." In: *Physiological Review* 90.3, pp. 983–1012.

Hamacher, V. (2004). "Signalverarbeitungsmodelle des elektrisch stimulierten Gehörs". PhD thesis, pp. 1–220.

Hancock, K. E. and B. Delgutte (2004). "A physiologically based model of interaural time difference discrimination". In: *The Journal of Neuroscience* 24.32, pp. 7110–7117.

Hartmann, R, G Topp, and R Klinke (1984). "Discharge patterns of cat primary auditory nerve fibers with electrical stimulation of the cochlea". In: *Hearing Research* 13.1, pp. 47–62.

Hartmann, R. and R. Klinke (1990). "Response characteristics of nerve fibers to patterned electrical stimulation". In: *Cochlear Implants.* Springer, pp. 135–160.

Heinz, M. G. and J. Swaminathan (2009). "Quantifying envelope and fine-structure coding in auditory nerve responses to chimaeric speech". In: *Journal of the Association for Research in Otolaryngology (JARO)* 10.3, pp. 407–423.

Horne, C. D. F., C. J. Sumner, and B. U. Seeber (2016). "A Phenomenological Model of the Electrically Stimulated Auditory Nerve Fiber: Temporal and Biphasic Response Properties." In: *Frontiers in Computational Neuroscience* 10, p. 8.

Imennov, N. S. and J. T. Rubinstein (2009). "Stochastic population model for electrical stimulation of the auditory nerve". In: *IEEE Transactions on Biomedical Engineering* 56.10, pp. 2493–2501.

Javel, E. (1990). "Acoustic and electrical encoding of temporal information". In: *Cochlear Implants*. Springer, pp. 247–295.

Javel, E. and R. K. Shepherd (2000). "Electrical stimulation of the auditory nerve: III. Response initiation sites and temporal fine structure". In: *Hearing Research* 140.1, pp. 45–76.

Javel, E. and N. Viemeister (2000). "Stochastic properties of cat auditory nerve responses to electric and acoustic stimuli and application to intensity discrimination". In: *The Journal of the Acoustical Society of America (JASA)*.

Jørgensen, S. and T. Dau (2011). "Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing". In: *The Journal of the Acoustical Society of America (JASA)* 130.3, p. 1475.

Joris, P. and T. C. Yin (2007). "A matter of time: internal delays in binaural processing". In: *Trends in Neurosciences* 30.2, pp. 70–78.

Joris, P. X. (2003). "Interaural time sensitivity dominated by cochlea-induced envelope patterns." In: *The Journal of Neuroscience* 23.15, pp. 6345–6350.

Joris, P. X. and T. C. T. Yin (1992). "Responses to amplitude-modulated tones in the auditory nerve of the cat". In: *The Journal of the Acoustical Society of America (JASA)* 91.1, p. 215.

Joris, P. X., D. H. Louage, L. Cardoen, and M. van der Heijden (2006). "Correlation Index: A new metric to quantify temporal coding". In: *Hearing Research* 216-217, pp. 19–30.

Joris, P, C. Schreiner, and A Rees (2004). "Neural processing of amplitude-modulated sounds". In: *Physiological Review*.

Joshi, S. N., T. Dau, and B. Epp (2014). "Modeling auditory-nerve responses to electrical stimulation". In: *7th Forum Acusticum*, pp. 1–5.

Joshi, S. N., T. Dau, and B. Epp (2017). "A model of electrically stimulated auditory-nerve fiber responses with peripheral and central sites of spike generation". In: *Journal of the Association for Research in Otolaryngology (JARO)* 18.1, pp. 1–19.

Kan, A and R. Y. Litovsky (2015). "Binaural hearing with electrical stimulation". In: *Hearing Research* 322, pp. 127–137.

Kiang, N. Y. and E. C. Moxon (1972). "Physiological considerations in artificial stimulation of the inner ear." In: *Annals of Otology, Rhinology & Laryngology* 81.5, pp. 714–730.

Kreft, H. a., G. S. Donaldson, and D. a. Nelson (2004). "Effects of pulse rate on threshold and dynamic range in Clarion cochlear-implant users." In: *The Journal of the Acoustical Society of America (JASA)* 115, pp. 1885–1888.

Laback, B. and P. Majdak (2008). "Binaural jitter improves interaural time-difference sensitivity of cochlear implantees at high pulse rates." In: *The Proceedings of the National Academy of Sciences (PNAS)* 105.2, pp. 814–7.

Laback, B., P. Majdak, and W.-D. Baumgartner (2007). "Lateralization discrimination of interaural time delays in four-pulse sequences in electric and acoustic hearing[a])". In: *The Journal of the Acoustical Society of America (JASA)* 121.4, pp. 2182–2191.

Laback, B., I. Zimmermann, P. Majdak, W.-D. Baumgartner, and S.-M. Pok (2011). "Effects of envelope shape on interaural envelope delay sensitivity in acoustic and electric hearing." In: *The Journal of the Acoustical Society of America (JASA)* 130.3, pp. 1515–1529.

Laback, B., K. Egger, and P. Majdak (2015). "Perception and coding of interaural time differences with bilateral cochlear implants". In: *Hearing Research* 322, pp. 138–150.

Liberman, M. C. and M. E. Oliver (1984). "Morphometry of intracellularly labeled neurons of the auditory nerve: correlations with functional properties". In: *Journal of Comparative Neurology* 223.2, pp. 163–176.

Litovsky, R. Y. (2016). "Binaural and Spatial Hearing in Implanted Children". In: *Pediatric Cochlear Implantation.* Springer, pp. 163–175.

Litovsky, R. Y. et al. (2006). "Bilateral Cochlear Implants in Children: Localization Acuity Measured with Minimum Audible Angle". In: *Ear & Hearing* 27.1, pp. 43–59.

Litovsky, R. Y., G. L. Jones, S. Agrawal, and R. van Hoesel (2010). "Effect of age at onset of deafness on binaural sensitivity in electric hearing in humans". In: *The Journal of the Acoustical Society of America (JASA)* 127.1, pp. 400–414.

Litvak, L. M., B. Delgutte, and D. K. Eddington (2003). "Improved temporal coding of sinusoids in electric stimulation of the auditory nerve using desynchronizing pulse trains". In: *The Journal of the Acoustical Society of America (JASA)* 114.4, pp. 2079–2098.

Louage, D. H., M. van der Heijden, and P. X. Joris (2004). "Temporal properties of responses to broadband noise in the auditory nerve". In: *Journal of Neurophysiology* 91.5, pp. 2051–2065.

Lowen, S. B. and M. C. Teich (1992). "Auditory-nerve action potentials form a nonrenewal point process over short as well as long time scales." In: *The Journal of the Acoustical Society of America (JASA)* 92.2 Pt 1, pp. 803–806.

Macherey, O., R. P. Carlyon, A. van Wieringen, and J. Wouters (2007). "A dual-process integrator–resonator model of the electrically stimulated human auditory nerve". In: *Journal of the Association for Research in Otolaryngology (JARO)* 8.1, pp. 84–104.

Macherey, O., R. P. Carlyon, A. van Wieringen, J. M. Deeks, and J. Wouters (2008). "Higher sensitivity of human auditory nerve fibers to positive electrical currents". In: *Journal of the Association for Research in Otolaryngology (JARO)* 9.2, pp. 241–251.

Macherey, O., J. M. Deeks, and R. P. Carlyon (2011). "Extending the limits of place and temporal pitch perception in cochlear implant users". In: *Journal of the Association for Research in Otolaryngology (JARO)* 12.2, pp. 233–251.

Matsuoka, A. J., P. J. Abbas, J. T. Rubinstein, and C. A. Miller (2000). "The neuronal response to electrical constant-amplitude pulse train stimulation: evoked compound action potential recordings". In: *Hearing research* 149.1, pp. 115–128.

McAlpine, D., D. Jiang, and A. R. Palmer (2001). "A neural code for low-frequency sound localization in mammals". In: *Nature Neuroscience* 4.4, pp. 396–401.

McKay, C. M. and K. R. Henshall (2010). "Amplitude modulation and loudness in cochlear implantees". In: *Journal of the Association for Research in Otolaryngology (JARO)* 11.1, pp. 101–111.

Middlebrooks, J. C. (2008). "Auditory cortex phase locking to amplitude-modulated cochlear implant pulse trains." In: *Journal of Neurophysiology* 100.1, pp. 76–91.

Miller, C. A., P. J. Abbas, J. T. Rubinstein, B. K. Robinson, A. J. Matsuoka, and G. Woodworth (1998). "Electrically evoked compound action potentials of guinea pig and cat: responses to monopolar, monophasic stimulation". In: *Hearing Research* 119.1, pp. 142–154.

Miller, C. A., P. J. Abbas, B. K. Robinson, J. T. Rubinstein, and A. J. Matsuoka (1999). "Electrically evoked single-fiber action potentials from cat: responses to monopolar, monophasic stimulation". In: *Hearing research* 130.1, pp. 197–218.

Miller, C. A., B. K. Robinson, J. T. Rubinstein, P. J. Abbas, and C. L. Runge-Samuelson (2001a). "Auditory nerve responses to monophasic and biphasic electric stimuli". In: *Hearing research* 151.1, pp. 79–94.

Miller, C. A., P. J. Abbas, and B. K. Robinson (2001b). "Response properties of the refractory auditory nerve fiber". In: *Journal of the Association for Research in Otolaryngology (JARO)* 2.3, pp. 216–232.

Miller, C. A., N. Hu, F. Zhang, B. K. Robinson, and P. J. Abbas (2008). "Changes across time in the temporal responses of auditory nerve fibers stimulated by electric pulse trains". In: *Journal of the Association for Research in Otolaryngology (JARO)* 9.1, pp. 122–137.

Moore, B. C. (2003). "Coding of sounds in the auditory system and its relevance to signal processing and coding in cochlear implants". In: *Otology & Neurotology* 24.2, pp. 243–254.

Moore, B. C. (2007). *Cochlear hearing loss: physiological, psychological and technical issues.* John Wiley & Sons.

Morse, R. P. and E. F. Evans (1996). "Enhancement of vowel coding for cochlear implants by addition of noise". In: *Nature Medicine* 2.8, pp. 928–932.

Morse, R., D Allingham, and N. Stocks (2015). "A phenomenological model of myelinated nerve with a dynamic threshold". In: *Journal of Theoretical Biology* 382, pp. 386–396.

National Institute of Health, N. (2016). *Cochlear Implants.*

Negm, M. H. and I. C. Bruce (2014). "The effects of HCN and KLT ion channels on adaptation and refractoriness in a stochastic auditory nerve model". In: *IEEE Transactions on Biomedical Engineering* 61.11, pp. 2749–2759.

Noel, V. A. and D. K. Eddington (2013). "Sensitivity of bilateral cochlear implant users to fine-structure and envelope interaural time differences". In: *The Journal of the Acoustical Society of America (JASA)* 133.4, pp. 2314–2328.

Nourski, K., P. Abbas, C. Miller, B. Robinson, and F. Jeng (2006). "Effects of remaining hair cells on cochlear implant function, 15th quarterly progress report". In: *NIH contract* 1, pp. 2–1005.

Nowak, L. and J Bullier (1998). "Axons, but not cell bodies, are activated by electrical stimulation in cortical gray matter I. Evidence from chronaxie measurements". In: *Experimental Brain Research* 118.4, pp. 477–488.

Ota, C. Y. and R. S. Kimura (1980). "Ultrastructural study of the human spiral ganglion". In: *Acta Oto-Laryngologica* 89.1-2, pp. 53–62.

Oxenham, A. J. and H. A. Kreft (2014). "Speech Perception in Tones and Noise via Cochlear Implants Reveals Influence of Spectral Resolution on Temporal Processing." In: *Trends Hear.* 18, pp. 1–14.

O'Gorman, D. E., J. A. White, and C. A. Shera (2009). "Dynamical instability determines the effect of ongoing noise on neural firing". In: *Journal of the Association for Research in Otolaryngology (JARO)* 10.2, pp. 251–267.

O'Gorman, D. E., H. S. Colburn, and C. A. Shera (2010). "Auditory sensitivity may require dynamically unstable spike generators: evidence from a model of electrical stimulation." In: *The Journal of the Acoustical Society of America (JASA)* 128.5, EL300–L305.

Plauška, A., J. G. Borst, and M. van der Heijden (2016). "Predicting binaural responses from monaural responses in the gerbil medial superior olive". In: *Journal of Neurophysiology* 115.6, pp. 2950–2963.

Pozzorini, C., R. Naud, S. Mensi, and W. Gerstner (2013). "Temporal whitening by power-law adaptation in neocortical neurons". In: *Nat. Neurosci.* 16.7, pp. 942–948.

Ramekers, D., H. Versnel, S. B. Strahl, E. M. Smeets, S. F. Klis, and W. Grolman (2014). "Auditory-nerve responses to varied inter-phase gap and phase duration of the

electric pulse stimulus as predictors for neuronal degeneration". In: *Journal of the Association for Research in Otolaryngology (JARO)* 15.2, pp. 187–202.

Rattay, F., P. Lutter, and H. Felix (2001). "A model of the electrically excited human cochlear neuron: I. Contribution of neural substructures to the generation and propagation of spikes". In: *Hearing research* 153.1, pp. 43–63.

Rattay, F., T. Potrusil, C. Wenger, A. K. Wise, R. Glueckert, and A. Schrott-Fischer (2013). "Impact of morphometry, myelinization and synaptic current strength on spike conduction in human and cat spiral ganglion neurons". In: *PloS one* 8.11, e79256.

Remme, M. W. H. et al. (2014). "Subthreshold resonance properties contribute to the efficient coding of auditory spatial cues." In: *The Proceedings of the National Academy of Sciences (PNAS)*, pp. 1–10.

Rubinstein, J. T., B. S. Wilson, C. C. Finley, and P. J. Abbas (1999). "Pseudospontaneous activity: Stochastic independence of auditory nerve fibers with electrical stimulation". In: *Hearing Research* 127.1-2, pp. 108–118.

Rubinstein, J. T. (1991). "Analytical theory for extracellular electrical stimulation of nerve with focal electrodes. II. Passive myelinated axon." In: *Biophysical Journal* 60.3, p. 538.

Rubinstein, J. T., C. A. Miller, H. Mino, and P. J. Abbas (2001). "Analysis of monophasic and biphasic electrical stimulation of nerve". In: *IEEE Transactions on Biomedical Engineering* 48.10, pp. 1065–1070.

Rutherford, M. A., N. M. Chapochnikov, and T. Moser (2012). "Spike Encoding of Neurotransmitter Release Timing by Spiral Ganglion Neurons of the Cochlea". In: *The Journal of Neuroscience* 32.14, pp. 4773–4789.

Scott, L. L., P. J. Mathews, and N. L. Golding (2005). "Posthearing developmental refinement of temporal processing in principal neurons of the medial superior olive." In: *Journal of Neuroscience* 25.35, pp. 7887–95.

Seeber, B. U. and H. Fastl (2008). "Localization cues with bilateral cochlear implants". In: *The Journal of the Acoustical Society of America (JASA)* 123.2, pp. 1030–1042.

Shackleton, T. M., R. Meddis, and M. J. Hewitt (1992). "Across frequency integration in a model of lateralization". In: *The Journal of the Acoustical Society of America (JASA)* 91.4, p. 2276.

Shannon, R. V., F. G. Zeng, V Kamath, J Wygonski, and M Ekelid (1995). "Speech recognition with primarily temporal cues." In: *Science* 270.5234, pp. 303–4.

Shannon, R. V. (1992). "Temporal modulation transfer functions in patients with cochlear implants". In: *The Journal of the Acoustical Society of America (JASA)* 91, pp. 2156–2164.

Shepherd, R. K. and E. Javel (1999). "Electrical stimulation of the auditory nerve: II. Effect of stimulus waveshape on single fibre response properties". In: *Hearing research* 130.1, pp. 171–188.

Shepherd, R. K., N. A. Hardie, and J. H. Baxi (2001). "Electrical stimulation of the auditory nerve: single neuron strength-duration functions in deafened animals". In: *Annals of Biomedical Engineering* 29.3, pp. 195–201.

Smith, D. W. and C. C. Finley (1997). "Effects of electrode configuration on psychophysical strength-duration functions for single biphasic electrical stimuli in cats". In: *The Journal of the Acoustical Society of America (JASA)* 102.4, pp. 2228–2237.

Smith, Z. M. and B. Delgutte (2007). "Sensitivity to interaural time differences in the inferior colliculus with bilateral cochlear implants". In: *The Journal of Neuroscience* 27.25, pp. 6740–6750.

Smith, Z. M., B. Delgutte, and A. J. A. Oxenham (2002). "Chimaeric sounds reveal dichotomies in auditory perception." In: *Nature* 416.6876, pp. 87–90.

Spoendlin, H and A Schrott (1989). "Analysis of the human auditory nerve". In: *Hearing research* 43.1, pp. 25–38.

Stern, R. M., A. S. Zeiberg, and C. Trahiotis (1988). "Lateralization of complex binaural stimuli: a weighted-image model". In: *The Journal of the Acoustical Society of America (JASA)* 84.1, pp. 156–165.

Swaminathan, J. and M. G. Heinz (2012). "Psychophysiological analyses demonstrate the importance of neural envelope coding for speech perception in noise." In: *The Journal of Neuroscience* 32.5, pp. 1747–56.

Undurraga, J. A., R. P. Carlyon, J. Wouters, and A. van Wieringen (2013). "The polarity sensitivity of the electrically stimulated human auditory nerve measured at the level of the brainstem". In: *Journal of the Association for Research in Otolaryngology (JARO)* 14.3, pp. 359–377.

van den Honert, C. and J. T. Mortimer (1979). "The Response of the Myelinated Nerve Fiber To Short Duration". In: *Annals of Biomedical Engineering* 7, pp. 117–125.

van den Honert, C. and P. Stypulkowski (1984). "Physiological properties of the electrically stimulated auditory nerve. II. Single fiber recordings". In: *Hearing research* 14.3, pp. 225–243.

van der Heijden, M., D. H. Louage, and P. X. Joris (2011). "Responses of auditory nerve and anteroventral cochlear nucleus fibers to broadband and narrowband noise: implications for the sensitivity to interaural delays". In: *Journal of the Association for Research in Otolaryngology (JARO)* 12.4, pp. 485–502.

van Hoesel, R. (2011). "Bilateral Cochlear Implants". In: *Auditory Prostheses, New Horizons*. Vol. 39. 5, pp. 13–57.

van Hoesel, R. J. (2007). "Sensitivity to binaural timing in bilateral cochlear implant users". In: *The Journal of the Acoustical Society of America (JASA)* 121.4, pp. 2192–2206.

van Hoesel, R. J. (2012). "Contrasting benefits from contralateral implants and hearing aids in cochlear implant users". In: *Hearing Research* 288.1, pp. 100–113.

van Hoesel, R. J., G. L. Jones, and R. Y. Litovsky (2009). "Interaural time-delay sensitivity in bilateral cochlear implant users: effects of pulse rate, modulation rate, and place of stimulation". In: *Journal of the Association for Research in Otolaryngology (JARO)* 10.4, pp. 557–567.

van Wieringen, A., O. Macherey, R. P. Carlyon, J. M. Deeks, and J. Wouters (2008). "Alternative pulse shapes in electrical hearing". In: *Hearing research* 242.1, pp. 154–163.

Vandali, A. E., L. A. Whitford, K. L. Plant, G. M. Clark, et al. (2000). "Speech perception as a function of electrical stimulation rate: using the Nucleus 24 cochlear implant system". In: *Ear and hearing* 21.6, pp. 608–624.

Verveen, A. A. and H. E. Derksen (1965). "Fluctuations in membrane potential of axons and the problem of coding". In: *Kybernetik* 2.4, pp. 152–160.

Verveen, A. (1961). *Fluctuation in excitability*.

Wilson, B. S. and M. F. Dorman (2008). "Cochlear implants: a remarkable past and a brilliant future". In: *Hearing research* 242.1, pp. 3–21.

Wilson, B. S., C. C. Finley, D. T. Lawson, R. D. Wolford, D. K. Eddington, and W. M. Rabinowitz (1991). "Better speech recognition with cochlear implants". In: *Nature* 352.6332, pp. 236–238.

Yin, P., J. S. Johnson, K. N. O'Connor, and M. L. Sutter (2011). "Coding of amplitude modulation in primary auditory cortex." In: *Journal of Neurophysiology* 105.2, pp. 582–600.

Zhang, F., C. A. Miller, B. K. Robinson, P. J. Abbas, and N. Hu (2007). "Changes across time in spike rate and spike amplitude of auditory nerve fibers stimulated by electric pulse trains". In: *Journal of the Association for Research in Otolaryngology (JARO)* 8.3, pp. 356–372.

Zhou, N., C. T. Kraft, D. J. Colesa, and B. E. Pfingst (2015). "Integration of Pulse Trains in Humans and Guinea Pigs with Cochlear Implants". In: *Journal of the Association for Research in Otolaryngology (JARO)* 16.4, pp. 523–534.

A

Effects of relative and absolute frequency in the spectral weighting of loudness [e]

Abstract

The loudness of broadband sound is often modeled as a linear sum of specific loudness across frequency bands. In contrast, recent studies using molecular psychophysical methods suggest that low and high frequency components contribute more to the overall loudness than mid frequencies. In a series of experiments, the contribution of individual components to the overall loudness of a tone complex was assessed using the molecular psychophysical method as well as a loudness matching task. The stimuli were two spectrally overlapping ten-tone complexes with two equivalent rectangular bandwidth spacing between the tones, making it possible to separate effects of relative and absolute frequency. The lowest frequency components of the "low-frequency" and the "high-frequency" complexes were 208 and 808 Hz, respectively. Perceptual-weights data showed emphasis on lowest and highest frequencies of both the complexes, suggesting spectral-edge related effects. Loudness matching data in the same listeners confirmed the greater contribution of low and high frequency components to the

---

overall loudness of the ten-tone complexes. Masked detection thresholds of the individual components within the tone complex were not correlated with perceptual weights. The results show that perceptual weights provide reliable behavioral correlates of relative contributions of the individual frequency components to overall loudness of broadband sounds.

Keywords *Loudness perception, Broadband signals, Spectral integration, Acoustic modeling, Sound discrimination*

A.1   Introduction

Loudness is generally defined as the perceived magnitude of a sound that primarily changes with intensity (Florentine, 2011). Here, "primarily" is a key adjective, since loudness is also affected by other physical dimensions such as duration, frequency, and bandwidth (for review, see Jesteadt and Leibold, 2011). Almost all the sounds in our acoustic environment, including speech, are broadband in nature. Nevertheless, more research has been focused on the loudness of sinusoidal stimuli than broadband sounds. This discrepancy may be the reason why loudness discomfort remains a major source of dissatisfaction with hearing-aid prescription (Kochkin, 2010), despite use of loudness data in hearing aid fitting. Understanding the laws that govern the loudness of broadband sounds is, therefore, clinically important as well as ecologically relevant. This is one of a series of studies exploring the use of molecular psychophysics to assess the contributions of individual frequency bands to the total loudness of broadband sounds.

Zwicker et al., 1957 investigated the effect of the frequency spacing of tones within a multi-tone complex on the loudness of the complex. They showed that the loudness did not change when all the components remained within a single critical band. The loudness increased, however, when spacing of the tones occupied a bandwidth larger than a critical band. This trend of spectral summation of loudness was observed for noise stimuli as well. Zwicker, 1958 and Zwicker and Scharf, 1965 proposed a model of loudness perception for broadband sounds, which had a

significant impact on all later models. In their model, stimuli were transformed to produce geometric excitation patterns, which were then converted to specific loudness patterns using a function that described a power-law type relationship between the degree of excitation and loudness. In this and later models, specific loudness (N') was defined as the loudness per critical band. The overall loudness in sones was obtained by calculating the area under the curve for N' plotted against a frequency scale with units of critical bands, Barks (in the model of Zwicker, 1958) or equivalent rectangular bandwidths (ERBs) (Glasberg and Moore, 1990) in the model of Moore and colleagues (Moore et al., 1997). In summary, the loudness of a broadband sound is generally assumed to be a linear summation of specific loudness across auditory filters. This modeling framework has served as the basis for all modern psychoacoustical models of loudness perception (e.g., Chalupper and Fastl, 2002; Moore and Glasberg, 2004; Moore et al., 1997). The loudness model described by Moore and colleagues serves as an ANSI standard for calculation of loudness (ANSI, 2007) as well as a basis for hearing aid prescriptions and for their validation (e.g., NAL, CAMEQ) (Moore and Sęk, 2012).

Leibold et al., 2007 used molecular psychophysical methods to determine the relative contribution of each component of a five-tone complex to the loudness of the complex. In their study, the components were logarithmically spaced, centered at 1000 Hz; the lowest bandwidth tested was 46 Hz and the highest was 2119 Hz. Their analysis showed that for the conditions where bandwidth did not exceed a critical band, the perceptual weights for the components of a complex were relatively uniform. For conditions in which the bandwidth exceeded a single critical band, weights for components differed significantly from one another, suggesting differential contributions to the overall loudness. Leibold et al., 2007 used the loudness model of Moore and Glasberg, 2004 to predict the weighting patterns they observed. The model was successful in predicting the observed weights for all but the widest bandwidth condition, where the components were spread across several critical bands and did not mask one another. In that condition, the model predicted relatively uniform weights, but the observed weights continued to show the same emphasis on the lowest and highest components observed at intermediate bandwidths. The results for the widest bandwidth condition appeared to

contradict the assumption common to standard loudness models that the loudness of a broadband sound is a simple linear summation of loudness across the auditory filters.

To further investigate summation of loudness across the auditory filters, Jesteadt et al., 2014 used a six-tone complex, components of which were widely separated over a five-octave range (octave spaced frequencies, from 250 to 8000 Hz). Such a stimulus avoids the possible confounding effect of one component masking another and reducing its contribution to the overall loudness. The effect of overall level was investigated in two conditions, one in which the mean level of all the tones was equal in sound pressure level (SPL) and one where levels were corrected for threshold, resulting in components equal in sensation level (SL). In both cases, at low overall levels, low frequencies dominated the loudness. With increasing level, however, the low frequency emphasis was reduced and high frequencies received more weight. Although the loudness model (Moore and Glasberg, 2004) failed to predict the observed weighting function or any change in relative weights with increasing level, Jesteadt et al., 2014 showed that weights from the model correlated with the specific loudness predictions of the model, suggesting that the underlying specific loudness of the components of a broadband sound could be assessed behaviorally using the perceptual weights. An increase in high frequency weight with increasing level has also been reported for sample discrimination of intensity (Kortekaas et al., 2003; Leibold et al., 2009) and discrimination of spectral shapes (Lentz 2006).

Central to our investigation in this article is the question posed by the studies reviewed above, regarding the calculation of the specific loudness of individual frequency bands or auditory filter outputs in current loudness models. One can consider three possible explanations of the pattern of low- and high-frequency emphasis in the existing perceptual-weights data. The first is based on differences in growth of loudness with increasing level at different frequencies. For example, loudness grows more rapidly at low frequencies than at middle or high frequencies (Hellman and Zwislocki, 1968). The model would fail to reproduce the observed frequency emphasis in perceptual weights if it underestimates how fast loudness grows at a particular frequency. The second is to assume that the model makes

correct assumptions regarding growth of loudness, but that the loudness is indeed summed unequally across the auditory filters, and that the weights reflect those differences. The third is to assume that the weights do not reflect contributions to overall loudness but rather reflect sensitivity to the changes in the levels of individual components used to derive the weights.

Differences in the growth of loudness as a function of frequency are often estimated from equal loudness contours (e.g., Jesteadt and Leibold, 2011; Scharf, 1978). Although Glasberg and Moore, 2006 showed that their loudness model can predict the standard equal-loudness contours reasonably well, a direct comparison of their predicted contours and those in the current ISO (2003) standard [Jesteadt and Leibold, 2011, Fig. 5.5] suggests that the model may underestimate the steepness of the loudness function at low frequencies. Jesteadt et al., 2014 showed low frequency emphasis at low overall levels in both equal-SL and equal-SPL conditions. The lowest frequency component was presented at levels closer to quiet threshold than other components in the equal- SPL condition and at a higher SPL in the equal-SL condition. Therefore, it is unlikely that audibility of a tone or absolute level could account for the increased weight given to the lowest-frequency component in both conditions. Recently, Oberfeld et al., 2012 obtained perceptual weights for loudness of noise made up of multiple bands in a condition where levels of the noise bands were equated to produce equal loudness. In line with the results of Jesteadt et al., 2014, their listeners put highest weight on the lowest component of the complex. Oberfeld et al., 2012 demonstrated that the weights could not be accounted for by assuming different slopes of the loudness function at different frequencies. These results provide indirect support for the second and third explanations, in which the model fails to predict the weights because summation across the auditory filters carries unequal weighting or the weights reflect sensitivity to change rather than contribution to total loudness. The series of experiments reported below was designed to investigate the unequal weighting of loudness across the filters and its dependence on stimulus parameters.

A.2   Experiment I:Effect of a Component's Relative Position on its Specific Loudness

The components of the six-tone complex used by Jesteadt et al., 2014 were spread over a wide frequency range to avoid masking of one component by another. It is therefore unlikely that the emphasis on the lowest and highest components observed by Jesteadt et al., 2014 was due to reduced masking of those components. It might have been due to their edge position in the complex for some reason other than masking or to the specific frequencies of those components. For example, Moore and Ohgushi, 1993 tested an effect of spacing between components of an inharmonic-tone complex on their audibility. They showed that when components were separated by equal ERB units, the lowest and the highest edge partials were most audible in all the conditions. Audibility of the inner partials, however, also increased with increasing spacing. While the method used in their study is not directly comparable to ours, their thesis regarding a "salience" of the partials is certainly applicable. This experiment was designed to investigate the effects of relative and absolute position of the components of a complex tone on the perceptual weight in a loudness comparison task.

A.2.1   Listeners

Six listeners with normal hearing participated in the study. All listeners had audiometric thresholds below 15 dB hearing level (HL) at the octave audiometric frequencies. All listeners were compensated for their time of participation in the study. The experimental procedures were approved by the institutional review board at the Boys Town National Research Hospital.

A.2.2   Stimuli

Two ten-tone complexes, components of which were two-ERB apart, were used to study the frequency specificity of the contributions of individual components to total loudness. The low-frequency complex included components at 208, 313,
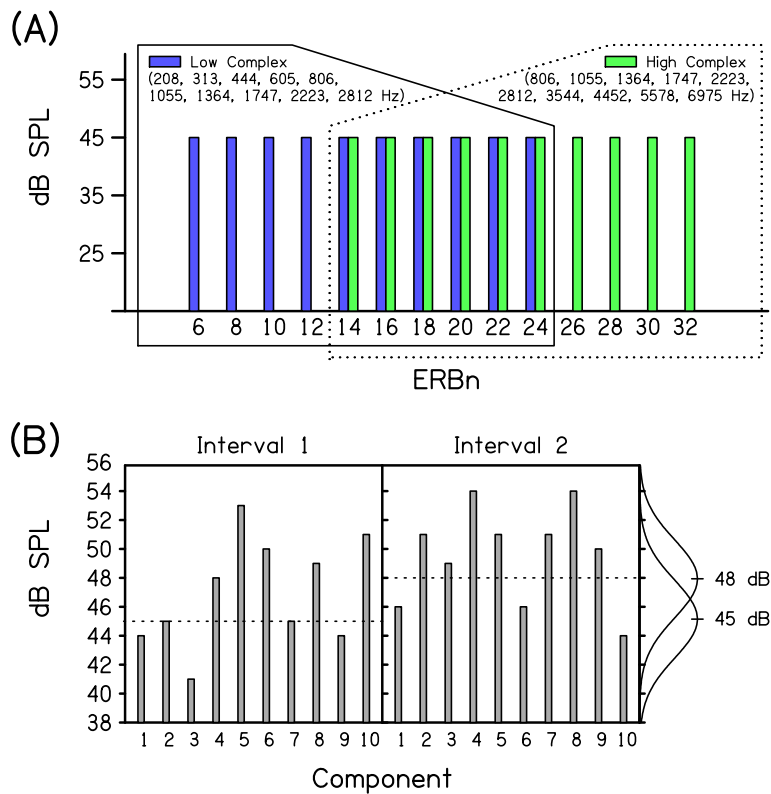
Figure A.1: **(A)** Spectral configuration of the two ten-tone complexes. The lowest frequency of the High complex was the fifth component in the Low complex. This stimulus configuration was designed to test the effect of absolute frequency and relative position of the component in a tone complex on the perceptual weight it received. **(B)** Illustration of the task used to obtain perceptual weights. Level of each component in the "signal" and "non-signal" interval was selected from a distribution with a mean of 48 and 45 dB, respectively. Listeners were asked to pick the interval that was louder.

444, 605, 806, 1055, 1364, 1747, 2223, and 2812 Hz. The high-frequency complex included components at 806, 1055, 1364, 1747, 2223, 2812, 3544, 4452, 5578, and 6975 Hz. The fifth component of the low complex (806 Hz) was the first (lowest) component of the high complex. The frequency overlap of the two complexes is shown in Fig. A.1(A). This stimulus configuration made it possible to distinguish between absolute frequency, represented by the low and high complexes, and

relative frequency, represented by position within the complex. The stimuli were 500-ms long with 10-ms onset and offset cosine-squared ramps. Stimuli were created as individual sinusoids in Matlab with a sampling rate of 44,100 Hz with 24-bit resolution. For each interval of every trial, the level of each sinusoid was selected from a rectangular distribution with a range of 16 dB in 1-dB steps and the ten sinusoids were then added in random phase. The low-frequency complex and the high-frequency complex were presented in separate blocks. The order of presentation of the blocks was randomized.

A.2.3   Methods

Listeners were seated in a double-walled sound attenuating booth. Stimuli were generated using 24-bit digital-to-analog converters (Digital Audio Labs, Card-Deluxe) and presented monaurally to the left ear through Sennheiser 250-HD headphones.

In a two-interval forced choice (2IFC) task, listeners were asked to pick the interval that was louder. For the "signal" interval, the mean of the level distribution for the individual components was 48 dB SPL (Overall mean level of 58 dB); for the "non-signal" interval it was 45 dB SPL (Overall mean level of 55 dB). A visual description of a sample trial is shown in Fig. A.1(B). The two intervals, each 500-ms long, were separated by a 500-ms inter-stimulus-interval. Each trial began with a 500-ms light emitting diode (LED) light on a response box indicating the beginning of the trial. Separate lights to cue the listener highlighted each interval. There was no gap between the initial warning light offset and the onset of the light indicating the beginning of the first interval. Stimulus onset was simultaneous with the onset of the interval lights. Although each trial could be marked with a "correct" and "incorrect" interval, listeners received no feedback regarding their response. Data were collected in blocks of 100 trials and appropriate breaks were given to the listeners. Each listener completed 1000 trials for each of the two ten-tone complexes. Blocks for the two ten-tone complexes were run in a random order.

Moore and Glasberg, 2004 model was tested for its ability to predict listener's responses. It was used because it represents the ANSI standard, has been used

for development and evaluation of hearing aid fitting strategies and was used
by Leibold et al., 2007 and Jesteadt et al., 2014. The model was provided with a
headphone response of Sennheiser HD-250 headphones, measured on a flat plate
coupler and was run with the same configuration of the stimuli on every trial that
each listener received. The model estimated the loudness of the two stimuli on
every trial as an artificial observer and voted for the louder interval. This method
avoided any confounds such as differences in weights due to a rove distribution or
changes in actual levels due to headphone calibration issues. Responses by the
model were then analyzed in the same way as responses by the individual listeners.

A.2.4   Results

For both conditions tested here, listeners performed at a minimum of 70% correct
in the sample discrimination task, with mean of 77% correct for the low complex
and 78.9% correct for the high complex. Perceptual weights for the components
were obtained using the difference in the level of each component between the
two intervals (interval 2 - interval 1) to predict the listener's responses in a linear
model, as shown in equation A.1 (Lutfi, 1995; Richards and Zhu, 1994), where x_i is
the difference in level between the two intervals for component w_i x_i represents
the weight given to that component, and C is additive internal noise.

$$D = \sum_{i=1}^{i=10} w_i \; x_i + C \tag{A.1}$$

Weights were calculated separately using multiple linear regression for trials
when the correct interval was first or second, and then averaged. Because the
signal was in the same interval for all trials included in a given regression analysis,
only the differences in the levels of the components could contribute to systematic
differences in the responses (Berg, 1990). The same method was used to obtain
the weights predicted by the loudness model.

Weights were normalized to sum to unity, since our goal was to investigate
the "relative" contributions of the components of the complex. Mean normalized

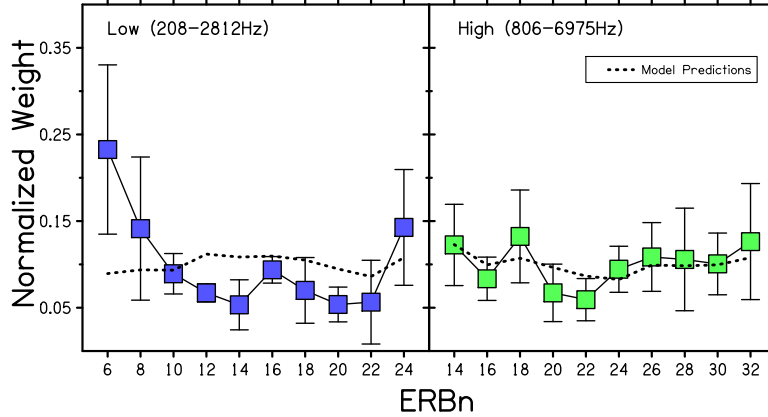observed weights as well as the model predictions are shown in Fig. A.2.

Figure A.2: Mean of the normalized perceptual weights and their standard deviations from experiment 1. The weights were obtained using a multiple regression model to predict a listener's responses on every trial using the difference in level between interval 2 and interval 1 for each individual component. Weights predicted by the loudness model are shown with a dashed line.

This experiment was designed to determine whether frequencies below 500 Hz receive higher weight because lower frequencies contribute more to total loudness than current models predict or because their position at the lower edge of the complex makes them more salient. For the low complex, the lowest and highest frequency components both received greater weight than was predicted by the model. For the high complex, the lowest frequency component received greater weight than all but one other component, but the model actually predicted this effect. The similarity of the patterns of weights in the two panels suggests that relative position is important. The effect of relative position is clearer in Fig. A.3, where the normalized weights are plotted for individual subjects. Subjects 1 and 3 assign lower relative weight to the lowest component when it is part of the low complex than when it is part of the high complex and subject 6 shows no difference in weights for the two complexes. The importance of the sixth component in the mean data for the high complex is due to subject 4.

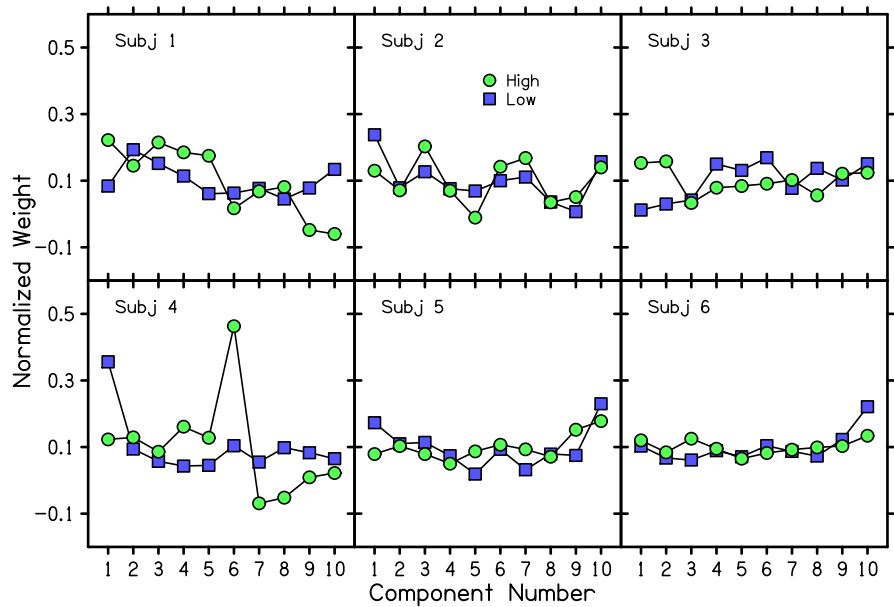A repeated measures analysis of variance (ANOVA) comparing the weights as-

Figure A.3: Normalized perceptual weights for six individual subjects in experiment 1, plotted as a
function of component number within the low and high complexes.

signed to the ten components of each complex based on their relative frequencies
showed a significant effect of relative frequency within the complex [$F(9,54.97) =$
2.12; p = 0.043], but no significant interaction of relative frequency and absolute
frequency [$F(7,55.22) = 0.56$; p = 0.78][b]. For the six overlapping frequencies, statis-
tical analysis indicated no significant difference in the overall weight assigned to
those frequencies in the low and high complexes [$F(1,9.43) = 1.56$; p = 0.242], no sig-
nificant differences among the weights assigned to the six components [$F(5,23.24)$
= 2.07; p = 0.105] and no significant interaction of relative frequency and absolute
frequency [$F(5,23.24) = 0.63$; p = 0.68]. The six frequencies were not weighted
differently when they were the lowest six components than when they were the

highest six components.

A.3 Experiment II: Relationship Between Audibility and Specific Loudness

The slope of the loudness function is frequency dependent (Scharf, 1978). The information provided by the level rove may be greater for frequencies where the slope of the loudness function is steeper than for frequencies where it is shallower. Although this is a plausible assumption, data in the literature suggest that neither frequency (Jesteadt et al., 1977) nor the slope of the loudness function (Hellman et al., 1987) affects performance in intensity discrimination. Berg, 1990 and Lutfi and Jesteadt, 2006 found that more intense tones receive more weight when tones are presented sequentially in a sample discrimination task, even when the more intense tones carry less information. Turner and Berg, 2007 and Oberfeld et al., 2013 ruled out forward masking as a contributor to this level dominance, but greater audibility could still be a factor.

Leibold et al., 2007 investigated the relationship between the audibility of a component and the perceptual weight it received in a loudness comparison task. They found that edge frequencies of closely spaced tone complexes had lower thresholds than frequencies in the middle of the complex. As the spacing between the components increased, thresholds became approximately equal among the components. The correlation between the audibility thresholds and the perceptual weight weakened with increasing spacing between the tones. The widest bandwidth tested in their study, however, was only 2119 Hz. Jesteadt et al., 2014 obtained weights for components that were presented at equal SL, avoiding the effect of threshold differences across components. Oberfeld et al., 2012 obtained weights for noise bands that were matched in loudness, avoiding any "loudness" domi-

---

[b] The ANOVAs reported here were run with PROC GLIMMIX using sas/STAT software, v. 9.4. Adjustments for multiplicity were made with the simulation technique (Westfall et al., 2011). The Kenward-Roger adjustment was used, as is standard in repeated measures, resulting in fractional degrees of freedom.

nance. Both studies, however, found that the listeners put higher weight on the low frequency component of their stimuli. These results could not be reconciled with level dominance, unless the response to each component is affected when presented simultaneously with other components. For example, low-frequency maskers more readily mask high-frequency signals, a phenomenon known as the upward spread of masking. Mutual masking might alter or enhance the audibility of certain components of the complex tones, producing level dominance that is reflected in the perceptual weights.

This experiment tested the relationship between listener's masked thresholds for each component in the complex and the weights obtained in experiment 1. If masked thresholds show a systematic trend, where components that receive highest weight have lowest masked threshold and vice versa, it would suggest that the weights represent level dominance created by mutual masking.

A.3.1   Methods

The six listeners from experiment 1 participated in this experiment, but only four of the six completed all of the conditions. The stimuli used in the experiment were identical to stimuli used in experiment 1. Masked thresholds for each component in the presence of the other nine components of the complex were obtained using a 2IFC adaptive procedure to estimate the 70.7 % point on a psychometric function (Levitt, 1971). In the two-interval task, the "incorrect" interval contained the nine non-target components of a complex, each at 45 dB SPL before correcting for the headphone response, while the "correct" interval contained the target tone, along with the other nine components. The level of the non-signal complex was roved randomly across trials with a range of +/- 3 dB with 1 dB steps to avoid use of overall loudness as a cue for detection of the addition of a signal frequency. The two intervals, each 500-ms long, were separated by 500 ms of silence. Listening intervals were marked by lights and each trial was preceded by a 500-ms light indicating the beginning of a trial, as in experiment 1. Listeners were instructed to choose the interval that contained a signal tone. Correct-answer feedback was provided after the listener's response.
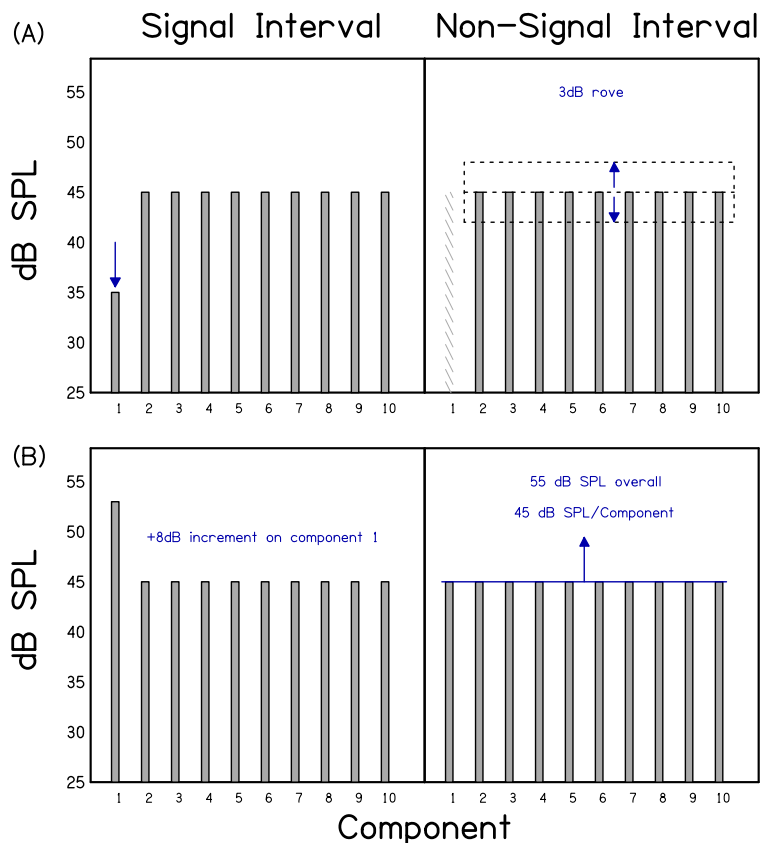
Figure A.4: **(A)** An illustration of the task used to obtain the masked threshold for individual components. The signal interval included the signal tone at a level which was changed adaptively. The non-signal interval contained only 9 tones presented at equal level with mean of 45 dB/component and their levels were roved across trials in 3 dB range. **(B)** Illustration of the loudness matching task used in experiment 3. The standard complex included one component that was either incremented or decremented by 8 dB. The level of the comparison tone complex was changed adaptively to match the loudness of the standard complex.

Each block of trials began with the target tone presented at 60 dB SPL, so that the listener could clearly hear out the signal tone. The step size was 5 dB until after the second reversal and 2 dB for the remaining trials. Each block consisted of 50 trials and four threshold estimates were obtained for each component. The task

is described in Fig. A.4(A). Thresholds were obtained for all the components in a random order.
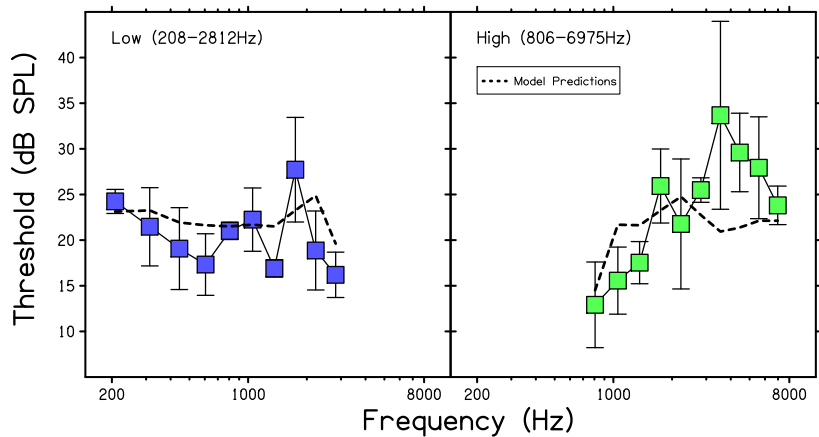
A.3.2   Results

Figure A.5: Mean masked thresholds and their standard deviation for individual components of the ten-tone complexes in experiment 2. Predictions of the partial loudness model with a criterion of 5 phons are shown with dashed lines.

Average thresholds for each component of both complexes did not differ much across the listeners. Therefore the arithmetic means across listeners are shown in Fig. A.5. Moore et al., 1997 described a version of the loudness model that can be used to predict the loudness of one component of a complex stimulus in the presence of other components and noted that partial loudness can be used to estimate threshold, if one assumes that the signal at threshold has some uniform low loudness. They proposed 2 phons as an appropriate value. The partial-loudness model was used to predict the masked thresholds for all the components for individual subjects, for assumed threshold criteria of 1 to 10 phons. A criterion value that minimized the least square error between model predictions and observed thresholds was then determined for individual listeners. The optimum values for loudness at threshold for the four listeners ranged from 4 to 6 phons. Mean

thresholds predicted by the partial loudness model using a loudness criterion of 5 phons are shown in Fig. A.5. This criterion provides the best fit to the mean data of listeners for all the frequencies in both complexes. This value is higher than the 2-phon criterion proposed by Moore et al., possibly due to the level rove in the non-signal interval.

The pattern of thresholds for individual components clearly differed for the low and high complexes. Thresholds for the lowest components in the low complex were elevated because absolute thresholds were higher for frequencies below 500 Hz. Thresholds near the high-frequency edge of both complexes were lower than those closer to the center of the complex. The thresholds of the six components common to both complexes did not differ significantly as a group across the two complexes [$F(1,7.102) = 0.79$; $p = 0.403$], but did differ across the six components [$F(5,13.03) = 4.89$; $p = 0.0098$]. The interaction of relative frequency and absolute frequency was not significant [$F(5,13.03) = 2.12$; $p = 0.127$]. A repeated-measures ANOVA of data for all ten components comparing thresholds components in the low- and high-frequency complexes in terms of their relative position showed significant differences between the low and high complexes [$F(1,18.2) = 6.01$, $p = 0.024$], a significant effect of component position [$F(9,40.56) = 2.84$, $p = 0.011$] and a significant complex by position interaction [$F(9,40.56) = 4.11$, $p = 0.0008$]. As seen in Fig. A.5, the partial loudness model correctly predicts that thresholds of the lowest components of the low complex are elevated relative to the lowest components of the high complex. This is consistent with the loudness model prediction (Fig. A.2) that the lowest components of the low complex will make a smaller contribution to overall loudness than higher frequency components in the same complex. The perceptual-weights data do not agree with this prediction. An analysis of the perceptual weights from experiment 1 for the four listeners with complete data in experiment 2, using threshold as an additional predictor of the perceptual weights, showed no significant difference in overall weight between the complexes or relative position within the complexes and a non-significant inter-action of relative and absolute frequency, as in the earlier analysis. This analysis showed no effect of threshold on perceptual weight [$F(1,18.62) = 0.30$; $p = 0.59$].

A.4 Experiment III: Contributions to Total Loudness Estimated from Loudness Matching

Results of experiment 1 suggest that perceptual weights obtained at lower-edge-frequencies in the loudness task are determined by their relative position within a broadband signal. Results of the second experiment show that the perceptual weights observed in the first experiment are not due to the audibility or masking of individual components by one another. The threshold configurations differ significantly from one complex to the other while the weights do not. In line with suggestions from Jesteadt et al. (2014) and Oberfeld et al. (2012), results of the first two experiments suggest that the global loudness of broadband sounds is based on an unequally weighted sum of specific loudness across the auditory filters.

If all the frequencies do not contribute equally to the overall loudness, this should also be seen in loudness matching. For example, an X dB increment or decrement of a component that receives higher weight should produce more change in overall loudness than the same increment or decrement of a component that receives lower weight. In this experiment, we determined the change in loudness level of ten-tone complexes resulting from 8-dB increments or decrements to individual components.

A.4.1 Methods

The same six listeners from experiment 1 participated in this experiment. The stimuli used in the experiment were identical to stimuli used in experiment 1. The effects of an 8-dB increment or decrement in the level of a single component (component number 1, 2, 5, 6, 9, and 10) on the overall loudness of the complex were tested in two separate conditions. An 8-dB increment on one component produced a 1.85-dB increase in the overall level of the complex while an 8-dB decrement produced a 0.58-dB decrease.

Loudness matching was carried out in a 2IFC task using interleaved adaptive tracks (Jesteadt, 1980). The test complex, depending on the condition, contained either an increment or a decrement of 8 dB on one component of the ten-tone

complex. The remaining nine components were presented at 45 dB SPL. The comparison complex contained all of the components at equal level and that level was adapted to determine the level required for equal loudness. The listener's task was to indicate the interval that was louder [Fig. A.4(B)]. Data were collected in 100-trial blocks, 50 trials for each of the two interleaved tracks. On one track, the beginning level of the comparison complex was 35 dB/component, and was increased adaptively, while on the other track, the beginning level was 55 dB/component and was decreased adaptively. The initial step size of 8 dB was reduced to 2 dB after two reversals. The 70.7 % and 29.1 % points were estimated from the means of the reversals after the decrease in step size, and the average of the two was used to determine the level of the complex with ten equal-level components required to match the loudness of the complex where the level of one component had been altered. Each listener completed two loudness matches for each of the six components, in random order.

Each trial consisted of a 500-ms warning interval and two observation intervals, each 500-ms long, which were separated by 500-ms of silence. An LED light on a response box marked each interval. Listeners received no feedback regarding their responses.

A.4.2   Results

The mean loudness matches across listeners for an increment or decrement of 8 dB on various components of a multi-tone complex are shown in Fig. A.6, along with predictions from the loudness model. Upward pointing triangles show conditions where 8 dB was added to the component and downward pointing triangles show conditions where 8 dB was subtracted from the component. It is not surprising that complexes with one component incremented by 8 dB resulted in higher loudness matches than those with one component decremented by 8 dB. The data were analyzed, therefore, in terms of the absolute magnitude of the deviations from the 55 dB midpoint. A repeated measures ANOVA was used to assess the effects of absolute frequency (low vs high complex), relative frequency (the six positions within each complex), the direction of the change in level (-8 dB vs +8 dB) and the

pairwise interactions of those three main effects. The magnitude of the deviations
from the 55 dB midpoint did not differ for the low vs high complexes [$F(1,16.41)$
$= 1.03$; $p = 0.326$], but the effect of an 8 dB decrement was greater than that of an
increment [$F(1,16.41) = 6.05$; $p = 0.025$] and the effect of relative position within
each complex was significant as well [$F(5,54.24)= 8.42$; $p < 0.0001$]. The difference
between the effect of an increment or decrement did not differ for low vs high
complexes [$F(1,17.18) = 0.88$; $p = 0.360$] or as a function of relative position within
the complexes [$F(5,54.24) = 1.09$; $p = 0.376$], but the effect of relative position was
marginally greater for low complexes compared to high complexes [$F(5,54.24) =$
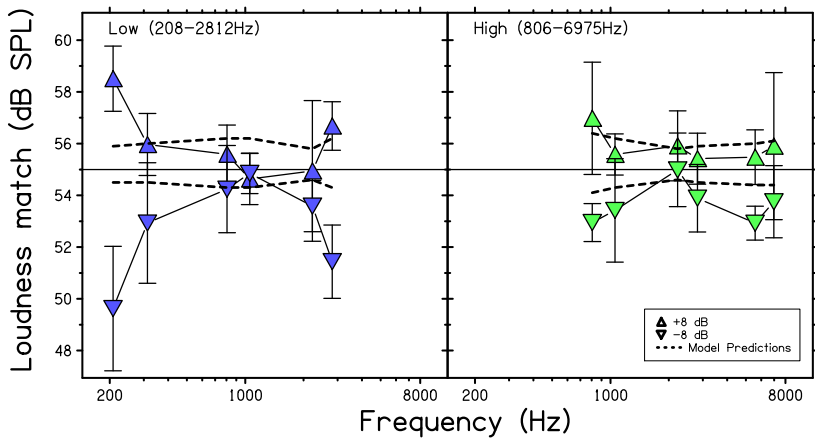2.17; $p = 0.071$].

Figure A.6: Mean levels (and their standard deviations) of comparison complexes required to match
the loudness of standard complexes in experiment 3. Upward triangles show conditions where the
standard complex included one component with an increment of 8 dB and downward triangles show
conditions with a decrement of 8 dB. Predictions of the loudness model are shown with dashed lines.

The shape of the loudness matching pattern is consistent with the shape of
the pattern of perceptual weights observed in experiment 1 and the earlier studies
(Jesteadt et al. 2014 ; Leibold et al. 2007). The edge components that receive higher
weight also produced a larger change in the overall loudness, compared to the
components in the middle of the same complex. Statistical analyses incorporating
both perceptual weights and matching data, however, failed to show any significant

relation between the two measures.

A.5   Discussion

Perceptual weights for loudness were obtained for two different sets of stimuli in experiment 1. Even though the task used to obtain the perceptual weights was subjective in the perspective of the listener, the presence of an average difference in level between the two observation intervals made it possible to designate a "correct" and an "incorrect" interval. In the correct interval, the level of each component was drawn from a distribution whose mean was $\mu + \Delta$ dB [Fig. A.1(B)], with $\Delta = 3$ dB. The only difference between a pure "sample discrimination" task and the task used in experiment 1 is that listeners did not receive feedback regarding the "correct" answer. Jesteadt et al. (2014) showed that the reliability of listeners weighting patterns could be assessed with percent correct performance in the "loudness comparison" task and that moving from loudness judgments in a task with no net difference in level between the two observation intervals to loudness comparison did not alter normalized perceptual weights. Using percent-correct responses, we calculated d' values for individual listeners for experiment 1. The mean value was 1.03 for the low complex and 1.20 for the high. For further comparison, we calculated performance of the ideal observer using Equation A.2. In this equation, N is the number of components, Δ is the increment in dB, and σ is the standard deviation of the rove [c].

$$d'_{ideal} = \sqrt{\sum_{i=1}^{N} \left(\frac{\Delta_i}{\sigma_i} \right)^2} \tag{A.2}$$

For a given a rectangular distribution with a range of 16 dB, $\sigma = 4.62$ and Eq. (2) predicts $d' = 2.05$ for both of the tone-complexes. The difference between the

---

[c] Equation (2) is not strictly correct because the stimulus distributions were uniform rather than Gaussian. Because the calculation involves 20 such distributions, however, 10 in each observation interval, use of a Gaussian approximation is appropriate.

predicted and the observed performance could be attributed to either the larger amount of internal noise in listeners or lack of feedback. A cube root rule (Lutfi and Gilbertson 2011) predicts $d' = 1.40$, still higher than the observed values.

Performance may be lower in this task than in other sample discrimination tasks because the components are too close to one another in frequency to be processed independently by the auditory system. This would also contribute noise to the assessment of perceptual weights because changes in the effective level of any given component would be altered by the changes in level of adjacent components. To test this, a second set of weights predicted by the model was obtained by using a version of the model developed by Chen et al. (2011) to estimate the "excitation" in dB for the auditory filter centered on each component, for every trial each listener received. The model of Chen et al. (2011) provides direct access to the estimated excitation pattern but is similar to the model of Moore and Glasberg (2004) in other respects. Weights were then obtained by using the difference between the excitation value for each component between interval 2 and interval 1 to predict the difference in overall loudness estimated by the model for the two intervals. This allowed us to use values that represented an "internal representation" of the stimuli as predicted by the model rather than differences in input level. The weights calculated using this method did not differ from the weights obtained using the level difference between the two intervals, and therefore are not shown. A similar strategy to calculate the perceptual weights using excitation patterns has been used previously by Lentz (2006) . Her study also reported no difference between the weights calculated using excitation patterns and the weights calculated using physical levels of the individual components.

A.5.1   Specific Loudness and Perceptual Weights

Although a calculation of specific loudness is an important stage in modelling the overall loudness of broadband sounds, direct measurements of specific loudness have not been possible. Previously, contributions of frequency components have been measured by matching the loudness of band-limited signals to the loudness of various reference signals. For example, Pollack (1952) determined the loudness-

level of noise bands of various bandwidths and frequency regions and concluded
that the mid-frequencies make larger contributions to the overall loudness of broad-
band sounds than high or low frequencies. Equal loudness contours are derived
using a similar method and show that higher levels of low frequency signals are
required to match the loudness of reference mid frequency signals. These results
are generally interpreted as indicating that low frequencies carry lower specific
loudness than the mid-frequencies, when presented at equal levels. This notion
contrasts with the results of recent studies that have obtained relative weights using
molecular psychophysical methods in loudness comparison tasks (Jesteadt et al.,
2014; Oberfeld et al., 2012). Oberfeld et al. (2012) considered effects of differences
in slopes of loudness functions used to predict the spectral weights for loudness of
broadband sounds and found that the steeper loudness functions cannot account
for the "edge" emphasis observed. Jesteadt et al. (2014), based on their model
analysis, suggested that spectral weights obtained in a loudness comparison task
may represent the relative specific loudness of the spectral components. To test
this, we obtained loudness matches between equal-level multi-tone complexes
and complexes where the level of a test component was either increased or de-
creased by 8 dB. As seen in Fig. A.6, the results of this experiment are qualitatively
consistent with the perceptual weights obtained in experiment 1. A change in the
level of the components that receive higher weight also results in greater change
in the loudness of the overall complex. These results provide a foundation for
interpretation of the perceptual weights obtained in this and previous studies and
confirm the hypothesis of Jesteadt et al. (2014) that perceptual weights reflect
relative specific loudness of spectral components of broadband sounds.

The edge frequencies in multi-tone stimuli have been shown to be either more
salient or less salient depending on the psychophysical task. Moore and Ohgushi
(1993) found that detecting a presence of mistuned frequencies near the edge of
an inharmonic tone complex was easier than in the middle of the tone complex.
In a profile analysis task, however, listeners performed more poorly when a signal
was added to the edge frequencies than the central frequencies of the complex
tones (Green and Berg, 1991). Results presented here demonstrate edge effects in
the contribution of individual components to the loudness of broadband sounds.

The masked thresholds obtained in experiment 2 rule out the effect of audibility of the edge components as a reason for such effects. The edge emphasis observed in perceptual weights from experiment 1 and loudness matching results from experiment 3 suggest some mechanism that enhances the contribution of spectral edges to the loudness of broadband sounds.

A.5.2   Peripheral vs Central Auditory Processing

Loudness is commonly described as a product of the overall neural excitation produced by a sound. Importantly, this definition does not specify a stage in the auditory system at which the loudness is being calculated from the excitation. Many investigators have attempted to relate psychophysical loudness functions to input-output characteristics of the cochlea. It is reasonable to assume that the perceived loudness reflects the effective output of cochlear processing, and therefore these studies have achieved a certain amount of success in relating loudness functions for "narrowband signals" to specific physiological measures such as otoacoustic emissions and auditory brainstem responses (Epstein and Silva 2009; Neely et al. 2003; Rasetshwane et al. 2013; Schlauch et al. 1998; Thorson et al. 2012). These schemes, however, have never been tested for signals that are broadband in nature. Earlier studies have shown that auditory nerve responses cannot account for the trends observed spectral summation of loudness (Pickles 1983; Relkin and Doucet 1997). Recent neuroimaging studies using fMRI to investigate the neural correlates of loudness perception found that the auditory brainstem activation showed a linear relationship with the physical levels of the broadband signals while the nonlinearities observed in spectral summation were only seen in the auditory cortex activation (Langers et al., 2007 ; Röhl et al., 2011 ; Röhl and Uppenkamp, 2012). They concluded that the spectral summation of loudness takes place in the auditory cortex. Ernst et al. (2008) presented a tone in background noise at several signal-to-noise ratios and measured cortical activation at various regions of interest using fMRI. Their data showed that separate regions were activated when the tone was at a level close to its threshold (lateral Heschl's gyrus) compared to above threshold (Planum temporale). They concluded that the activation in the

Planum temporale was likely related to the increase in specific loudness of the tone presented in noise.

A motivation for this study was to explore the use of perceptual weights as a more objective measure of loudness for broadband sounds that might be clinically useful for the fitting of hearing aids. The results suggest that measures of the loudness of narrowband signals presented in isolation may not be applicable to real world sounds, which are broadband in nature. Our results support the hypothesis of Leibold et al. (2007) that the perceptual weights for widely spaced stimuli reflect the influence of more "central" processing stages that are involved in spectral summation of loudness. It is unclear, however, at what stage in the auditory pathway this dichotomous coding of loudness may arise. Further investigation in this matter would require physiological recordings at more central stages of auditory processing that might be responsible for the weighting of loudness across frequencies.

A.5.3   Implications for Models of Loudness and Intensity Perception

The matching data suggest that loudness changes produced by a component as part of a larger broadband complex are significantly different than when measured for that same component in isolation. Further experiments that can explicitly test the effect of such spectral context on loudness of the individual components may allow us to understand the rules that govern loudness of broadband sounds. In light of these results, it becomes clear why loudness models can reproduce the equal loudness contours and slopes of the loudness functions with accuracy, yet fail to reproduce the perceptual weights observed in the data. The models have been calibrated with data that predominantly reflect the loudness of narrowband sounds in isolation. Moore's model has been revised to reflect more realistic assumptions regarding nonlinearities in cochlear processing (Chen et al. 2011), yet these modifications do not improve the predictions of perceptual weights. This supports our hypothesis that the contribution of the individual components to overall loudness is influenced by additional processes central to the auditory nerve,

aspects that are missing in the loudness models. Inclusion of a more central stage that will account for changes in contributions to loudness of frequencies presented in the context the other sounds may improve the predictions of the loudness model over a wide range of conditions.

A.6   Conclusions

An overall goal of the experiments reported here was to further investigate the feasibility of using perceptual weights to assess the relative contribution of specific frequency bands to the overall loudness of broadband sounds. Psychophysical data and model simulations are presented in three experiments. Based on the data reported here and the review of the literature, we conclude as follows:

- Specific loudness is weighted across frequency channels before spectral integration of loudness.

- This weighting is dependent on the relative position in the complex signal and not the absolute frequency of the components.

- Such weighting is likely due to more central stages of auditory processing that provide gain to the spectral edges.

- Perceptual weights provide a reliable and fast estimate of the quantity that is referred to as specific loudness.

- Loudness models fail to predict these weights, possibly due to a lack of central processing stages.

Acknowledgements

References

ANSI, A. (2007). "S3. 4–2007. Procedure for the Computation of Loudness of Steady Sounds". In: *American National Standards Institute, New York*.

Berg, B. G. (1990). "Observer efficiency and weights in a multiple observation task." In: *The Journal of the Acoustical Society of America (JASA)* 88.1, pp. 149–58.

Chalupper, J. and H. Fastl (2002). "Dynamic Loudness Model (DLM) for Normal and Hearing-Impaired Listeners". In: *Acta Acustica united with Acustica* 88, pp. 378–386.

Florentine, M. (2011). "Loudness". In: *Loudness*. Springer, pp. 1–15.

Glasberg, B. R. and B. C. J. Moore (2006). "Prediction of absolute thresholds and equal-loudness contours using a modified loudness model". In: *The Journal of the Acoustical Society of America (JASA)* 120.2, p. 585.

Glasberg, B. R. and B. C. Moore (1990). "Derivation of auditory filter shapes from notched-noise data". In: *Hearing Research* 47.1-2, pp. 103–138.

Hellman, R. P, B Scharf, M Teghtsoonian, and R Teghtsoonian (1987). "On the relation between the growth of loudness and the discrimination of intensity for pure tones". In: *The Journal of the Acoustical Society of America (JASA)* 82, pp. 448–453.

Hellman, R. and J. Zwislocki (1968). "Loudness determination at low sound frequencies". In: *The Journal of the Acoustical Society of America (JASA)* 43.1, pp. 60–64.

Jesteadt, W, C. C. Wier, and D. M. Green (1977). "Intensity discrimination as a function of frequency and sensation level." In: *The Journal of the Acoustical Society of America (JASA)* 61.1, pp. 169–177.

Jesteadt, W. and L. J. Leibold (2011). "Loudness in the laboratory, Part I: Steady-state sounds". In: *Loudness*. Springer, pp. 109–144.

Jesteadt, W., D. L. Valente, S. N. Joshi, and K. K. Schmid (2014). "Perceptual weights for loudness judgments of six-tone complexes[a]". In: *The Journal of the Acoustical Society of America (JASA)* 136.2, pp. 728–735.

Kochkin, S. (2010). "MarkeTrak VIII: Consumer satisfaction with hearing aids is slowly increasing". In: *Hearing Journal* 63.1, pp. 19–20.

Kortekaas, R., S Buus, and M. Florentine (2003). "Perceptual weights in auditory level discrimination." In: *The Journal of the Acoustical Society of America (JASA)* 113.6, pp. 3306–3322.

Leibold, L. J., H. Tan, S. Khaddam, and W. Jesteadt (2007). "Contributions of individual components to the overall loudness of a multitone complex[a])". In: *The Journal of the Acoustical Society of America (JASA)* 121.5, pp. 2822–2831.

Leibold, L. J., H. Tan, and W. Jesteadt (2009). "Spectral weights for sample discrimination as a function of overall level." In: *The Journal of the Acoustical Society of America (JASA)* 125.1, pp. 339–346.

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

Lutfi, R. A. (1995). "Correlation coefficients and correlation ratios as estimates of observer weights in multiple-observation tasks". In: *The Journal of the Acoustical Society of America (JASA)* 97.2, p. 1333.

Lutfi, R. A. and W. Jesteadt (2006). "Molecular analysis of the effect of relative tone level on multitone pattern discrimination". In: *The Journal of the Acoustical Society of America (JASA)* 120.6, pp. 3853–3860.

Moore, B. C. and K Ohgushi (1993). "Audibility of partials in inharmonic complex tones." In: *The Journal of the Acoustical Society of America (JASA)* 93.1, pp. 452–461.

Moore, B. C. J. and B. R. Glasberg (2004). "A revised model of loudness perception applied to cochlear hearing loss". In: *Hearing Research* 188.1-2, pp. 70–88.

Moore, B. C. J. and A. Sęk (2012). "Comparison of the CAM2 and NAL-NL2 Hearing Aid Fitting Methods". In: *Ear & Hearing* 34.1, pp. 83–95.

Moore, B. C., B. R. Glasberg, and T. Baer (1997). "A model for the prediction of thresholds, loudness, and partial loudness". In: *The Journal of the Audio Engineering Society (AES)* 45, pp. 224–240.

Oberfeld, D., W. Heeren, J. Rennies, and J. Verhey (2012). "Spectro-Temporal Weighting of Loudness". In: *PLoS One* 7.11.

Oberfeld, D., M. Kuta, and W. Jesteadt (2013). "Factors limiting performance in a multitone intensity-discrimination task: Disentangling non-optimal decision weights and increased internal noise". In: *PLoS One* 8.11.

Richards, V. M. and S. Zhu (1994). "Relative estimates of combination weights, decision criteria, and internal noise based on correlation coefficients". In: *The Journal of the Acoustical Society of America (JASA)* 95.1, pp. 423–434.

Scharf, B. (1978). "Loudness". In: *Handbook of perception* 4, pp. 187–242.

Turner, M. D. and B. G. Berg (2007). "Temporal limits of level dominance in a sample discrimination task". In: *The Journal of the Acoustical Society of America (JASA)* 121.4.

Westfall, P., R. Tobias, and R. Wolfinger (2011). *Multiple Comparisons and Multiple Tests Using SAS*.

Zwicker, E and B Scharf (1965). "A model of loudness summation." In: *Psychological Review* 72.1, pp. 3–26.

Zwicker, E. (1958). "Über psychologische und methodische Grundlagen der Lautheit". In: *Acustica* 8.1, pp. 237–258.

Zwicker, E., G Flottorp, and S. S. Stevens (1957). "Critical Band Width in Loudness Summation". In: *The Journal of the Acoustical Society of America (JASA)* 29.5, pp. 548–557.

Contributions to Hearing Research

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

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

Vol. 13: *Claus Forup Corlin Jespersgaard*, Listening in adverse conditions: Masking release and effects of hearing loss, 2012.

Vol. 14: *Rémi Decorsière*, Spectrogram inversion and potential applications for hearing research, 2013.

Vol. 15: *Søren Jørgensen*, Modeling speech intelligibility based on the signal-to-noise envelope power ration, 2014.

Vol. 16: *Kasper Eskelund*, Electrophysiological assessment of audiovisual integration in speech perception, 2014.

Vol. 17: *Simon Krogholt Christiansen*, The role of temporal coherence in auditory stream segregation, 2014.

Vol. 18: *Márton Marschall*, Capturing and reproducing realistic acoustic scenes for hearing research, 2014.

Vol. 19: *Jasmina Catic*, Human sound externalization in reverberant environments, 2014.

Vol. 20: *Michał Fereczkowski*, Design and evaluation of individualized hearing-aid signal processing and fitting, 2015.

Vol. 21: *Alexandre Chabot-Leclerc*, Computational modeling of speech intelligibility in adverse conditions, 2015.

Vol. 22: *Federica Bianchi*, Pitch representations in the impaired auditory system and implications for music perception, 2016.

Vol. 23: *Johannes Zaar*, Modeling the consequences of hearing loss and hearing-aid processing on consonant perception, 2016.

Vol. 24: *Johannes Kasbach*, Correlations between physical and perceptual parameters of acoustic scenarios. Implications for auditory modelling and sound field design, 2016.

Vol. 25: *Gustav Locesi*, Characterization and prediction of perceptual consequences of individual hearing loss, 2016.

Vol. 26: *Suyash Narendra Joshi*, Modelling auditory nerve responses to electrical stimulation with cochlear implants, 2016.

The end.

To be continued…

DTU Electrical Engineering
Department of Electrical Engineering