BOSTON UNIVERSITY
COLLEGE OF ENGINEERING

Dissertation

CROSS-FREQUENCY COINCIDENCE DETECTION IN THE PROCESSING OF COMPLEX SOUNDS

by
XUEDONG ZHANG

B.E., Southeast University, P.R. China, 1995
M.S., Southeast University, P.R. China, 1998

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
2004

Approved by

First Reader _____
H. Steven Colburn, Ph.D.
Professor of Biomedical Engineering

Second Reader _____
Barbara Shinn-Cunningham, Ph.D.
Professor of Cognitive and Neural Systems and Biomedical Engineering

Third Reader _____
Gerald Kidd, Ph.D.
Associate Professor of Communication Disorders

Fourth Reader _____
David C. Mountain, Ph.D.
Professor of Biomedical Engineering

Fifth Reader _____
Laurel H. Carney, Ph.D.
Professor of Bioengineering & Neuroscience,
Institute for Sensory Research, Syracuse University

Acknowledgments

I am pleased to have this opportunity to express my gratitude for those people who have made this dissertation possible, and have made the past six years of my life as a graduate student a rich experience.

The first and foremost, I would like to thank my thesis supervisor, Dr. Laurel Carney, for her continuous support and invaluable guidance throughout the development of this dissertation. Her continual encouragement, insightful discussions, and extremely valuable detailed comments on the earlier version of this manuscript have made it possible for me to finish this dissertation in time. During these years, her enthusiasm and integral view on research, as well as her genuine concern for her students, have made a deep impression on me and set a great example for me. I owe her lots of gratitude for showing me how to do quality research, and giving me advice and friendship in many other aspects of life. Working with her in these years has been an enjoyable experience that I will never forget.

I would like to thank Steve Colburn for invaluable support during these years. He is always available when I need his help. His kindness and brilliance make the discussion with him both pleasant and intriguing. I would also like to thank the other members of my PhD committee: Barbara Shinn-Cunningham, Gerald Kidd, and David Mountain, who asked good questions in committee meeting and took effort in reading and providing valuable comments on earlier versions of this thesis.

I am also thankful to people in my lab and Boston University Hearing Research Center (HRC) for a helpful and friendly atmosphere. They are Chris Mason (sorry, I missed the birthday cake), Michael Heinz, Susan Moscynski, Mary Evilsizer, Ling Zheng, Qing Tan, Paul Nelson, Neil Letendre, and Yi Zhou. In addition, weekly internet meeting with Mike Anzalone, Sean Davis, Satish, and Yan Gai in Syracuse University have made me feel not isolated.

I thank my parents for giving me the life in the first place, and their unconditional support and encouragement to pursue my interests.

Last but not least, I thank my wife Yunyun, for her love and patience during the PhD period.

CROSS-FREQUENCY COINCIDENCE DETECTION IN THE PROCESSING OF COMPLEX SOUNDS

(Order No.)

XUEDONG ZHANG

Boston University College of Engineering, 2004

Major Professor: H. Steve Colburn, Professor of Biomedical Engineering

Thesis Supervisor: Laurel H. Carney, Syracuse University

ABSTRACT

Responses of coincidence-detecting neurons are a direct function of the temporal structure of their input patterns. Quantitative studies of coincidence-detection provide insight into how neural processing of temporal information contributes to psychophysical performance. This study explored in detail the response properties of model coincidence-detection cells that receive inputs from auditory-nerve (AN) fibers. It also focused on the role of these model cells in coding of complex sounds related to psychophysical tasks for which temporal cues are believed to be important.

Performance of model cells was evaluated quantitatively for different model parameters, including the width of the coincidence window, the number of input AN fibers, the characteristic frequencies (CFs) of the input AN fibers, and mixed strengths of the inputs. Results suggest that model cells with low CFs are very sensitive to the phase relationship of the input AN responses. The response properties of the model cells were also compared with results of physiological studies, and the coincidence-detection model predicts several response properties that were previously believed to be difficult to explain.

Models for psychophysical detection and discrimination were designed based on population responses of model coincidence cells. Quantitative predictions of masked detection suggest that the most sensitive model cells for detection are the cells whose input AN responses are out of phase when a tone is added to the noise. The temporal structure in AN responses changes with signal-to-noise ratio and does not change as the overall level changes; thus, this model predicts psychophysical performance better than energy-based models under conditions in which the overall level of the stimulus varies randomly from trial to trial. The comparison of the coincidence-detection model and models based on other cues (e.g. envelope detector and channel theory) and implications for the theory of complex sound processing are also discussed.

Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Goals

The problem of understanding information processing in the auditory system, or any biological system, is usually approached from a multidisciplinary perspective that incorporates an integrated study of behavior, neurophysiology and computational modeling. Psychophysical data demonstrate the computational problems, or specific auditory functions, that require explanation, and neurophysiological analysis constrains the possible solutions to a given computational problem. Computational modeling has its unique and significant role among these approaches. A successful computational model should be sufficiently specified by the underlying physiological mechanisms and thus should provide insight into the functional significance of the neural circuit being studied. It should also yield computational results that are comparable to the psychophysical data and thus bridge our understanding of the relationship between auditory function and the underlying implementation in the form of neural mechanisms.

The modeling study presented here explores the potential importance of a physiologically realistic mechanism, coincidence-detection, from the perspectives of both physiology and psychophysics. The fundamental questions that this study addresses are how the coincidence-detection mechanism affects the physiological response properties of a neural circuit, and how models based on the coincidence-detection mechanism could be used to account for the psychophysical performance associated with complex sound processing. After a brief summary of previous work on this topic, a description of how each chapter relates to the overall goals of this dissertation is provided in this introduction.

1.2 Background

1.2.1 Coincidence detection: Theoretical significance and neurophysiological evidence

Recent studies show that information can be represented by the coherent patterns in the responses of different neurons (Singer and Gray, 1995; Gray 1999; Roy and Alloway, 2001). Temporal correlations among different neurons are crucial for processing if information is coded in the temporal structure of neural activity. Interest in coincidence detection is aroused because such temporal codes can be employed by the nervous system if neurons are sensitive to coincidence (Abeles, 1982; Konig et al., 1996). Coincidence-detecting cells respond only to synchronously arriving spikes and most other spikes have no

effect on their responses. The coding capacity and dynamic performance of coincidence detection are different from those that occur in traditional temporal integration. Coincidence detection uses the timing of neuronal discharges and provides an elegant and highly economic way to bind distributed neurons into functionally coherent assemblies, and to select subsets of responses for further joint processing.

It is widely accepted that the time dimension in hearing is important since acoustic information is essentially carried by time-varying signals. The auditory system provides an ideal place to study how temporally structured information is processed by the central nervous system (CNS). In the auditory system, one impressive example of neurons acting as coincidence detectors and responding to temporal structure of input activity is in the medial superior olive (MSO), which is involved in binaural hearing. It is well know that interaural time differences (ITDs) between the arrival times of acoustic signals reaching the two ears provide an important cue for the spatial location of sound, and discrimination thresholds for ITD are very small (~10us, Durlach and Colburn, 1978). Physiologically, the principal neurons of the MSO receive bilateral synaptic input from the axons of spherical bushy cells of both left and right anterior ventral cochlear nuclei (AVCN), which respond precisely to the fine timing of the input stimulus. These cells act like coincidence detectors and are sensitive to tens of microseconds differences in their afferent signals (Goldberg and Brown, 1969; Yin and Chan, 1990).

Recent studies in the AVCN also show there are cells that have response properties that can be explained by a coincidence-detection mechanism. Each primary auditory-nerve (AN) fiber has an ascending branch to the AVCN in the cochlear nucleus. AVCN neurons can be classified based on morphology (e.g., spherical bushy cell; globular bushy cell; octopus cell; and multipolar/stellete cell; see review by Cant, 1992) or based on physiological response properties (e.g. primary-like (PL); primary-like-with-notch (PLn); chopper; onset lockers (O_L); onset choppers; pausers; and buildups; see Blackburn and Sachs, 1989). Nearly all the cells from the AVCN receive convergent inputs from AN fibers. Onset cells have distinctive response properties consistent with the assumption that these cells only respond to a large number of synchronized inputs. Carney (1990) showed that several neural response types in the AVCN are sensitive to temporal discharge patterns across the AN population and are consistent with a monaural cross-frequency coincidence mechanism. Studies of phase-locking in the axons of the trapezoid body (TB) show enhanced synchronization compared with AN fibers (Joris et al., 1994ab). The improvement in synchronization at low frequencies can be predicted with a coincidence-detection model that receives convergent inputs from two or more AN fibers (Joris et al., 1994a). In addition, studies of intracellular recordings in the MSO show that the membrane properties of MSO cells (Smith, 1995) are similar to those of bushy cells in the AVCN (Manis and Marx, 1991; Wu and Oertel, 1987; Rothman and Manis, 2003a), which indicates that these cells may have similar processing abilities, including coincidence detection.

1.2.2 Psychophysical studies show the importance of spatio-temporal cues in processing complex sounds

While some basic functions of the auditory system can be characterized using simple stimuli such as pure tones (e.g. level and frequency discrimination), studies of physiological and psychophysical responses to broadband stimuli have had a profound influence on the development of auditory theory. Studies of complex sound processing reveal nonlinear effects of the auditory periphery and interactions between different parts of the auditory system. Also, these studies are important for applications of signal processing because most natural stimuli are broadband.

Fletcher (1940) found that the detectability of a tone is affected only by the energy of a masking noise in a certain bandwidth; the noise outside this bandwidth does not affect the detectability of a tone. This work inspired the idea that narrow-band filtering exists in the first stage of auditory processing and a sound signal is filtered into different channels for further processing. The channel theory has been the basis of most modern models in auditory theory.

The work of Green (1988) and that of Hall and his colleagues (1984) showed that broadband sound processing is much more complex than the assumption that information is processed in different channels independently. Cross-channel comparisons are used by the subject to decode cross-channel information, and temporal fluctuations of the masker affect a subject's performance. When multiple equal-amplitude components are presented, the threshold for detecting a level increment of the center component decreases with an increasing number of components (the frequency ratio between components are the same, thus the total spectral range increases with the number of components; see Green, 1988). When envelope fluctuations are correlated across different frequency bands, the detection of a tone in a fluctuating noise masker is enhanced (Hall et al., 1984).

Further studies have explored the effects and interaction of different cues used in psychophysical tasks. Kidd et al. (1989) compared the shape of the threshold-bandwidth function for tone-in-noise detection and for tone-in-noise discrimination. In a fixed-masker-level paradigm, their results showed that level cues are enough to account for human performance, whereas in the roving-level situation, the detection process must use temporal and/or across-frequency cues to account for the results. Even in the fixed-level task, the results for different bandwidths revealed that human subjects use level-invariant cues to improve performance in the level-discrimination task.

Gilkey (1986) used a molecular approach to study human performance in detection of tones in wideband noise. He found that linear combinations of the outputs of several detectors centered at different frequencies provide better fits to the data than the output of a single channel detector. Richards and Nekrich (1993) assessed level-dependent and level-invariant cues (in a single frequency channel) for detection of a tone added to a narrowband noise. Different observers seem to employ the cues in different ways. Detection of a tone added

to noise may be performed using two or more cues, and a single observer may alter strategies as the experimenter changes stimulus parameters. These results suggest that spatio-temporal cues are important for even simple masking experiments.

1.2.3 Previous modeling work on temporal processing of complex sounds

Durlach et al. (1986) extended the traditional channel model and combined it with signal detection theory (SDT) to process broadband complex sounds. In this framework, the same optimal processor was used to discriminate overall level and spectral shape. The central processor used the inter-channel correlation to decrease the variance in the decision variable that was caused by external noise. Thus the sensitivity index, d', increased (threshold decreased) as the bandwidth of the noise increased and more channels were excited. Further investigation has concentrated on the effects of unequal variance across the channels and the bandwidth of the channels (Bernstein and Green, 1987; Lentz and Richards, 1997). Studies of pitch and timbre also provide ideas related to how the auditory system processes complex sounds (Lyon and Shamma, 1996). Feth and Stover (1987) developed a model for complex sound processing based on pitch changes in complex sounds. Changes in spectral shape often produce a noticeable change in the pitch of stimuli. They found that the pitch change at the threshold of profile analysis experiments is constant as the number of components changes, suggesting that subjects use these apparent pitch changes as a cue. Another class of complex sounds that has been a topic of renewed interest is amplitude-modulated stimuli. Studies of amplitude-modulation suggest that there are neurons tuned to different modulation rates (Kay, 1982), i.e. these neurons may act as a “modulation filter bank”, and the envelope fluctuation in each auditory filter is extracted by these modulation filters (Dau et al., 1997a).

Although these previous modeling efforts extended our understanding of the auditory system, there is much to do to construct an auditory theory of complex sound processing. There is still no successful theory that relates physiological responses to psychophysical behavior in these tasks, and the temporal correlation between different auditory filter outputs is ignored in most previous modeling studies. Furthermore, the level, temporal fluctuation and spectral shape of stimuli do not change independently in most psychophysical experiments. Thus a multiplicity of effective cues exists, and it is hard to say which cues are actually used in various auditory information processing tasks. Because the nervous system may use some fundamental operations to process information, exploring the capability of mechanisms such as coincidence detection to process different cues is important to construct a unified theory of complex sound processing. Since coincidence detection can group different neurons and process temporal information, it is possible that all of these cues could be represented by different coincidence-detecting cells, allowing further processing by the central nervous system.

1.3 Organization of this thesis

The following four chapters present results of studies of coincidence detection as a physiologically realistic mechanism for processing complex sound in the auditory nervous system. Chapters 2 and 3 focus on the study of neural response properties of coincidence-detecting cells that receive convergent AN inputs. Chapters 4 and 5 focus on the potential neural implementation of perceptual functions based on coincidence detection. Each chapter is written in the form of a journal paper, with individual introduction and discussion sections.

Chapter 2 studies the response statistics (regularity, synchronization, and output rate vs. input rate function) of an integrate-and-fire (I&F) model to stationary and non-stationary periodic input. The I&F model receives sub-thresholds inputs and thus behaves as a coincidence detector. A computational method to calculate response statistics was derived, and the relationship between model parameters and responses statistics was explored in detail.

Chapter 3 describes the response properties to pure tones and tones in noise of model coincidence-detecting cells that receive convergent AN inputs. Tuning properties, response areas and rate-level functions were simulated. The effects of the synapse configuration of input AN fibers was demonstrated by studying model cell responses with different synaptic configurations. The potential correlations between model predictions and physiological neuron responses are discussed. The relationship between different coincidence-detection models (shot-noise and cross-correlation models) is also discussed in focusing on predictions of performance for a masked-detection task.

Chapters 4 and 5 examine the processing capability of the coincidence-detection model for psychophysical masked-detection tasks. Chapter 4 focuses on tone-in-noise detection and Chapter 5 studies model performance for a level-discrimination-in-noise task. A matrix of coincidence-detecting model cells was constructed based on a population of coincidence-detection cells that received inputs from AN fibers tuned to different frequencies. The model responses for different stimulus conditions were studied first, and then the performance of each model cell was evaluated. Based on the population responses, detectors were designed, and model performance was examined by simulating the same tracking algorithms as used in psychophysical experiments.

The performance of other model detectors was also evaluated in Chapters 4 and 5. These model detectors include 1) the energy-based channel model, which was based on the output energy of several different auditory filters, 2) the envelope detector, based on the envelope statistics (peakiness) of the auditory-filter outputs. The relationship between these model predictions and possible cues used in masked detection are then discussed.

The last chapter provides some brief comments on the significance of the present work and suggests directions for future work.

Chapter 2. Response properties of an integrate-and-fire model that receives sub-threshold input

2.1 Abstract

A new computational technique for calculation of the inter-spike interval and poststimulus time (PST) histogram of an integrate-and-fire (I&F) model for arbitrary input was developed. The effects of the model parameters on the response statistics were studied systematically. Specifically, the probability distribution of the membrane potential was calculated as a function of time, and the mean inter-spike interval and PST histogram were calculated for arbitrary inputs. For stationary inputs, the regularity of the output was studied in detail for various model parameters. For non-stationary inputs, the effects of the model parameters on the output synchronization index were explored. The results show that enhanced synchronization in response to low-frequency stimuli required a large number of weak inputs, and irregular responses and a linear input-output relationship required strong (but sub-threshold) inputs with a small time constant. A single synapse configuration with mixed-amplitude inputs can respond to stationary inputs irregularly and also have enhanced synchronization to low frequency inputs. Both of these response properties have been reported in different cells in the ventral cochlear nucleus.

2.2 Background and Introduction

One fundamental question in the study of nervous-system function is how a single neuron responds to and processes the information received from other neurons. In the auditory system, approximately 30,000 primary auditory-nerve (AN) fibers with different characteristic frequencies (CF) connect the auditory sensory organ, the cochlea or inner ear, and the cochlear nucleus (CN) in the brainstem. The CN receives all the information in the acoustic signal represented by the temporally structured spike discharges in the population of AN fibers, and is the first stage of information processing in the central auditory system. The CN contains a variety of cells that differ in their responses to a relatively homogeneous input and therefore presents a unique opportunity for quantitatively studying input-output transformations by neurons and the relations between a neuron's function and its underlying mechanisms.

While the importance of spike timing in the millisecond range in cortical areas is still a topic of intense debate (e.g. König et al., 1996), the importance of temporal coding for auditory perception, especially sound localization, has been widely accepted (e.g. Joris et al., 1998). It is thus of great interest to understand temporal coding and processing along the auditory pathway and its

underlying mechanisms. Considerable progress has been made over the decades. Temporal information is first encoded in the discharge pattern of AN fibers, which are phase-locked to the temporal features of the acoustic waveform up to 4-5 kHz (Kiang *et al.*, 1965; Johnson, 1980). In the ventral division of the cochlear nucleus (VCN), the bushy cells appear to be specialized to preserve and even enhance the temporal information encoded in AN fibers (Joris *et al.*, 1994ab). Temporal information is further transmitted to the superior olivary complex where cells are sensitive to interaural timing differences from their binaural inputs.

The capability that bushy cells have to preserve or even enhance timing information relies on the synaptic configuration and the membrane properties of the cells. Bushy cells receive large somatic AN terminals, called the end bulbs of Held (see review by Cant, 1992), that differ in number and size: spherical bushy cells (SBC) have fewer and larger end bulbs, while globular bushy cells (GBC) have more and smaller end bulbs. The somatic inputs bypass the dendritic low-pass filtering and thus have a very short time constant in their synaptic current. Bushy cells also have short membrane time constants (Oertel 1983, 1985) caused by the activation of a low-threshold potassium conductance at the resting potential. Recent experimental (Manis and Marx, 1991; Rothman and Manis, 2003c) and modeling (Rothman *et al.*, 1993, 2003a,b) studies have provided a satisfactory explanation for how the membrane properties of the bushy cell contribute to its capability of preserving temporal information precisely.

Analysis of neural responses is very useful for estimating the parameters of synapse inputs (the number and size of the inputs) to bushy cells since these parameters are believed to be crucial to the bushy cell's input-output functions. SBCs usually have prepotential waveforms (Bourk, 1976; reviewed by Rhode and Greenberg, 1992) and have discharge patterns similar to those of the AN fibers, thus they are referred to as primarylike (PL), suggesting that SBCs may receive suprathreshold inputs. Modeling studies by Rothman *et al.* (1993) suggest that bushy cells with primarylike-with-notch (PLn, associated with GBCs) responses to high-frequency tones at CF must also receive suprathreshold inputs to maintain their irregularity. Evidence that bushy cells receive subthreshold inputs also is available. Some bushy cells demonstrate *Onset* discharge patterns that can be successfully modeled using many weak subthreshold inputs (Rothman *et al.*, 1993; Kipke and Levy, 1997; Kalluri and Delgutte, 2003a,b). Enhanced phase-locking has been observed (Joris *et al.*, 1994a,b) in low-CF bushy cells in the AVCN in response to CF tones and in high-CF PLn response types in response to low-frequency tones, consistent with models receiving sub-threshold inputs.

The reports that high-CF bushy cells respond irregularly to CF tones and have enhanced-sync responses to low-frequency tones (Joris et al., 1994a) suggest that the same synaptic configuration (number and size of the AN inputs) must be capable of both input-output relationships. In the studies from Rothman and his colleagues (Rothman *et al.*, 1993; Rothman and Young, 1996;

Rothman and Manis, 2003a), a compartmental model was used to explore the model responses with different synaptic configurations. Their results support the hypothesis that sub-threshold inputs are capable of producing enhanced-sync response to low-frequency tones, and suprathreshold inputs are more suitable to describe the PL and PLn responses of bushy cells. They also suggested that the different arrangements of synaptic inputs may affect the input-output relationships of bushy cells. However, the relative importance of such arrangements was not clear based on their results, and they did not elucidate whether the desired responses can be attributed to different synapse configurations without considering the complex (nonlinear) effects of neural dynamics.

Statistical analysis of neural activity based on integrate-and-fire (I&F) models to obtain estimates of physiological and anatomical parameters of neurons has a long history in theoretical neuroscience (Tuckwell, 1988) due to its simplicity and mathematical tractability. The I&F model was also the first neural model to capture the essential properties of neural behavior: synaptic integration and threshold for responding. A generalization of this simple phenomenological model (known as the Spike Response Model, see Gerstner and Kistler, 2002) can emulate more physiologically realistic Hodgkin-Huxley type (channel) models (Kistler *et al.*, 1997) and has been widely used in the study of neural coding, synaptic plasticity, and pattern formation. I&F models have been used to study the regularity properties of spontaneous activity in auditory neurons (Molnar and Pfeiffer, 1968), the phase-locking properties of bushy cells (Joris *et al.*, 1994a), and the discharge pattern of *Onset* neurons in the CN (Kalluri and Delgutte, 2003a,b). Stochastic processes have been employed in modeling the responses of single neurons using the I&F model. Stein (1965), who proposed a discontinuous Markov process as a neural model that incorporated the exponential decay of the membrane potential, provided various insights to model properties. Computational methods have provided quantitative statistical descriptions of the model response (Molnar and Pfeiffer, 1968; Colburn and Moss, 1981). Previous methods for analyzing the I&F model have been limited to conditions with stationary inputs and new techniques are needed to explore model responses to non-stationary (phase-locked) input.

Kempter *et al.* (1998) investigated the coincidence-detection properties of an I&F model in response to periodic spike inputs. Their analysis concerns how the model response rate depends on the neural parameters, such as the number of synapses, the threshold, and the time course of the postsynaptic responses. They also explored the effects of these parameters on the neuron's ability to convert a temporal code into a rate code. An extended study (Burkitt and Clark, 2001) has also been conducted to evaluate the inter-spike interval (*ISI*) histogram and the period histogram for neural responses to ongoing periodic inputs. Both studies assume that there is a large number of small inputs to the model and that the membrane potential is approximated by a Gaussian random variable, and they limit their analyses to a model without refractoriness.

In the present study, we explored how the neural response statistics change with different synapse configurations using an I&F model. The model cell received convergent AN inputs, which were superimposed and modeled as a non-stationary Poisson point process. A computational method based on Stein's model is proposed to calculate the *ISI* histogram and post-stimulus time histogram (PST) of the I&F model accurately in response to an arbitrary stimulus waveform. The method presented here applies to the I&F model without any limitations on the model parameters and is especially efficient when there is small number of inputs with fast membrane decay time constants. The model parameters were systematically investigated using responses to both stationary and non-stationary inputs. Various response properties of the model cell were explored, including the rate response of the model cell, regularity in response to stationary inputs, and phase-locking in response to non-stationary input. The general conclusions about the effect of model parameters on the neural response statistics apply to all cells that receive convergent inputs, though the statistics we investigated here are of particular interest for the study of bushy cells in the VCN.

2.3 Method

2.3.1 The I&F Model:

The I&F model used in the present study is a simple I&F neuron with the following properties:

1. Each input spike (except those that arrived during the model's dead time) at time t_k from channel i generated an EPSP given by

$$V_{i,t_k}(t) = A_i \times e^{-(t-t_k)/\tau_i} . \qquad (2.1)$$

   The amplitude (A_i) and time constant (τ_i) of the EPSP ($V_{i,tk}$) represent the basic configuration of the synapse integration and were explored systematically, along with different input stimuli.
2. The membrane potential $V(t)$ (0 at resting) was the linear summation of all incoming EPSPs.
3. The model cell fired when the membrane potential $V(t)$ exceeded the threshold. The threshold was always set to 1 (so the EPSP amplitude represented the synapse strength relative to threshold).
4. The membrane potential $V(t)$ was reset to zero after firing, with a dead time of 0.7 ms (spikes that arrived during the deadtime do not generate EPSPs).

How the membrane potential $V(t)$ changed with time after the firing was very important to the neural response statistics. When the neuron was not discharging (assuming the dead time period ended at t = 0, that is, the previous spike time is at t = -0.7 ms) and the model input was a Poisson stationary

process with arrival rate R, the mean and variance of the model's membrane potential values were given by (Stein, 1965)

$$\mu_v = RA\tau(1-e^{-t/\tau}), \tag{2.2}$$

and

$$\sigma_v^2 = RA^2(\tau/2)(1-e^{-2t/\tau}). \tag{2.3}$$

2.3.2 Stimulus description and superposition of AN inputs

The discharge pattern of the AN fiber can be described as a nonhomogeneous Poisson process modified to include refractory effects (Johnson and Swami, 1983). Since the EPSPs are integrated linearly, inputs from multiple AN fibers that produce EPSPs with identical amplitudes and time constants can be superimposed. The equivalent input can then be described by a nonhomogeneous Poisson process (Cox, 1962) as the number of input fibers increases. Figure 2-1 illustrates the change of the inter-spike interval (ISI) distribution of the superimposed input with different numbers of input AN fibers. Each model AN fiber's discharge times were produced by a renewal process that simulated a stationary input (100 sp/sec) modified by refractoriness (Carney, 1993). The ISI curves (calculated based on 100,000 simulated spikes) are plotted on normalized axes so that they are comparable with each other (the solid line represents the ISI for a Poisson process). The simulation shows that the superimposed input could be approximated by a Poisson process (i.e. the effect of input refractoriness on ISI could be ignored) when there were more than five independent AN-fiber inputs.

For the model cell that received stationary input from multiple AN fibers, we treated the total input spike train as a Poisson process with rate R. When the input was periodic, the total input spike train was described as a nonhomogeneous Poisson process with an instantaneous rate of firing $s_{AN}(t)$ given by (see Colburn, 1973; Colburn et al., 2003)

$$s_{AN}(t) = R\frac{e^{\varphi\sin(2\pi ft)}}{f\int_0^{1/f} e^{\varphi\sin(2\pi ft)}dt}, \tag{2.4}$$

where the exponential function in the numerator represents the periodic signal, with φ and f determining the strength of phase locking and frequency of the input. The parameter R is the mean firing rate of the non-stationary Poisson process (the exponential function is normalized

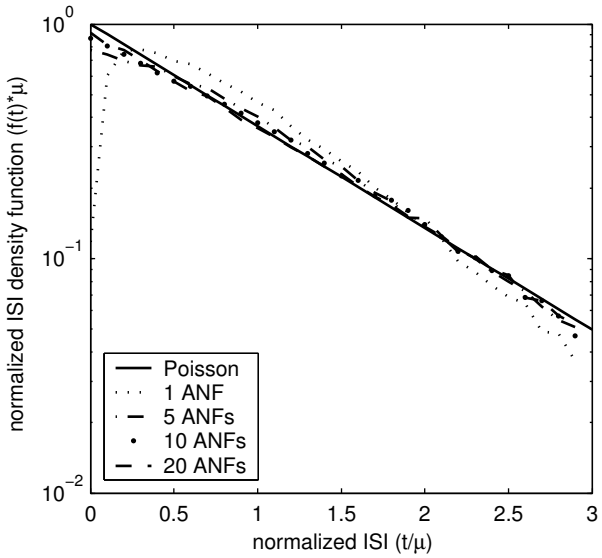


Figure 2-1: *ISI* (inter-spike interval) distribution of a superposition model with different numbers of independent AN inputs. Each input model AN fiber had a stationary Poisson response of 100 sp/sec modified by refractoriness (Carney, 1993). The input spikes were interleaved, and the *ISI* was calculated based on 10,000 simulated discharges. Both axes were normalized (either multiplied or divided by the mean interval-spike time μ) to make the *ISI* distributions comparable. The simulation shows that as the number of independent inputs increased (to values larger than 5), the combined input spikes could be approximated by a simple Poisson process (solid line).

by the denominator, which is the modified Bessel function $I_0[g]$ described in Colburn et al., 2003).

2.3.3 Analytical calculations of the ISI and PST for the I&F model with stationary and non-stationary inputs

Stein (1965) proposed a discontinuous Markov process model to describe the statistics of the membrane potential for the I&F model mentioned above. Molnar and Pfeiffer (1968) used this model to numerically (computationally) calculate the ISI of the output for the case with stationary input. The following analysis and computational results extend this method to include both the ISI and PST of the model output.

We define $F[V_x,t]$ as the cumulative probability that the membrane potential $V(t)$ is less than potential V_x at time t: $F[V_x,t]$=Prob ($V(t) \le V_x$), and $F_c[V_x,t]$ as the conditional cumulative distribution for this probability given that the voltage does not exceed the threshold of 1 at time *t*. For both distributions, it is assumed that the previous output spike time is at t=0 and the potential $V(t')$ is always less than 1 for $0<t'<t$. Thus, $F[V_x,t]$ can be derived from the distribution $F_c[V_x,\cdot]$ before time *t*. For a stationary Poisson input with rate *R*, the probability of an input spike occurring in a short time interval Δ can be represented as $R\Delta$. We then can express the function $F[V_x, t+\Delta]$ in terms of $F_c[V_x, t]$ based on the transition of the Markov process model (Stein, 1965) as

$$F[V_x,t+\Delta] = (1-R\Delta)F_c[V_x e^{\Delta/\tau},t] + R\Delta F_c[V_x e^{\Delta/\tau} - A,t]\,, \qquad (2.5)$$

where *A* and τ are the amplitude and time constant of the input EPSP.

For a threshold voltage equal to 1, the probability that the model cell will have an output spike in the time interval from *t* to $t+\Delta$ is equal to $1-F[V_x=1,t]$. The relationship between the corresponding probability density functions $f[V_x,t]$ and $f_c[V_x,t]$ can be described as $f_c[V_x,t] = f[V_x,t]/F[1,t]$ for $x < 1$. $F_c[V_x,t]$ is thus given by

$$F_c[V_x,t] = F[V_x,t]/F[1,t] \text{ for all } V_x< 1, \text{ and}$$

$$F_c[V_x,t] = 1 \text{ for all } V_x >=1. \qquad (2.6)$$

The output of the I&F model can be described as a renewal process (Cox, 1962) with a hazard function $\rho(t)$, which is defined as the rate of a renewal (spike) event that occurs at time *t* and is determined by

$$\rho(t) = (1-F[1,t])/\Delta\,. \qquad (2.7)$$

The inter-spike interval (ISI) of the model output with stationary input can be specified by

$$f_{ISI}(t)=S(t)\rho(t), \qquad (2.8)$$

where $S(t)$ is the survival function of the renewal process, or the probability that there is no renewal (spike) event between 0 and t. The S(t) can be written in terms of the hazard function as

$$S(t) = e^{-\int_0^t \rho(x)dx}. \qquad (2.9)$$

The above analysis can be easily extended to the situation where the input is a nonhomogeneous Poisson process described by $R(t)$ with a previous output spike time at t_0 (The arguments of $S(\cdot)$, $\rho(\cdot)$, and $f_{ISI}(\cdot)$ are intervals since the previous spike time $t_0=0$ (in Eq. 2.5-2.9), whereas the arguments of $P(\cdot)$ and $R(\cdot)$ are times). In this case, the ISI calculated above becomes the density of the first-passage time to threshold and is represented as $f_{ISI}(t-t_0 \mid t_0)$. The unconditional firing probability $P(t)$ (which is an estimate of the PST histogram) of the model output to the input $R(t)$ can described as (Cox, 1962)

$$P(t) = \int_{-\infty}^{t} P(x) f_{ISI}(t-x \mid x)dx, \qquad (2.10)$$

where x represents the spike time before time t.

The calculation of P(t) from the above equation is not possible computationally because the duration over which the integral is computed is not limited.

We now assume that the cumulative conditional probability of the membrane potential $F_c[x,t \mid t_0]$ (where t_0 is the previous spike time) is determined by the input spikes during the preceding time period $(t-T, t)$, where $T >> \tau$. This is a reasonable assumption since the potential contributed by spikes before $t-T$ decays with a time constant τ and can be neglected compared to the potential contributed by recent spikes if $T >> \tau$. For all previous spike times for which $t_0<t-T$, the cumulative probabilities of the membrane potential, $F_c[x,t \mid t_0]$ and $F[V_x,t \mid t_0]$, can be approximated as $F_c[V_x,t \mid t-T]$ and $F[V_x,t \mid t-T]$, and the hazard function $\rho(t-t_0 \mid t_0)$ derived from equation Equation 2.7 can be approximated by $\rho(T \mid t-T)$. Together with Eq. 2.10 the unconditional firing probability $P(t)$ can be rewritten as

$$\begin{aligned} P(t) &= \int_{-\infty}^{t} P(x) f_{ISI}(t-x \mid x)dx \\ &= \int_{-\infty}^{t} P(x) S(t-x \mid x)\rho(t-x \mid x)dx \\ &\approx \int_{-\infty}^{t-T} P(x) S(t-x \mid x)\rho(T \mid t-T)dx + \int_{t-T}^{t} P(x) S(t-x \mid x)\rho(t-x \mid x)dx \\ &= \rho(T \mid t-T)\int_{-\infty}^{t-T} P(x) S(t-x \mid x)dx + \int_{t-T}^{t} P(x) S(t-x \mid x)\rho(t-x \mid x)dx \end{aligned}. \qquad (2.11)$$

The integral on the 2nd line was separated into two integrals on the 3rd line, and for all spike times previous to x < t-T, ρ(t-x|x) was approximated by ρ(T|t-T), which was the hazard function at time t given a previous spike at t-T. The second integral in the final line of the above equation has a limited duration, and thus the numerical calculation based on Eqs. 2.7-2.10 is possible. The first integral in the final line of the above equation can be further simplified as

$$
\begin{aligned}
&P_{residue}(t+\Delta) \triangleq \int_{-\infty}^{t+\Delta-T} P(x)S(t+\Delta-x\,|\,x)dx \\
&\approx \int_{-\infty}^{t-T} P(x)S(t+\Delta-x\,|\,x)dx + P(t-T)S(T\,|\,t-T)\Delta \\
&\approx \int_{-\infty}^{t-T} P(x)[S(t-x\,|\,x)+\Delta\cdot dS(t-x\,|\,x)]dx + P(t-T)S(T\,|\,t-T)\Delta \quad , \qquad (2.12) \\
&= \int_{-\infty}^{t-T} P(x)S(t-x\,|\,x)[1+\Delta\cdot\frac{dS(t-x\,|\,x)}{S(t-x\,|\,x)}]dx + P(t-T)S(T\,|\,t-T)\Delta \\
&= (1-\rho(T\,|\,t-T)\Delta)P_{residue}(t) + P(t-T)S(T\,|\,t-T)\Delta
\end{aligned}
$$

where the last step of the derivation is based on the relationship between the survival function *S*(t) and the hazard function ρ(t) (Cox, 1962),

$$\frac{dS(t)}{S} = -\rho(t)\,. \qquad (2.13)$$

The final line in Eq. 2.12 can be described by a differential equation and calculated numerically

$$dP_{residue} = -P_{residue}\rho(T\,|\,t-T) + P(t-T)S(T\,|\,t-T)dt\,. \qquad (2.14)$$

Using the above relationships, P(t) can be calculated knowing R(t) and the mean ISI for a non-stationary input from time t_1 to t_2 can be represented by

$$\overline{f}_{ISI}(t) = \frac{\int_{t1}^{t2} P(x) f_{ISI}(t\,|\,x)dx}{\int_{t1}^{t2} P(x)dx} \qquad (2.15)$$

and calculated numerically (this is not shown in detail because we are not interested in the ISI for the non-stationary input).

For a model receiving inputs having EPSPs with two different amplitudes but the same time constant, the Markov process of the I&F model (Eq. 2.5) can be described as

$$F[x,t+\Delta] = P_{00}F_c[xe^{\Delta/\tau},t] + P_{10}F_c[xe^{\Delta/\tau} - A_1,t] \\ +P_{01}F_c[xe^{\Delta/\tau} - A_2,t] + P_{11}F_c[xe^{\Delta/\tau} - A_1 - A_2,t] \quad , \qquad (2.16)$$

where P_{00}, P_{10}, P_{01}, P_{11} represent the joint probability of input spikes from two channels with different EPSP amplitudes (A_1 and A_2) in the interval from t to $t+\Delta$. The derivation above can then be extended to calculate the PST and ISI of the model response to arbitrary inputs with mixed amplitude EPSPs. The same technique can be applied to allow multiple spikes to arrive in a time window Δ (such that a large Δ can be used to approximate the Poisson process), making this computation more efficient.

2.4 Results

2.4.1 Predictions for a model that receives stationary inputs

The steady-state response of an AN fiber to a CF tone at a high frequency is generally assumed to be a stationary point process (Siebert, 1964; Kiang et al., 1965). The response of a neuron receiving stationary inputs can be modeled successfully as a stationary renewal process fully characterized by the inter-spike interval of the mean, μ, and standard deviation, σ, of the process (Cox, 1962). The mean rate of the model output is defined as

$$\text{Rate} = 1/\text{mean interval} = 1/\mu,$$

and the quantitative measure of the response regularity is described by the coefficient of variation (*CV*)

$$CV = \sigma/\mu.$$

This regularity measure of the cell response is important since it may represent different underlying processing mechanisms, and it has been used as one of the criteria to classify different unit types in the CN (Young et al., 1988; Blackburn and Sachs, 1989). A cell with a *CV* value close to 1 is considered irregular, and its response can be treated as a process essentially similar to the Poisson process ($\sigma=\mu$). It is more realistic to model a cell with a deadtime-modified (τ_d) Poisson process, and the measure of *CV* for such a process is affected by the response firing rate R_{out} ($1/R_{out} = \mu = \sigma + \tau_d$, and $CV = \sigma/\mu = 1 - R_{out}\tau_d$) (Rothman et al., 1993). To reflect the more fundamental nature of the underlying process, the modified coefficient of variation (*CV'*) of the cell responses (Rothman et al., 1993) is used as a measure of the cell regularity

$$CV' = \sigma/(\mu-\tau_d). \qquad (2.17)$$

The mean rate and *CV'* measurements of the I&F model responses to stationary inputs with various model parameters are shown in Figs. 2.2-2.3. The input to the I&F model was a stationary Poisson process, and the calculations were based on the equations described in Section 2.

Figure 2-2 illustrates the effect of the time constant (τ) of the EPSP (with a fixed amplitude of 1/3) on the model cell responses. The response rate of the model cell was plotted as a function of input strength (input discharge rate R multiplied by the EPSP amplitude) in Fig. 2-2(a). With a large time constant, the model output was more affected by the integration (energy) of the input EPSPs, and the model response rate changed more linearly with the input strength. When the time constant was short, the model cell's response was dominated by the coincidence-detection mechanism, and the probability of discharge in an effective time window w was approximated by (Stein, 1965)

$$P_f = \frac{R^k w^{k-1} e^{-Rw}}{(k-1)!}, \qquad (2.18)$$

where k equals the number of input spikes that are required to arrive within the time window w (proportional to τ) to generate an output spike. The model response rate increased rapidly (and nonlinearly) as the input rate R increased.

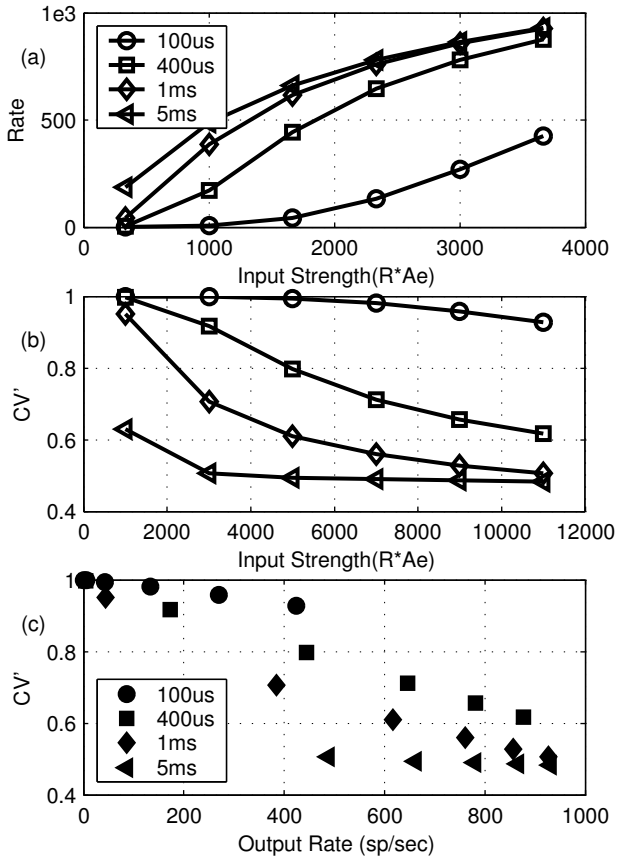


Figure 2-2: Responses for models with different EPSP time constants to stationary inputs. The EPSP amplitude was fixed at 1/3 for all models. The abscissa in (a) and (b) is input strength, which is defined as input rate multiplied by the EPSP amplitude. (a) Model response rate as a function of input strength. For high input strength, the output rate was limited by the deadtime of I&F model. For small input strength, input-output rate function was determined by the EPSP time constant. The model rate responses initially changed linearly with input strength for large EPSP time constants and increased nonlinearly for small EPSP time constants. (b) Regularity measure (*CV'*) of model response as a function of input strength. The *CV'* generally dropped as the input strength increased, but remained high (> 0.65) for model cells with short time constants (up to 400 μs). (c) Regularity measure replotted as a function of model response rate. The increase in irregularity of the model response was caused by the EPSP time constant and not by the drop in the model response rate.

The regularity measure (*CV'*) of the model response is plotted in Fig. 2-2(b) as a function of input strength. The *CV'* of the model response decreased as the input rate increased for all time constants. However, model cells with short-time constants (up to 400 μs) were still classified as irregular (*CV'* of the model response was higher than 0.65). When the time constant of the model cell was long, the model cell discharged more regularly (with small *CV'*). This is because the probability of discharge of the model cell was greatly affected by the integration time constant of the membrane potential after the deadtime, as the mean and variance of the model potential increased slowly after the deadtime (See Eq. 2.2 and Eq. 2.3). Figure 2-2(c) shows the regularity measure of the model cell response as a function of the model response rate. For a fixed output response rate, model cells with large time constants were more regular than cells with short time constants, showing that the decrease of the irregularity in Fig.2-2(b) with increasing EPSP time constant for a fixed input rate was not caused by the increase of the response rate.

Model responses for different amplitude input EPSPs are shown in Fig. 2-3. The time constant of the model EPSP was fixed at 100 μs and the amplitude was always below threshold (otherwise the model response process would have been the same as the input process, modified by refractoriness). The model response rate as a function of input strength is plotted in Fig. 2-3(a). Since the input strength is defined as EPSP amplitude multiplied by the input rate, the input rate for each model cell (with different amplitude EPSP) was different at each abscissa value, but the total energy of the input was the same. The response rate of the

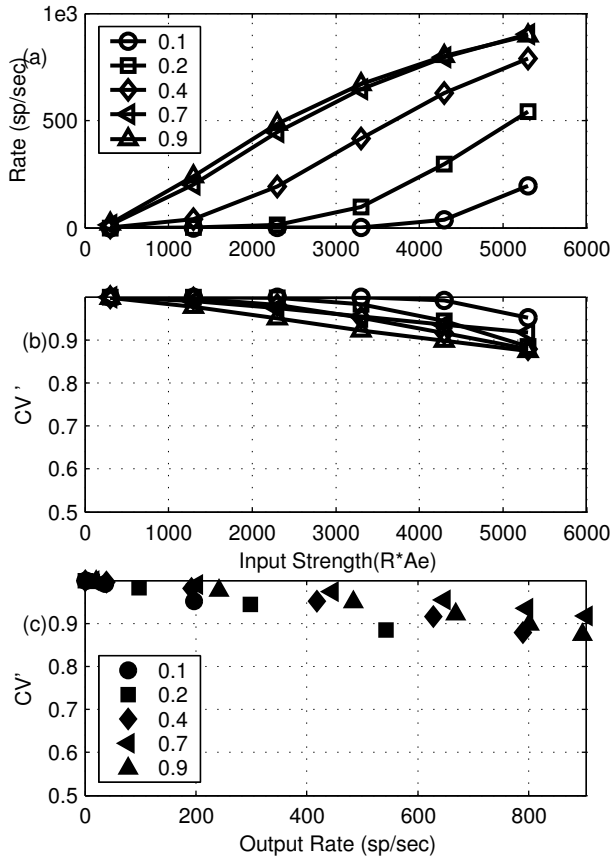


Figure 2-3: Responses of models receiving stationary inputs with different EPSP amplitudes and a fixed time constant (100 μs). Data are plotted in the same way as in Fig. 2-2. (a) The input-output rate function was more linear for models with large EPSP amplitude and more nonlinear for models with small EPSP amplitude. (b) Regularity measure *CV'* of model responses. The change of model EPSP amplitude did not affect the *CV'* of the model responses; the *CV'* remained high, presumably because of the short EPSP time constant used in the computation. (c) *CV'* replotted as a function of model response rate. When model cells had same output rate, there was no clear monotonic relationship between *CV'* and the model EPSP amplitude.

model output increased as the input strength increased, however, the model cells with small inputs required greater input strength to generate the same response rate. Weak inputs required a larger number (k) of input spikes in an effective time window w and the rate tended to change nonlinearly as the input strength increased, as expected from Eq. 2.18. The CV' of the model response (Fig. 2-3(b)) dropped as the input strength increased, but remained high for all model cells, regardless of EPSP amplitude. When the regularity measure is plotted as a function of the output response rate in Fig. 2-3(c), it is clear that there is not a simple relationship between CV' and the strength of the synapse input (amplitude of the EPSP). The CV' changed non-monotonically with increased amplitude of the model EPSPs. This result shows that the regularity of model cells that received sub-threshold inputs was determined primarily by the time constant of the input EPSPs.

2.4.2 Predictions for the model that receives synchronized input

The most prominent feature of AN fiber responses to low-frequency tones is that the discharges phase lock to the stimulus frequency up to about 5 kHz (Johnson, 1980). Enhanced phaselocking has been reported in VCN bushy cells (Joris et al. 1994) and can be modeled as a consequence of converging sub-threshold AN inputs (Joris et al., 1993; Rothman et al. 1996; Rothman et al., 2003). The combined input from convergent AN discharges to the I&F model was represented by a single non-stationary (periodic) Poisson process, as described in association with Eq. 2.4. The PST of the model response to such an input is also periodic and can be calculated numerically based on the methods described in Section 2.3. The degree of phase locking of the model response was quantified by the synchronization index (SI), which is defined as $SI = B/A$, where B is the fundamental frequency (stimulus frequency) component and A is the DC component of the Fourier series of the response PST histogram (Johnson, 1980).

The responses for I&F models with different time constants are plotted as a function of input synchronization index (SI) in Fig. 2-4. The model EPSP amplitude was fixed at 1/3, and the combined input had a constant rate of 5400 spikes/sec. The input stimulus had a frequency of 500 Hz, and its SI was varied systematically. The model rate responses are illustrated in the top panel of Fig. 2-4. The model response with a short time constant changed dramatically when the input SI increased (the input spikes were more synchronized). For a large time constant, the model response rate mainly depended on the total energy of the input and did not change dramatically as the input synchronization index changed (in fact, the rate dropped as more input spikes arrived in the deadtime period of the model and did not contribute to the model potential). The SI measure of the output response is plotted in the bottom panel of Fig. 2-4; the dotted line is the result for which the output SI equals the input SI. The model responses with short time constants had more enhanced synchronization than model responses with large time constants,

and the *SI* measure was not affected by model time constant when the inputs were highly synchronized.

Figure 2-5 is similar to Fig. 2-4 except that the stimulus frequency was 2000Hz. With a high input frequency, the response rate of the model

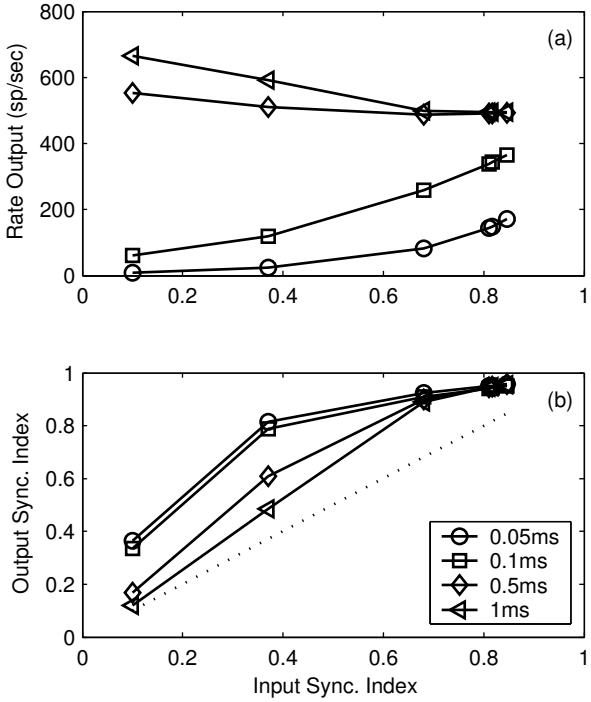


Figure 2-4: Responses for models with different EPSP time constants (see legend) and a fixed amplitude EPSP (1/3) to non-stationary inputs. The input waveform had a frequency of 500 Hz and a fixed average rate of 5400 spikes/sec (Eq. 2.4). The results were plotted as a function of input synchronization index (*SI*). (a) Rate responses for different models. For a small time constant (see legend), the model response rate increased dramatically when the input *SI* increased (a timing code was converted to a rate code in this situation). For a large time constant, the model response rate did not change much as the input *SI* changed. In fact, the response rate dropped as more input spikes arrived in the deadtime period of the model and did not contribute to the model potential. (b) The *SI* measure of the model output; the dotted line represents the output *SI* equaling the input *SI*. The model responses with short time constants had more enhanced synchronization than model responses with large time constant, and the *SI* measure is nearly independent of model time constants when the inputs were highly synchronized.

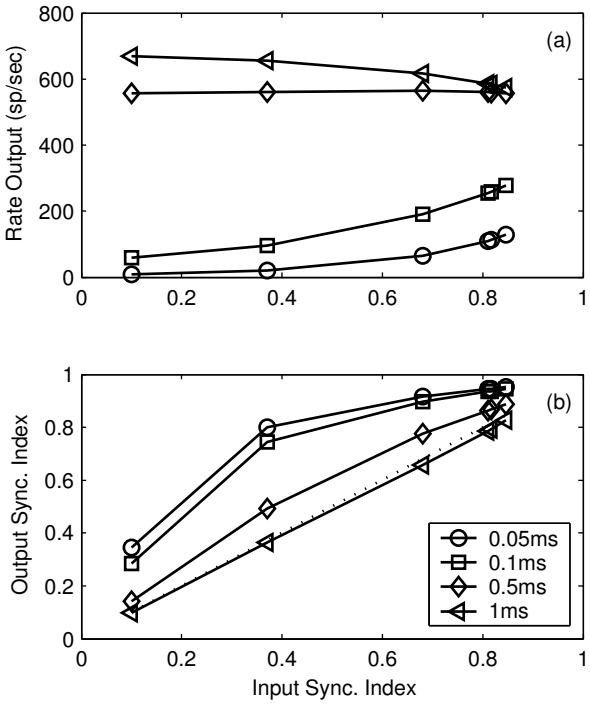


Figure 2-5: Similar plot to Fig. 2-4 except that the stimulus frequency was 2000 Hz. The output *SI* measure was greatly affected by the model time constant for both highly and weakly synchronized inputs.

was similar to the result in Fig. 2-3, but the output *SI* measure was affected by the model time constant for both highly and weakly synchronized inputs. The response of the model with a large time constant (left triangle) had a degraded synchronization compare to the input *SI* measure (dotted line). The reduction of synchronization due to the large time constant was most effective at high frequencies. This may explain, in part, the physiological observation that the *SI* of the CN cell is enhanced with respect to AN fibers at low frequencies but is lower than that of AN fibers at mid frequencies (Blackburn and Sachs, 1992, Joris et al, 1994).

The responses of a model with different EPSP amplitudes (with a fixed time constant of 100μs) are plotted as a function of input *SI* in Fig. 2-6. The stimulus frequency was 500 Hz, and the input strength (EPSP amplitude multiplied by the mean input rate) was fixed at 1800. With strong inputs (a large model EPSP amplitude) the model response rate (top panel) changed slowly as the input *SI* measure changed, and the synchronization enhancement (re: input *SI*, bottom panel) was also lowest for the model with the largest EPSP amplitude. When the input had a high *SI* measure (usually true for low-frequency inputs), the amplitude of the inputs had a larger effect on the output *SI* measure than did the model time constant (See Figs. 2-4 and 2-6).

The synchronization of actual AN fibers, which are the inputs to CN cells, change systematically as a function of the stimulus frequency. Model cell responses to inputs with realistic synchronization at each stimulus frequency are illustrated in Fig. 2-7. The input spike rate was

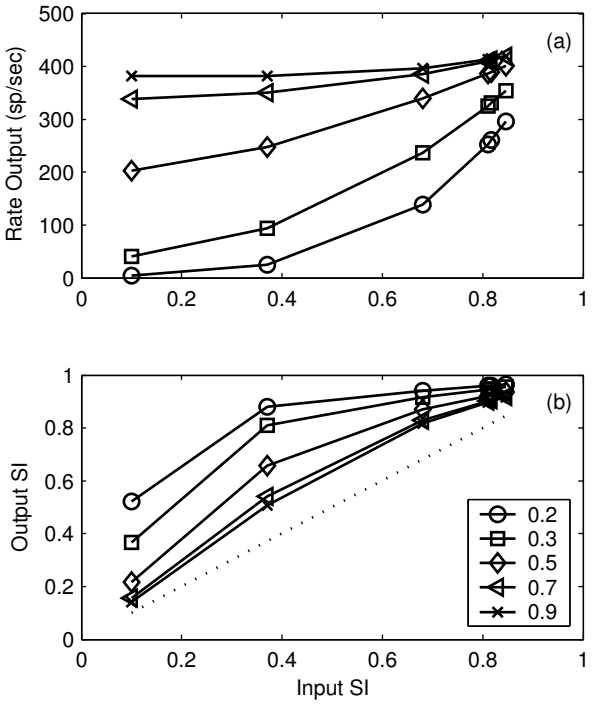
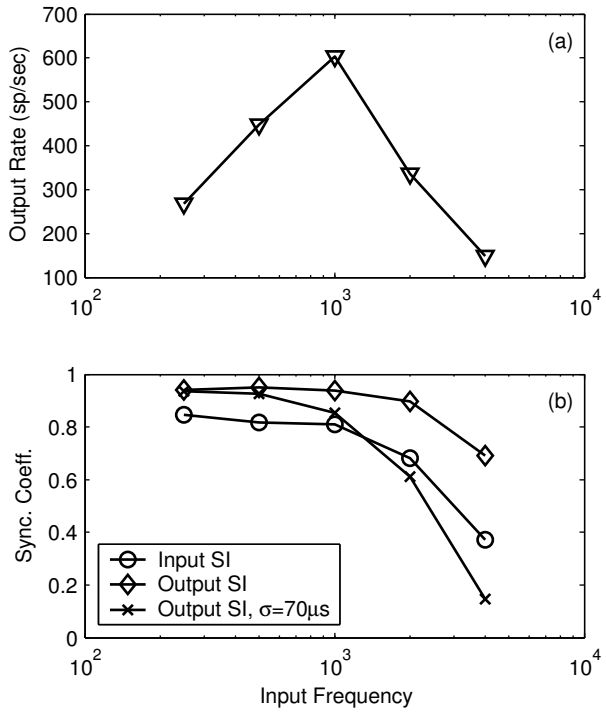


Figure 2-6: Responses of models with different EPSP amplitudes and a fixed EPSP time constant of 100μs to non-stationary inputs. The results were plotted in the same way as Fig. 2-4. The stimulus had a frequency of 500 Hz with fixed input strength of 1800. Since the input strength was defined as the EPSP amplitude multiplied the input rate, the input rate doubled when the model amplitude was decreased by half. In this way, the energy of the model inputs was kept constant for different input *SI*s. (a) Model rate responses. (b) The *SI* measure of the model responses. The EPSP amplitude had a larger effect on the degradation of the output SI when the input *SI* was high as compared with the effects of time constant illustrated in Fig. 2-4.

fixed at 8000 spikes/sec, and each input spike generated a model EPSP with an amplitude of 0.3 and a time constant of 100μs. The *SI* measure for each input frequency was plotted in the bottom panel (circles) of the figure and fitted to AN fiber data (Johnson 1980; Rothman and Manis, 2003a). Model response rate (top panel) changed non-monotonically as the input frequency increased. At low frequencies, the deadtime (absolute refractoriness) of the model prevented multiple discharges in each cycle, and the rate increased as stimulus frequency increased, since the model cell fired in each cycle. The model response rate then dropped with increasing frequency, because the stimulus inputs were less synchronized. The *SI* measure of the model output (diamonds in the bottom panel of Fig. 2-7) was higher than the input *SI* across all the frequencies. If we assume that a time jitter of 70 μs (Kopp-Scheinpflug et al., 2002, p. 11006) was added to each output spike (we assumed the time jitter[1] had a normal distribution with standard deviation 70 μs; the calculation was based on the convolution of the output PST and this normal distribution), the *SI* measure of the model output was more comparable to the physiological observation in some CN cells (e.g. see Joris et al., 1994a).

2.4.3 Effects of mixed-amplitude inputs on model responses

As illustrated in the above results, the model cell response to stationary inputs required a short time constant to maintain appropriate irregularity, and required strong inputs for a linear input-output rate



[1] The time jitter applied here may be caused by the dynamics of spike generation, refractoriness or just degraded timing due to the strong inputs (the effective coincidence window for strong inputs is larger than that for weak inputs).

Figure 2-7: Responses as a function of stimulus frequency for a model with EPSP amplitude of 0.3 and a time constant of 100μs. The input spike rate was fixed at 8000 spikes/sec, and the input *SI* varied with frequency systematically to fit the AN fiber data (Johnson 1980; Rothman and Manis, 2003a, plotted as circles in the bottom panel). (a) Model response rate changed non-monotonically as input frequency increased. (b) *SI* of the model output with) and without (diamonds) time jitter added (crosses and diamonds, respectively). The time jitter had a normal distribution with standard deviation 70 μs; the calculation was based on the convolution of the output PST and this normal distribution.

function (e.g. similar to high-CF PL responses to CF tones). However, the model cell response to non-stationary inputs required a large number of weak inputs to create enhanced synchronization, and required a large time constant to be more responsive to both synchronized and non-synchronized inputs without showing a reduction of the enhancement of the synchronization at low frequencies (when the input *SI* measure is high, see Fig. 2-4). We hypothesized that cells with mixed-amplitude inputs would respond to high-frequency stimuli (i.e. stationary inputs) irregularly and also show enhanced phase-locking to low-frequency stimuli (i.e. nonstationary inputs).

The responses of models with mixed-amplitude EPSP parameters to stationary inputs across different input strengths are plotted in Fig. 2-8. The time constant of the models was fixed at 400 μs, and the other parameters for each model (plotted in different symbols) are described in the figure legend. For the model with mixed-amplitude inputs, the amplitudes of the weak and strong inputs were fixed at 0.17 and 0.7, respectively, and the rate for both strong and weak inputs changed with the input strength. The responses of models for both mixed-amplitude inputs and single-amplitude inputs had high values of regularity (bottom panel of the plot), consistent with the finding the *CV* depends primarily on the time constant of the model (see Fig. 2-2). For the models with similar response output rates, the model with mixed-amplitude inputs tended to respond more linearly to the change of the input strength (top panel). Increasing the EPSP amplitude for the model with the same inputs made the model respond more linearly, but this manipulation

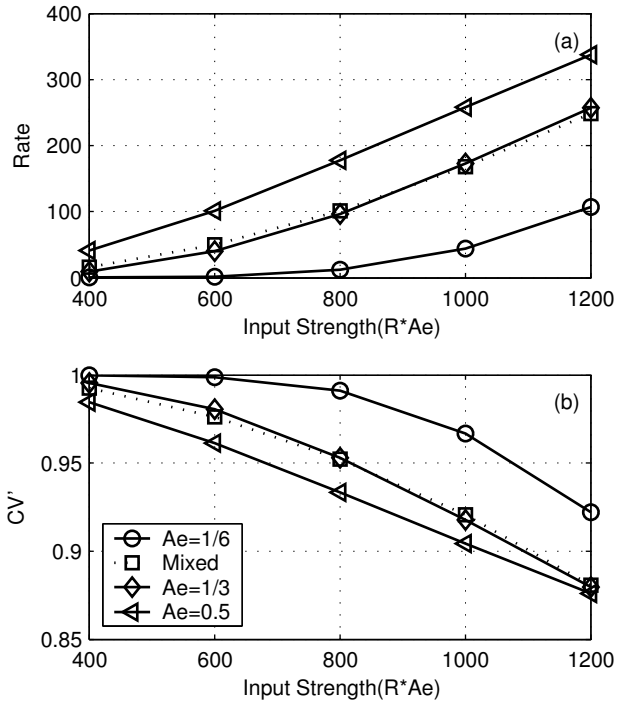


Figure 2-8: The responses for models with different synaptic configurations to stationary inputs. All model EPSPs had a fixed time constant of 400μs. The results for models with same-amplitude inputs were plotted with a solid line with different symbols (see legend for EPSP amplitude). The model with mixed-amplitude inputs (dotted line with squares) had EPSPs with amplitudes of 1/6 and 0.7, and the ratio of weak input rate to strong input rate was fixed at 28.8:1 (the results were plotted against the weak input strength for this model). (a) Model response rate. Compared to the same-amplitude input model with similar responses, the model with mixed-amplitude inputs had a higher response rate with small inputs and a lower response rate with large inputs. (b) *CV'* measure of the model responses. All model responses had *CV'* measures that would be classified as irregular cells, presumably because of the short EPSP time constant used.

reduced the enhancement of phase-locking in response to low-frequency stimuli.

Figure 2-9 shows the responses of models with different parameters to synchronized inputs. The models with same-amplitude inputs had a constant input strength of 1200. For models with mixed-amplitude inputs, the input strength was 1200 for weak inputs with an amplitude of 0.17, and the strength was 1600 for weak inputs with an amplitude of 0.08. The rate of strong inputs for the mixed-amplitude model was fixed at 250 spikes/second; the strong input amplitude was 0.7. Other parameters for the model and stimulus are described in the legend. The rate responses (top panel) for all models changed non-monotonically as a function of stimulus frequency. For the mixed-amplitude model, the response rate (plotted with asterisks and downward triangles) to the low-frequency inputs depended on the synchronized weak inputs; as stimulus frequency increased, the inputs were less synchronized and the model response depended more on the strong inputs. The *SI* measure of the model responses with mixed-amplitude inputs dropped more quickly as the stimulus frequency was increased than that of the models with same-amplitude inputs. At low frequencies, the *SI* measure of the model response stayed high since the output was dominated by the discharges generated by the weak inputs; at high frequencies, the model response was determined by the strong input, and thus the synchronization of the input was not enhanced. In general, the cell with mixed-amplitude inputs had a more linear input-output rate function in response to high-frequency tones than model cells with same-amplitude inputs, and the *SI*

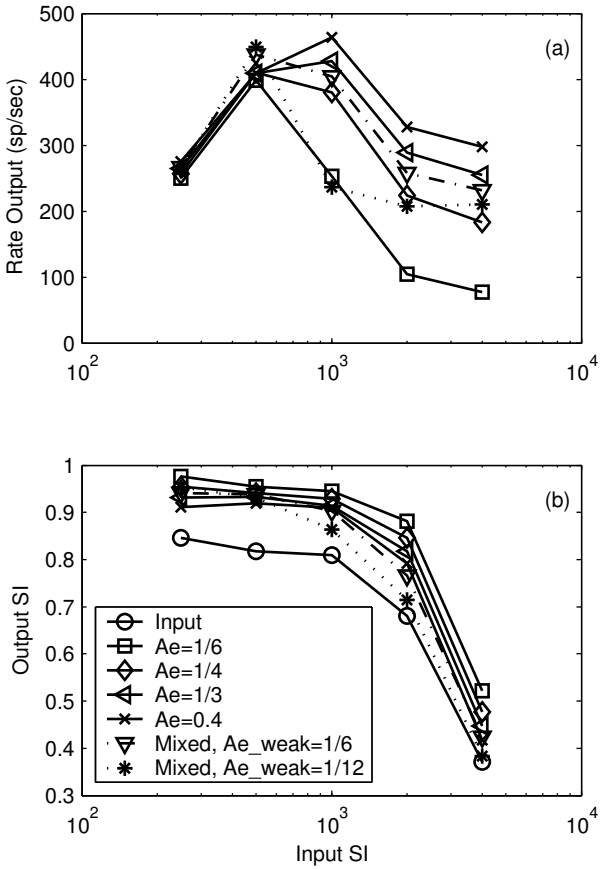


Figure 2-9: Responses for models with different synaptic configurations to non-stationary inputs. The results are plotted in the same way as in Fig. 2-7. The solid lines (see legend for model EPSP amplitude) represent the results for same-amplitude input models. The results for mixed-amplitude input models are plotted with a dotted line (stars) and a dot-dashed line (downward triangles). All model EPSP time constants were fixed at 400 μs. The input strength was fixed at 1200 for the same-amplitude input models and was 1200 and 1600 for the mixed-amplitude input models with weak amplitudes of 1/6 and 1/12 respectively. The strong inputs for both mixed-amplitude input models had a rate of 250 spikes/sec and each input EPSP had an amplitude of 0.7. (a) Model rate responses. (b) *SI* measure of the model responses. The *SI* measure of the stimulus input (fitted to AN fiber data) was plotted with circles. The *SI* measure for mixed-amplitude input models stays high in response to low frequencies and drops more quickly than that of the same-amplitude input model as the input frequency increased.

degraded more rapidly as stimulus frequency increased[2]. Both properties are desirable for the physiological responses of the globular bushy cells in the CN.

2.5 Discussion

Calculation of the PSTs and ISIs of I&F models with non-stationary inputs

Statistical analysis of neural activity, together with stochastic neuron models, has proven to be very useful for estimating physiological and anatomical parameters of neurons and elucidating the different functions of various neurons (see Tuckwell, 1988). In addition to the discrete Markov process discussed here, other stochastic neuron models have been proposed, including the Ornstein-Uhlenbeck Process (OUP) approximating diffusions, and partial differential equations modeling the spatial extent of neurons (especially for dendrites) (Tuckwell 1989, Chapter 5). However, little progress has been made to provide a satisfactory analytical solution for the first passage time problem for these models, and researchers have generally either analyzed their models with limited ranges of parameters (Kempter *et al.*, 1998) or resorted to Monte Carlo simulations. The numerical method proposed in this study provides a new way to calculate the statistics of the neuron model with more accuracy and efficiency than using Monte Carlo simulations, without the compromise of using only stationary inputs or limiting the model's parameter space. Because equations 2.11-2.14 only depend on the assumptions that the neuron can be modeled as a renewal process and that only recent inputs (discharges) determine the response, the method can be generalized in different ways as long as the conditional first passage time can be calculated numerically: 1) Relative refractoriness can be incorporated by changing the firing threshold as a function of time (assuming the previous discharge time occurs at time zero); 2) Inhibitory effects may be incorporated based on Eq. 2.16, in which the amplitude of inhibitory EPSPs is negative and arriving inhibitory discharges decrease the model potential; 3) The assumption that the membrane decays exponentially simplifies our analysis by allowing the EPSPs of incoming spikes to be combined, without having to keep track of the history of input spike times. It is possible to use more realistic EPSP waveforms (e.g. alpha-functions) by separating the model potential (hyper-polarization) contributed by recently incoming EPSPs and by EPSPs in the exponential decay tail; 4) Noise that is intrinsic to the neuron can be introduced as a diffusion of the potential distribution at each step of calculation.

[2] In the previous section, we assumed a constant time jitter to degrade the *SI* at high frequency. If the *SI* in response to high frequencies is already degraded, a smaller time jitter is required, which thus maintains the timing information at low frequencies.

Regularity of the model cell response to stationary inputs; Effects of time constant, synapse amplitude, and refractoriness

Regularity analysis of the model responses suggested that a small value of the EPSP time constant was important to prevent the cell from regular firing; this prediction agrees with findings in physiological studies (Blackburn and Sachs, 1989; Young et al., 1988). The EPSP inputs to a bushy cell have a very short time constant since the somatic synapse bypasses any dentritic filtering and the low-threshold potassium channels reduce the effective membrane time constant (Rothman and Manis, 2003a). All three response types associated with bushy cells (PL, PLn, and On-L) demonstrate irregular discharge patterns (Rothman *et al.*, 1993), regardless of possible differences in their input synapse strengths. In contrast, the chopper response type, which usually has a regular response pattern, is believed to be related to stellate cells in VCN which have large dentritic trees contacted by AN fibers (Young *et al*, 1988) and long duration EPSPs (Oertel, 1983). Regularity is also affected by the relative refractoriness of the cell responses (which is not corrected in the calculation of *CV'*), especially when the mean inter-spike interval is comparable to the duration of refractoriness. Our simulations for model AN responses showed that the *CV'* of a Poisson process modified by relative refractoriness decreased dramatically as model response rate increased (not shown). This result is consistent with the simulations reported by Rothman *et al.* (Fig. 9A in Rothman *et al.*, 1993) using a channel-based (Hodgkin-Huxley-like) model. The model responses had the lowest regularity measure when the input EPSP was just above the absolute threshold, where the refractoriness effect was strongest. Response regularity was higher for models with just sub-threshold inputs, since the combination of two required inputs was much higher than the absolute threshold. Rothman *et al.* (1993) argued that a secure input, which generates an EPSP much higher than the threshold, is necessary to maintain response irregularity for PLn cells, because the strong input decreases the relative refractory period. They also argued that the regular response of the model onset cell, which may not be physiological realistic, could be improved via inclusion of inhibition. It is possible though, as illustrated in their later study (Rothman and Young, 1996), that an inhibitory mechanism can also be used to increase the irregularity of the PLn model cell responses without the requirement of a strong suprathreshold input. Other mechanisms are also possible to reduce the effect of refractoriness (Rothman and Manis, 2003a) and thus increase the irregularity of the model cell responses.

Effect of EPSP amplitude on the input-output rate function

The input-output rate function of the model response was strongly affected by the amplitude of the model EPSPs. The input rate and amplitude had different effects on the statistics of the model potential distribution. While increasing both the input rate and EPSP amplitude increased the expected value (mean) of the membrane potential, the variance of the potential was

proportional to the square of the EPSP amplitude but had a linear relationship with the input rate. With the same input strength (EPSP amplitude multiplied by the input rate), the potential of the model with larger EPSP amplitudes had larger variance, and the model cell response depended more on the fluctuations of the potential. The input-output rate function of model cell tended to be exponential when the relative potential variance was small and to be linear when there were large potential fluctuations (Tuckwell and Richter, 1978). This prediction has important implications for the synapse conditions of bushy cells. The input discharge rate to bushy cells changes dramatically during tone bursts, as a result of onset adaptation in high-spontaneous-rate AN fibers. The fact that bushy cells with PL or PLn response type have response rates during tone bursts that are similar to those of the input AN fibers suggests that they receive at least one large input. Further, the Onset response type bushy cells that have a nonlinear input-output rate function may receive many small inputs. These predictions agree with the morphological correlates of the different cell types in the CN (see review by Cant, 1992). The PL responses are usually observed in SBCs, which have one or two large synapses known as endbulbs of Held. The PLn and Onset response type units are more closely related to the GBC, which receives small (compared to the endbulb of Held) endbulbs that are varied in number and size.

Enhanced phase locking and its relation to EPSP amplitude and time constant

Increasing the EPSP amplitude, which increased the potential fluctuation, degraded phase locking of the model response. With a large number of small inputs, the membrane potential usually followed the expected value of potential with a small variance, and the model potential could be treated as deterministic. The model cell fired very precisely around the time when the expected value of the potential crossed the threshold. This conclusion may also apply to the channel-based model, in which all the EPSPs are linearly summed. Of course, the small variance in the potential may be disturbed by other nonlinear properties, such as refractoriness. Realistically, for large inputs that generate EPSPs just above threshold, the timing of the action potentials was affected by the amplitude of the EPSPs (Rothman et al., 1993), and this relationship degraded the phase-locking of the model cell to the synchronized inputs (especially to the mid-frequency inputs, see Rothman *et al.*, 1993). Small inputs, in fact, helped to increase the precise timing of action potentials, since action potentials generated early in the periodic cycle, when there was a low rate of small inputs, had a large delay, and action potentials generated later had a small delay. This was illustrated in our study of mixed-amplitude inputs when the strong input amplitude was near but still below threshold.

Enhanced phase-locking was not affected much by the EPSP time constant as long as the EPSP time constant was short (e.g., by about a factor of 4) compared to the cycle of the stimulus frequency. Of course, the time constant is still much smaller than observed in other neurons that are not specialized for temporal coding. The short time constant of the membrane

conductance has other effects on the precise timing of the neuron's response, such as refractoriness.

Implications of mixed-amplitude inputs for the Bushy Cell model

Results from this study show that model cells that receive mixed-amplitude inputs demonstrate desired response properties that have been observed in some cells in CN. The neurons encode or enhance the temporal information at low frequencies and also carry rate information at high frequencies. These properties made the model neurons more efficient in processing information in different conditions. The inputs to high-CF cells in the CN in response to complex sounds usually have temporal (envelope) fluctuations due to narrowband peripheral filtering. A cell that receives mixed-amplitude inputs can benefit from both spectral and temporal cues. The different number and size of the endbulbs may contribute to the different synapse configuration for bushy cells, and the dendrites that branch profusely within several hundred microns of the cell body (Rhode and Greenberg, 1991) could also provide weak inputs to help enhance timing information in response to complex sounds.

Potential effects of Inhibition on model responses

Inhibitory inputs to bushy cells have been shown to exist in physiological studies (Caspary[3] et al., 1994; Wu and Oertel, 1986). The function of inhibitory inputs on model response statistics can be interpreted in several ways. First, inhibitory inputs will have different effects on the mean and variance of the model potential if the inhibitory post-synapse potential (IPSP) is integrated linearly in the I&F model. The mean of the potential will decrease as the inhibitory input rate increases, while the variance of the membrane potential will be equal to the sum of the potential variances contributed by excitatory and inhibitory inputs. The model cell responses will depend primarily on the variance of the potential distribution (that is, the fluctuation of the voltage) when the inhibitory and excitory inputs are balanced. This situation is similar to what occurs when the amplitude of individual EPSPs increases; thus, including inhibition may make the cell's response rate vary more linearly with the input rate. Second, inhibition will have different effects on the peaks and valleys of a non-stationary input. IPSPs usually have a larger time constant; therefore, the integral of the IPSPs in response to non-stationary inputs will not fluctuate as much as the integral of the EPSPs. As a result, the model cell will tend to respond more at the peak of the synchronized inputs and the inhibition will contribute to enhanced phase-locking. Finally, inhibition in bushy cells will

[3] Caspary's study showed that inhibition has the same receptive field as excitation, and that the role of inhibition is generally not lateral inhibition, which is often described as a mechanism for sharpening the receptive field. This on-frequency inhibition can be interpreted as a modulation filter that extracts the envelope fluctuation in the inputs (Nelson and Carney, submitted). As discussed here, inhibition could also contribute to the enhanced timing of the cell responses in CN.

have several nonlinear effects on the membrane properties. Inhibition will effectively make the membrane time constant faster by adding membrane conductance, and it will also decrease the effective amplitude of EPSPs, thus reducing the amplitude of a secure synapse to an amplitude that is just above or even below the threshold.

In summary, while the membrane properties of a neuron define the cell's capacity to process the information carried in the input spikes, the synaptic configuration of each input determines how the information is actually processed. Cells that receive multiple inputs could have different synaptic configurations and respond differently to corresponding stimuli, and thus achieve processing functions that are advantageous for specific stimuli.

Chapter 3 Predicting responses of CN cells that receive convergent AN inputs based on coincidence-detection

3.1 Abstract

The basic response properties of model coincidence-detecting (CD) cells that receive converging auditory-nerve (AN) inputs are presented in this chapter. The responses to both pure tones and tones in wideband noise are presented. The responses of model cells with different synaptic configuration were explored and compared, and possible relations to physiological responses of cells in the anteroventral cochlear nucleus are discussed. Coincidence-detection cells with mixed-amplitude inputs showed response properties similar to those of CN cells (as well as those of AN fibers at low stimulus levels), and distinct response patterns in response to complex sounds that have been observed in physiological studies. The coincidence-detecting model cells responded to the spatio-temporal response pattern of their inputs and demonstrated a physiologically realistic mechanism for processing information in complex sounds that is carried in the spatio-temporal responses of the population of AN fibers.

3.2 Introduction

Neurons communicate with each other through discrete discharge events. The idea that information is represented by coherent patterns between different neurons has been a fundamental concept of modern neuroscience. To distinguish a given subset of cells participating in a particular representation based solely on discharge rate is expensive in terms of neuron numbers and sacrifices flexibility (König et al., 1996). It has been suggested, therefore, that synchronization of activity among a distributed population of neurons is very important in processing information in neural networks. One essential advantage of such population coding is that individual cells can participate at different times in the representation of different patterns; that is, the information can be represented by the spatio-temporal pattern of neuronal activities (Singer and Gray, 1995).

In the mammalian auditory system, the inner ear (cochlea) performs a mechanical frequency analysis; each place on the basilar membrane along the cochlear partition has a different resonant frequency and results in a sharply tuned and very sensitive input to the sensory hair cells. The afferent AN fibers that innervate sensory hair cells carry the information contained in different spectral channels to the brainstem through their temporal discharge patterns. The discharges are phase-locked to the stimulus waveform at low frequencies and reflect the envelope fluctuations of the stimulus when the stimulus

frequency is high. Sound information is thus encoded in the spatio-temporal pattern of discharges across the population of AN fibers, where the spatial dimension is frequency and refers to the orderly tonotopic map of tuned neurons in most auditory nuclei. The spatio-temporal patterns of AN fibers in response to various stimulus waveforms are determined primarily by the phase properties of auditory filters tuned to different frequencies. For example, the phase difference of outputs of two model auditory filters tuned at 850 and 950 Hz changed systematically between 0^0 and 180^0 as input frequency changes (Fig. 3-1). Advantages of coding schemes based on the temporal responses across a population AN fibers are that they are generally robust to changes in overall stimulus level and to non-linearities, such as saturation (Carney and Yin, 1988), and they are similar for fibers with different spontaneous rates (Johnson, 1980).

Physiological studies have shown that processing of spatio-temporal patterns may occur as early as in the cochlear nucleus (CN). Nearly all of the cells in the anteroventral cochlear nucleus (AVCN) receive convergent inputs from AN fibers. 1) Studies of phase-locking in the axons of the trapezoid body (TB), which originate from bushy cells in the AVCN, show enhanced synchronization as compared with AN fibers (Joris et al., 1994a). Synchronization of phase-locked responses is usually quantified by measuring vector strength as a function of input level and then finding the maximum value (R_{max}, Goldberg and Brown, 1969); typical R_{max} values for low-CF AN fibers are between 0.8~0.9. The study by Joris et al. (1994a) reported that 75% of fibers with CF less than 700Hz had $R_{max} > 0.9$. 2) Carney (1990) showed that some low-frequency cells in the AVCN are sensitive to the temporal discharge pattern across the AN population. In this study, complex stimuli were used that had fixed flat magnitude spectra, and the phase spectra of the stimuli were manipulated (a phase shift with variable slope at a particular frequency was introduced). Some cells in the AVCN showed sensitivity to changes

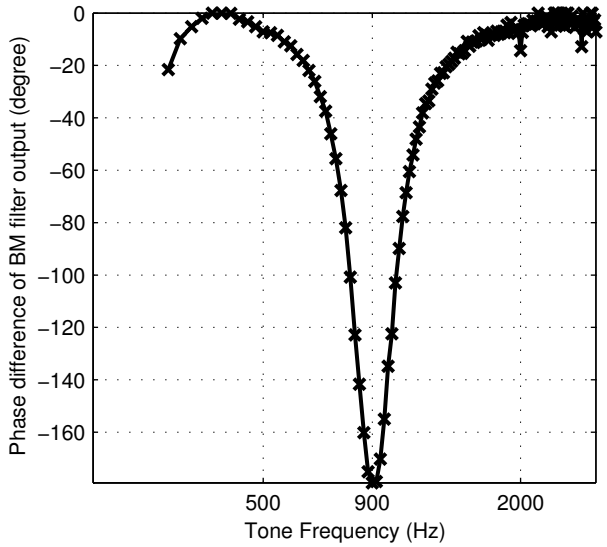


Figure 3-1: The phase difference between the outputs of two model auditory filters as a function of stimulus frequency. The CFs of two auditory filters were

tuned at 850 Hz and 950 Hz. The filters had out-of-phase responses with each other to 900 Hz tone (phase difference was around 180 degree). The responses of two filters became in-phase with each other as the signal moved away from 900 Hz.

in the slope of the phase transition (Figs. 6-11 in Carney 1990). These phase-sensitive responses come from several different cell types in the AVCN, and the responses are consistent with the hypothesis that these cells receive multiple inputs from AN fibers and respond based on a cross-frequency coincidence-detection mechanism.

Several models have been developed to process the spatio-temporal cues in AN population responses in the context of pitch perception (see review by Lyon and Shamma, 1996). The purely temporal information from the phase-locked responses can be extracted using models based on the auto-correlation mechanism. Cross-correlation models (Deng and Geisler, 1987; Carney, 1994; Heinz et al., 2001; Carney et al., 2002) are proposed to take advantage of temporal (phase-locked) information in one fiber and systematic phase relationships along the spatial organization of the fiber array. The use of cross-channel comparisons in cross-correlation models combines the spatial and temporal information together in one mechanism and thus provides an intuitive way to understand the processing mechanism. Shamma (1985) proposed a model based on a lateral inhibition mechanism that depends on the spatio-temporal patterns across the AN responses and shares some similarities with cross-correlation models; the model was originally developed for speech processing but can also be used for processing of other complex sounds.

Coincidence detection is a realistic physiological mechanism that performs a form of correlation computation and is very sensitive to temporal properties of the input. There is no sharp boundary between coincidence detection and temporal integration since the criterion for such a distinction depends on the relationship between the integration time and the mean interspike interval. Physiological experiments have shown that neurons in the medial superior olive (MSO) are sensitive to differences in the timing of their afferent signals on the order of tens of μs (Goldberg and Brown, 1969; Yin and Chan, 1990), providing strong support for a coincidence-detection mechanism. The fact that the nonlinear membrane properties of Type II AVCN neurons (bushy cells) are similar to those of cells in the MSO suggests that the cells in AVCN may also behave as coincidence detectors (Manis and Marx, 1991, Rothman et al., 2003ab). Using coincidence detection as an underlying physiological mechanism, the spatio-temporal cues embedded within AN population responses can be processed realistically and efficiently. To fully understand the implications of the coincidence-detection mechanism for neural response properties and for the functionality of information processing, more realistic models must be constructed and their response properties to complex stimuli should be studied.

In this chapter, a physiologically realistic "shot-noise", coincidence-detecting model cell that received inputs from convergent AN fibers is described. The basic response properties of two different model cells were explored in detail. One model cell was extremely sensitive to the spatio-temporal patterns of the inputs and represented the best coincidence-detecting cell for processing spatio-temporal information in a masked-detection task; another cell received mixed amplitude inputs and represented a more realistic synaptic

configuration for the cells in CN. The model responses to pure tones and tones-in-noise were studied in detail. How the model parameters affected the model responses and relations between the "shot-noise" model and other simple coincidence-detecting models were then explored.

3.3 Method

3.3.1 Computational Auditory-nerve Model

Carney (1993) developed a signal-processing model for auditory-nerve fibers that focused on the timing properties of AN responses. Zhang et al. (2001[4]) extended the model to more accurately simulate two-tone suppression and population (filter-band) responses found in physiological data. This model has proven useful in understanding the nature of the temporal cues provided to the CNS by the population of AN fibers (Heinz et al., 2000). The model was adapted to human by Heinz et al. (2001ab) to study human performance in some psychophysical experiments; this version of the model (model #1 in Heinz et al., 2001d) was used to simulate the responses of human AN fibers in the present study. The bandwidths of tuning of model human AN fibers (which are narrower than that of cat) were fitted to human ERB data. The population responses of low-threshold, high-spontaneous-rate (HSR) AN fibers (spontaneous rate was set to 60 sp/sec[5]) were simulated to produce either instantaneous firing rates $r(t)$ or discharge trains ($\mathcal{T}$) from the model, depending on the coincidence-detection model used.

3.3.2 Monaural Cross-frequency Coincidence Detection Model

Colburn described a simple coincidence-detection model in his studies of binaural detection phenomena (1973, 1977). The model receives two AN-fiber inputs from different ears with the same CF, and is assumed to discharge only when the two input fibers both discharge within a narrow temporal window. Heinz et al. (2001c) proposed a monaural, across-frequency coincidence model that has the same structure but receives inputs from AN-fibers from the same ear, that can have different CFs. The coincidence counts of the model can be represented by

$$C_{ij}\{\mathcal{T}^i,\mathcal{T}^i\} = \sum_{l=1}^{Ki}\sum_{m=1}^{Kj} f(t_l^{\,i} - t_m^{\,j}), \qquad (3.1)$$

[4] A recent study has been conducted to modify the synapse dynamics of this composite model. The new synapse model has a more realistic offset adaptation response, and the response to amplitude-modulated (AM) stimuli of the new composite model is more comparable with the physiological data. This study is reported in the Appendix of this dissertation.

[5] The spontaneous rate of model AN fibers is defined as the instantaneous firing rate before the discharge generator. For the simulated discharge trains, the spontaneous rate is lower because of refractoriness.

where $\mathcal{T}^i=\{t^i_1,\ldots,t^i_{Ki}\}$, and $\mathcal{T}^j=\{t^j_1,\ldots,t^j_{Kj}\}$ are discharge times of two input AN fibers from the i^{th} and j^{th} channels, respectively, K_i and K_j are the counts for the two channels, and f(.) is a rectangular coincidence window with unity height (see Eq. 2 in Heinz et al., 2001c).

If the AN discharge times $\mathcal{T}^i$ and $\mathcal{T}^j$ are independent responses of two stochastic model AN-fibers to some deterministic stimulus, with the output stochastic process as a non-stationary Poisson processes with time varying rate functions $r_i(t)$ and $r_j(t)$, the expected value of the coincidence counts during stimulus time duration T is given by (Colburn, 1977)

$$E[C_{ij}(\mathcal{T}^i,\mathcal{T}^j)]=\int_0^T\int_0^T f(x-y)r_i(x)r_j(y)dxdy\,. \qquad (3.2)$$

If the coincidence window is very short compared to the fluctuations of the instantaneous firing rate, the equation can be simplified to a cross-correlation operation on the two input AN spike trains.

This simplified coincidence-detection model does not address several situations that are common for cells observed in physiological studies: 1) cells usually receive more than two inputs (Liberman, 1991,1993), 2) cells receiving many input spikes from fibers tuned to the same CF would be expected to have some temporal summation between the matched-CF input spikes that would contribute to the cell output; therefore, a simple cross-CF coincidence window may not be sufficient to characterize the cell responses.

A more explicit, though still abstract and phenomenological, model to simulate the neural responses based on the coincidence-detection mechanism is the “shot-noise” neuron model (Colburn et al., 1990; see Chapter 2). In such a model (which is also used in the present study), each input discharge generates a small, exponentially decaying increment to the model membrane potential, analogous to an excitatory postsynaptic potentials (EPSP) in a real neuron. An output discharge was generated when the model potential exceeded a pre-defined threshold (which was always set to 1.0). Following each output spike, the membrane potential was reset to zero for 0.7 ms, called the "deadtime", to simulate a refractory period for the neuron (spikes that arrived during the deadtime did not generate EPSPs). The model was used to simulate different synaptic configurations that varied in several critical parameters: 1) the number of model AN inputs and the CFs of these AN fibers, and 2) the amplitude and time constant of the EPSPs for each input AN fiber. Two model cells were used in illustrating the response properties of the “shot-noise” model implementation of a coincidence-detecting cell (see Fig. 3-2). Model cell *A* received two AN inputs with CFs of 850 Hz and 950 Hz, respectively. Each input spike generated an EPSP with an amplitude of 0.7 and with a decay time constant of 300 μs. Model cell *B* received inputs with different EPSPs: strong inputs from a 900-Hz AN fiber that generated an EPSP with an amplitude of 0.8 and a time constant of 100μs; weak inputs from 10 AN fibers with CFs of 850 Hz and 950 Hz (5 fibers in each CF channel), each of which generated an

EPSP with an amplitude of 0.2 and a time constant of 300 μs. Model cell *A* represented the extreme case in which the input temporal pattern changed dramatically as stimulus frequency changed, whereas model cell *B* represented a more realistic synaptic configuration of cells in the *CN* (cochlea nuclei) receiving convergent AN inputs (see Chapter 2).

3.3.3 Simulations of Model Responses

For the simulation of model responses to pure tones, the stimulus was gated with a 5-ms cosine-squared function and the output discharges with latencies between 30 and 230 ms were counted for 200 repetitions of the stimulus. Model responses to tone input with different frequency and sound level were studied.

In the study of model cell responses to tones in wideband noise, the response area and rate-level function of the model cell in the presence of

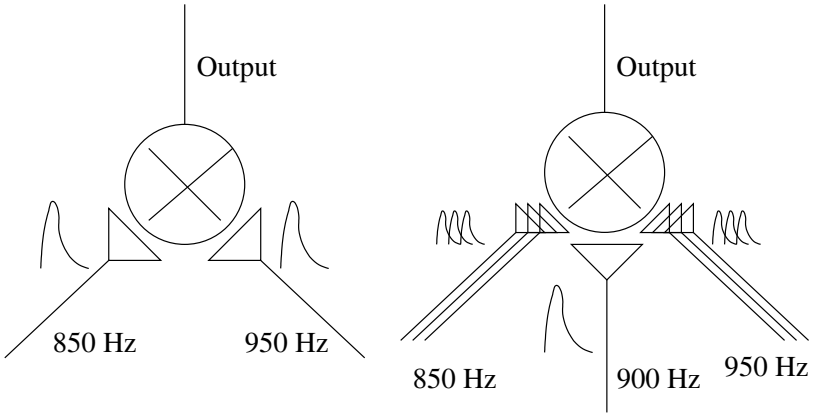


Figure 3-2: The synapse configuration of two coincidence detectors used in present study. A) Coincidence detector *A* received two AN inputs with CFs at 850 Hz and 950 Hz. Each input discharge generated an EPSPs with amplitude of 0.7 and time constant of 300 μs; B) Coincidence detector *B* received one strong input from a 900-Hz AN fiber and weak inputs from 10 AN fibers with CFs of 850 Hz and 950 Hz (5 independent AN fibers at each CF). The strong input generated EPSPs with amplitude of 0.8 and time constant of 100 μs. The weak inputs generated EPSPs with amplitude of 0.2 and time constant of 300 μs.

wideband noise were computed. The wideband noise was geometrically centered around 900 Hz and had a bandwidth of 3 kHz. The tone and noise were turned on simultaneously with a total duration of 250 ms, gated with a 20-ms cosine squared function. The response areas of model cells were computed at tone levels of 20, 40 and 60 dB SPL, with the noise spectral level fixed at 30 dB SPL. The rate-level functions of model cells were computed for noise spectral levels of 10, 20 and 30 dB SPL, with tone frequency fixed at 900 Hz. The model cell responses with latencies between 30 and 230 ms were counted. Fifty independent noises were generated for each simulation, and the response to each noise (plus tone) was calculated for 20 repetitions (so there were a total of 1000 repetitions for each stimulus condition).

The onset (first 30 ms) of model cell responses were not used in the present study. For input model AN fibers, the responses at onset were dominated by adaptation and had wider dynamic ranges than the steady state responses. The onset responses of model coincidence-detecting cells were affected primarily by the input AN fiber adaptation and were more level-dependent.

3.3.4 Comparison between Shot-noise Model and Cross-correlation

Constructing models for auditory theory usually involves two driving forces that sometimes contradict each other. On the one hand, the models should be complex and realistic enough to approximate neural implementations and predict physiological responses; in most case these models generate discharges that mimic the inputs and outputs of the neural system. On the other hand, models used to predict psychophysical experiments are usually based on signal processing schemes to simplify the operation. Using simple signal-processing models is simple and efficient, and can generate satisfactory approximations if the essential properties of the neural dynamics are captured.

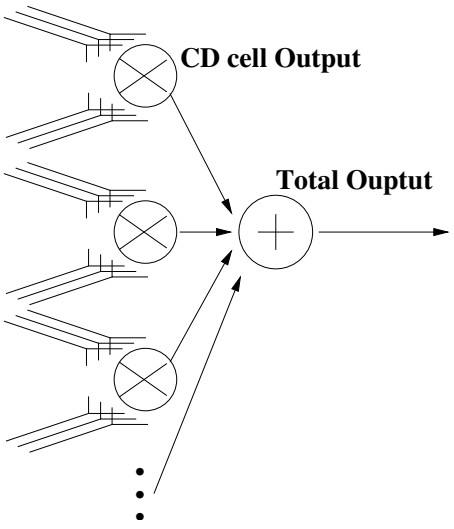
For the coincidence-detection model, the simple model (cross-correlation) in Eq. 3.2 has been used successfully to account for many binaural experiments (Colburn, 1995). The physiologically realistic coincidence-detection model ("shot noise" model) underlying this simple mechanism had a very low firing rate (a result of the very short coincidence window) that is not observed in physiological studies. While realistic response rates could be achieved by increasing the number of inputs, there was a question as to whether the cells were still sensitive to the input spatio-temporal patterns when they had a large number of inputs. To study whether the cross-correlation and shot-noise models were comparable with each other, the performance of models with different parameters were evaluated by simulating the tone-in-noise detection experiment from Kidd et al. (1989). Specifically, the signal to be detected was a 900-Hz pure tone, and the noise had a bandwidth of 3 kHz geometrically centered at the signal frequency and with a fixed spectral level of 35 dB SPL. The tone and noise were turned on simultaneously with a total duration of 250 ms, gated with a 20-ms cosine squared function.

To evaluate the effects of model parameters on the model's performance for detecting a tone in noise, a detector based on the model responses was constructed and measurement of the detector performance was defined. The detector used in the present study was based on the responses from a population of model CD cells, illustrated in Fig. 3-3. Each model CD cell received inputs from two channels (at 850 and 950 Hz, respectively), and each channel had N independent input AN fibers (all with EPSPs that had the same amplitude and time constant). The parameter N varied from 1 to 32. Responses from several identical but independent CD cells were summed to generate the total output discharge count. The number of CD cells was equal to $32/N$ to guarantee that the total number of AN inputs to the detector at each frequency was 32. The model CD cell responses were calculated during the steady-state response (100~200ms). The total output count C in response to noise n or tone-in-noise (n, L) was used as a decision variable. The strategy was to look for a drop in the responses to indicate when the signal was present[6].

The performance of the detector was evaluated by calculating the sensitivity measure (Heinz et al., 2001):

$$Q = \frac{E[C \mid n, L] - E[C \mid n]}{\sqrt{(Var[C \mid n] + Var[C \mid n, L])/2}}, \qquad (3.3)$$

where $C|n$ and $C|n,L$ represent the total response counts to noise only and noise plus signal , respectively. The signal level L (20 dB above the noise spectral level) was approximately equal to the subject threshold for tone-in-noise detection from Kidd *et al.* (1989). An alternative approach



[6] As the results indicated in following section, the model cell response decreased when a 900 Hz tone was added to the wideband noise because the input AN fiber responses became more out-of-phase with each other with the addition of the 900 Hz tone.

Figure 3-3: The model detector based on "shot-noise" coincidence detection model cells. Each CD model cell received inputs from two channels with CFs at 850 Hz and 950 Hz. Each channel of CD model cell received N independent AN inputs. Response spikes from several identical but independent CD cells were summed to produce the detector decision variable. The total number of AN inputs at each CF was fixed at 32, so the number of CD cells used was determined by the number of independent AN fibers in each channel to a single CD cell ($32/N$).

to evaluate the detector performance was used in which the model psychometric function was constructed by computing performance at different signal levels; this computation was based on 100 independent noise maskers.

3.4 Results

3.4.1 Responses of model Coincidence Cells to Pure Tones

A convenient summary of the model responses to pure tones was to plot the responses (e.g. rate, synchronization coefficient, phase etc.) as a function of the frequency and the sound level of the pure tone input. Other plots could then be derived from this 3-dimensional view of the data. For example, given the model responses *r(f,spl)* (rate as a function of frequency and sound level of a pure tone), the response area (*spl* is fixed), rate-level function (*f* is fixed) and tuning curve (contour curve at iso-rate threshold) could be derived.

Figure 3-4 summarizes the rate responses of model cell *A* to pure tones. The 3-dimensional view of the model cell's responses is plotted in the top panel, and the response area and threshold tuning curve of the model cell are plotted in the middle and bottom panels, respectively (Fig. 3-4). The model cell response increased as the tone level increased, with the response shape affected by both the tuning of the AN fibers (dashed and dotted lines in the bottom panel, Fig. 3-4) and their input phase relationship. When the stimulus frequency was around 900 Hz, the two input AN fibers both responded with high rates to the stimulus frequency, but the model cell had a low response rate because the inputs were out

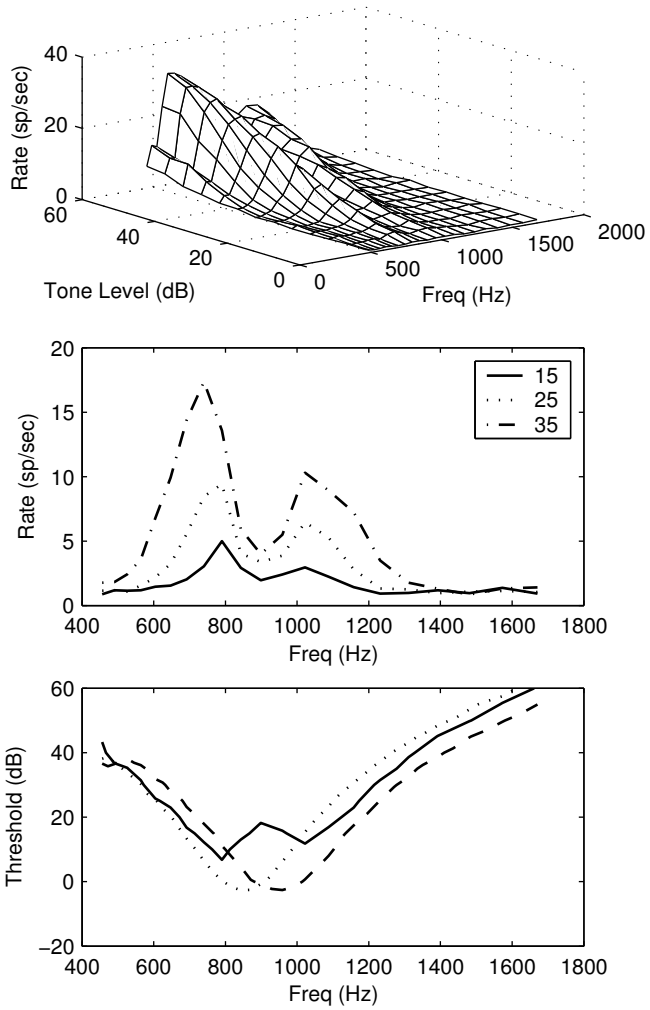


Figure 3-4: Responses of model cell *A* to pure tones. The stimulus was gated with 5-ms cosine-square function and the output spikes with latencies between 30 and 230 ms were counted for 200 stimulus repetitions. Top panel: Three-dimensional view of rate responses as a function of frequency and level of the stimulus. Middle panel: The response area of the model cell at tone levels of 15, 25, and 35 dB SPL. Bottom panel: the threshold tuning curves of the model cell (solid line) and of the input AN fibers (850 Hz (dotted) and 950 Hz (dashed)). The response threshold of the model cell was set at 2.5 sp/sec. The response threshold of the model AN fibers was set at 60 sp/sec (the spontaneous discharge rate of AN spikes including refractoriness was roughly 50 sp/sec). The responses were affected by both the AN fiber tuning and input phase relationship.

of phase. As a result, there were two peaks in the response area, neither of which was at the CFs of the input AN-fibers. Also, the threshold for the model cell at 900 Hz was somewhat higher than the threshold for other nearby frequencies apparently because of the coincidence-detection interactions between the two inputs whose phase response differ, the outline of the threshold tuning curve did not appear to be the simple union of the threshold tuning curves of the two input AN fibers.

The responses of model cell *B* to pure tones are plotted in Fig. 3-5. The responses were generated and plotted in the same way as in Fig. 3-4. The model cell responses at low levels were very similar to a 900-Hz AN fiber, as the cell responses were more dependent on the responses of the strong input AN fiber than on the temporal patterns in the other AN inputs at low levels. The response area at low levels and the threshold tuning curve of the model cell were more comparable to responses of neurons in the VCN (Rhode and Greenberg, 1991), which are similar to those of AN fibers. At high stimulus levels, the strong input AN fiber was saturated, there was relatively more inputs from other weak AN fibers, and the model cell responses began to be affected by the temporal patterns of the AN inputs. The response at 900 Hz at higher stimulus levels was lower than the responses to other frequencies, for which the inputs were more in phase with each other.

3.4.2 Responses of Model Coincidence Cells to tones in wideband noise

The response areas and rate level functions for model cells *A* and *B* to tones in wideband noise are plotted in Figs. 3-6 and 3-7 respectively. Since the input rate of AN fibers were all saturated at the noise spectrum

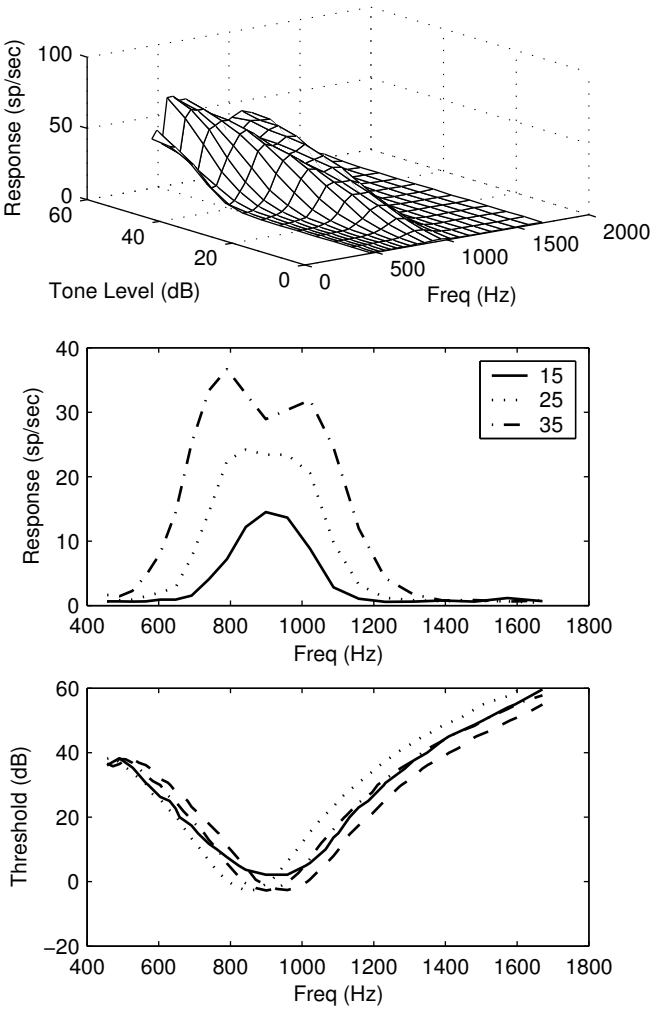


Figure 3-5: Responses of model cell *B* to pure tones. The data were computed and plotted in the same way as in Fig. 3-3. At low stimulus levels, the model cell responses were very much like the responses of a 900 Hz AN fiber. At high stimulus levels, the responses were more affected by the temporal patterns of input AN fibers.

level of 30 dB SPL, any changes of the model cell response area were caused by changes of the spatio-temporal patterns of the input AN fibers. At low tone levels (20 and 40 dB SPL), the response area was flat as a function of tone frequency because the temporal pattern of the AN fibers did not change; signal-to-noise ratio was so low that the responses were dominated by the noise. When the tone level was high (60 dB SPL), responses for both model cells changed systematically as the input tone frequency changed. At a tone frequency of 900 Hz, the AN inputs were more out of phase with each other and the response dropped. For tone frequencies away from 900 Hz, the AN inputs were more in phase with each other and the response increased. Such response properties not only made the model cell sensitive to changes in the spectral shape of the input stimuli but also made such sensitivity frequency specific.

The rate-level functions for model cells *A* and *B* were different. For model cell *A*, the responses always dropped when tone level increased, and the response was determined by the signal-to-noise ratio (the tone level referenced to the noise spectrum level). For model cell *B*, the responses at low noise spectrum levels increased with tone level, suggesting that the large 900-Hz AN input had a strong influence on the model response. At high noise levels, the model response dropped as tone level increased, suggesting that the temporal pattern of input AN fibers influenced the model response more strongly at high levels.

In general, the responses of model cell *B* were similar to those frequently observed in a recent study of cells in AVCN of gerbil,

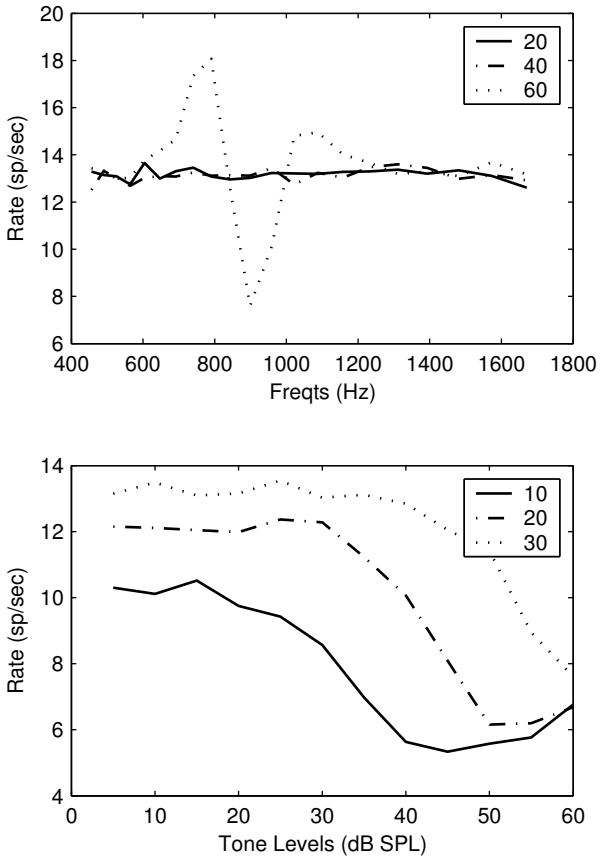


Figure 3-6: The response areas and rate-level functions for model cell *A* in response to tones in wideband noise. The wideband noise had a bandwidth of 3 kHz geometrically centered at 900 Hz. The tone and noise were gated with the 20-ms cosine squared function and turned on simultaneously with a total duration of 250 ms. The response areas of model cells were computed at tone levels of 20, 40 and 60 dB SPL, with noise spectral level fixed at 30 dB SPL. The rate-level functions of model cells were computed at noise spectral levels of 10, 20 and 30 dB, with tone frequency fixed at 900 Hz. The model cell responses with latencies between 30 and 230 ms were counted. Each simulation used 50 independent noise tokens, and the response to each noise (plus tone) was calculated for 20 repetitions (so there were a total of 1000 repetitions for each stimulus condition).

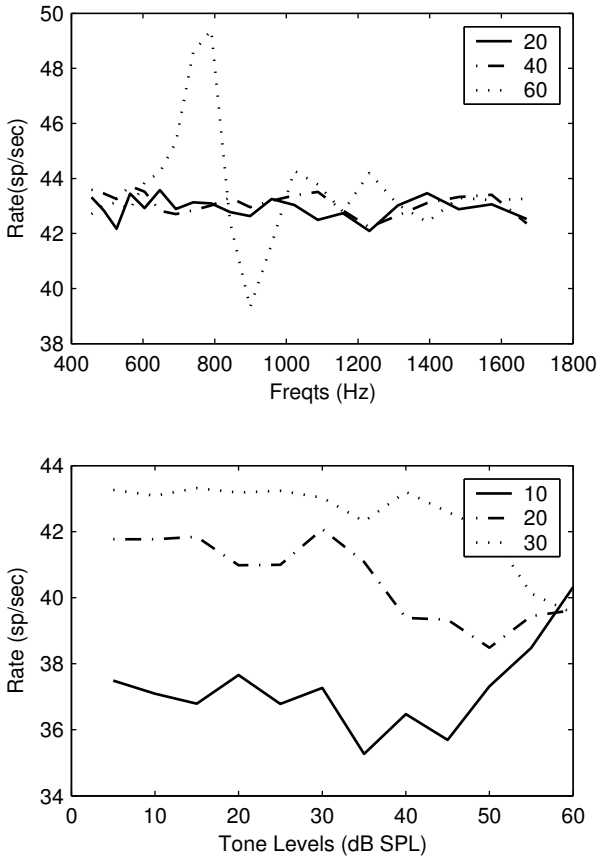


Figure 3-7: The response areas and rate level functions for model cell *B* in response to tones in wideband noise. The data were computed and plotted in the same way as in Fig. 3-6.

suggesting that such a synaptic configuration was realistic (see DISCUSSION).

3.4.3 Response sensitivity of coincidence detection models to tone-in-noise

Figure 3-8 plots the sensitivity measure of the model detector (*N*=1, i.e. each input channel of a CD cell contains one model AN fiber) for different EPSP parameters. The sensitivity is plotted as a function the EPSP time constant (τ) for various EPSP amplitudes (*Ae*) in the top panel, and sensitivity is replotted as a function of the duration of the coincidence window (defined as the longest time interval between two EPSPs that could generate an output spike, which was computed as $CW = \tau \times \log(\frac{Ae}{1-Ae})$) in the bottom panel. The sensitivity measure changed non-monotonically with the EPSP parameters. The most sensitive model detectors have a coincidence window around 170 μs. A shorter coincidence window did not guarantee better performance, because the response variability increased as response rate decreased.

For detectors with different numbers of independent input fibers (*N*=1, 4, 8), and with τ fixed at 200 μs, the EPSP amplitude that resulted in the highest sensitivity was computed. The rate-level function (top panel) and psychometric function (bottom panel) for these detectors are plotted in Fig. 3-9. The rate of each model coincidence cell was normalized for comparison; the maximum rates for different coincidence-detecting cells (with different *N*s, and different EPSP parameters) are labeled in the legend. The EPSP amplitude (*Ae*) that resulted in the greatest sensitivity was used in the computation (0.7, 0.4, and 0.3 for CD

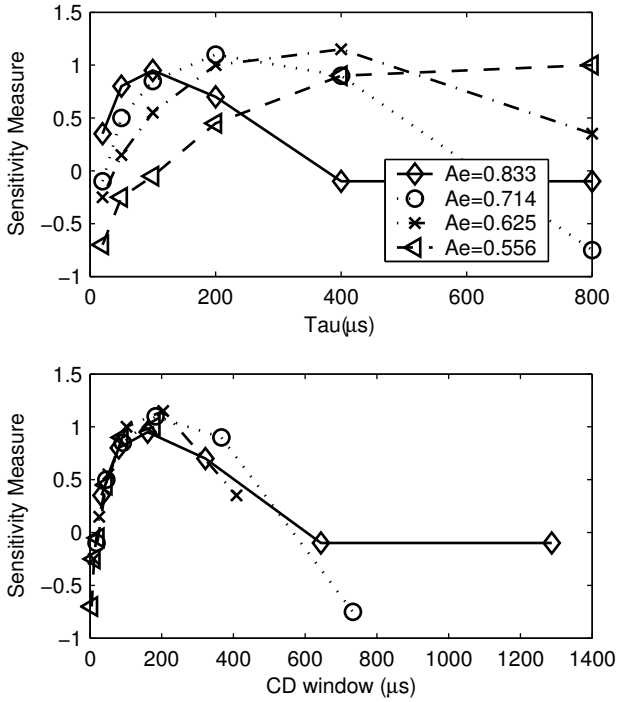


Figure 3-8: The sensitivity measure of the model detector (N=1) for various EPSP parameters. a) sensitivity measure as a function of time constant (τ) for different EPSP amplitude (Ae, see legend); b) sensitivity measure as a function of equivalent coincidence window. The coincidence window was computed as $CW = \tau \times \log(\frac{Ae}{1-Ae})$. The sensitivity measure changed non-monotonically with EPSP parameters. The shortest coincidence window did not result in the best sensitivity. The maximum sensitivity was achieved when the coincidence window was approximately 170 μs.

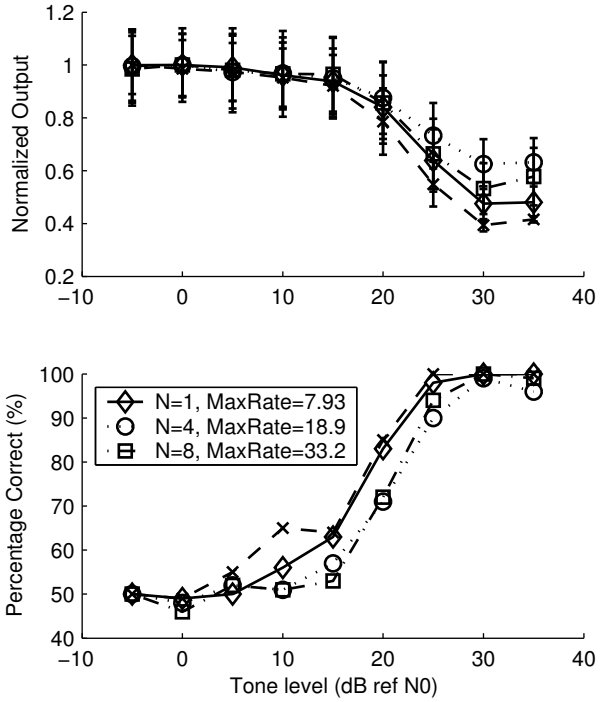


Figure 3-9: The normalized rate-level functions (top) and psychometric functions (bottom) for coincidence-detection models based on shot-noise model and cross-correlation model. The coincidence-detection model received AN inputs with CFs at 850 Hz and 950 Hz. The shot-noise model cell received 1, 4, or 8 independent AN inputs at each CF channel (plotted with diamonds, circle, and squares respectively). The EPSP time constant was fixed at 200 μs and the EPSP amplitude of each model cell changed (0.7, 0.4, and 0.3 respectively) to achieve maximum coincidence detection. The maximum response rate of model cell (see legend) increased as number of independent input AN fibers increased. The psychometric functions for three shot-noise models showed they were all very sensitive to the spatio-temporal pattern of the AN inputs. The performance of the cross-correlation model (cross symbols, not shown in legend) was very similar to performance of shot-noise model, suggesting it was a good approximation to the more physiological realistic coincidence detector for tone-in-noise detection.

models with N=1, 4, 8 respectively). The dashed lines with cross symbols represented results from the cross-correlation model (Eq. 3-2 with coincidence window of 20 μs). In general, when more inputs converged on one model cell, the maximum response rate increased and became more realistic, but the model cell was still sensitive to the spatio-temporal pattern of the AN inputs. Thus the cross-correlation model was a good approximation to the "shot-noise" coincidence-detection model with optimum coincidence-detection window in terms of tone-in-noise detection.

3.5 Discussion

3.5.1 Implications for Physiological Evidence of Cross-frequency Coincidence-detecting Cells

It is seldom the case that neurons receive only one synaptic input, and electrophysiological studies (Oertel, 1985; Manis and Marx, 1991; Rothman et al., 2003; Goldberg and Brown, 1969, Yin and Chan, 1990) have shown that many cells in the auditory brainstem are specialized for processing of timing information carried by the discharge patterns of their inputs. However, do cells in the brainstem exhibit distinctive response patterns that can be predicted by monaural cross-frequency coincidence-detection models? Based on their broad tuning it is generally agreed that the *Onset* cells in VCN receive convergent inputs from a wide range of AN CFs and that they respond based on coincidence detection; that is, they respond only at the stimulus onset, or to low frequency stimuli for which multiple input discharges arrive within a short time period (Oertel et al., 2000). Carney (1990) demonstrated that many cell types in AVCN were sensitive to changes in the phase spectra of complex stimuli (Huffman sequences), suggesting that these cell types received convergent AN inputs tuned to different CFs and were sensitive to the relative phase of these various inputs.

The results presented here show that model coincidence-detecting cells produce response properties similar to observed AN fiber responses at low levels, and produce the distinctive response patterns due to coincidence detection in response to high-level stimuli or in response to complex sounds. The response properties predicted in the present study have been observed for cells in the AVCN in recent studies of gerbil (unpublished observations). Among a total of 97 units recorded, more than 30% of units showed a decrease in response rate when a CF tone was added to the noise[7], and showed multiple peaks in the response areas at high levels. Figure 3-10 plots the responses of such a unit from that study. The response area to pure tones is plotted in the top panel; multiple peaks in the response area were observed at high stimulus

[7] 34 units out of 97 units recorded had a decrease in response rate when a tone was added to the noise out of 97; this response characteristic was observed in many cell types (no preference to a particular cell type).

levels. The response area for tones in noise (with noise spectrum level at 30 dB) is plotted in the middle panel. A decrease in responses was observed when a tone was added to the noise, and tones near CF caused the greatest decrement in cell responses. When the level of the CF tone increased, the cell responses to the tone in noise decreased (bottom panel of Fig. 3-10). These response properties are consistent with

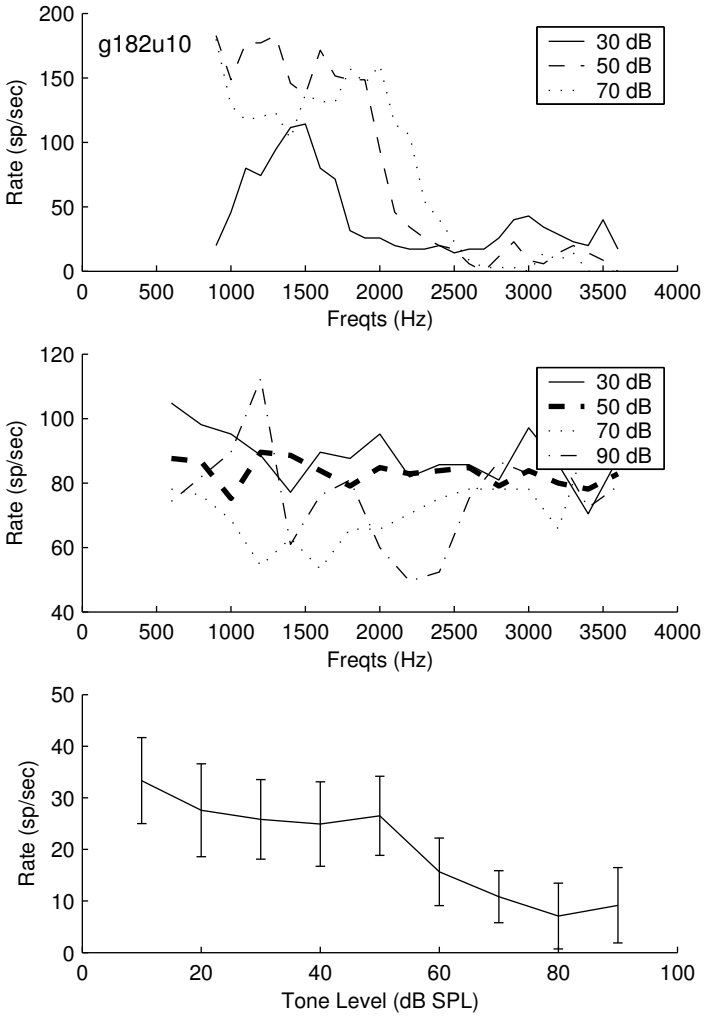


Figure 3-10: Responses of one primary-like cell (unit ID: g182u10) recorded from AVCN of the gerbil. Top Panel: Response area to pure tones at several levels. The response area had multiple peaks at high stimulus levels; Middle Panel: Response area for tones in noise, with fixed noise spectrum level at 30 dB SPL. The cell responses decreased when a tone was added, especially for tones near CF above 50 dB SPL; Bottom Panel: Rate-level function for tone in noise with fixed noise spectrum level at 30 dB SPL. As tone level increased (abscissa), the cell response decreased. These response properties could be predicted by model coincidence-detection cell which received convergent AN inputs.

responses of a model cell that received convergent AN inputs and responded based on a coincidence-detection mechanism.

The responses of a model coincidence-detection cell usually decreased when a CF tone was added to the noise (referred to as the phase-opponency mechanism, or PO, Carney et al., 2002). This decrease occurs because two AN fibers tuned very close to each other always had in-phase responses; in this case, there was little change in the correlation of their inputs when the tone was added. As the CF-difference between AN fibers increased, the phase between two AN responses to the tone signal (i.e. a signal with a frequency at the mean of the CFs of the two AN fibers) become out of phase first, which resulted in a decrease in model CD cell response. Two AN fibers with large CF-difference could have in-phase response to the signal but a high signal level is required to generate a change in the temporal correlation between the two AN responses.

From a functional point of view, both the PO mechanism and a lateral inhibition (lateral inhibition at low frequency is essentially a form of coincidence detection, see discussion in Shamma and Klein, 2000) could enhance the representations of the input spectral shape. The PO mechanism is purely temporal while the lateral inhibition mechanism is rate-based and requires non-saturated input rates (Rhode and Greenberg, 1994). Given that primarylike CN cells have dynamic ranges similar to AN fibers (May and Sachs, 1992), and that at low frequencies most cells have limited dynamic ranges (Winter and Palmer, 1991) as a result of little compression of the basilar membrane, the PO mechanism could be a more realistic way to process spectro-temporal information than lateral inhibition. Further studies with more controlled recordings from AVCN cells in the low CF regions could help to identify which of these two different mechanisms is responsible to CN responses. The coincidence cells should have multiple peaks in their response area at high levels and show a decrease in response rate when a CF tone is added to the complex sound at a level where the cells are already saturated.

3.5.2 Functional implications and relation to hearing impairment

Biologically inspired models based on coincidence detection have been used in explaining various auditory tasks. Some models require explicit delays, e.g., models for the processing of interaural time delays (Jeffress, 1948; Colburn and Durlach, 1978), or models for computing the correlograms for pitch (Licklider, 1951; Slaney and Lyon, 1993). Other models use cross-correlations to encode stimuli for specific tasks (Deng and Geisler, 1987; Shamma and Klein, 2000) where the delays are intrinsic within the spatio-temporal pattern of the AN responses. The present study, together with several other studies (e.g. Heinz et al., 2001; Carney et al., 2002), shows that fundamental features of the stimulus such as level information and spectrum shape could be conveyed in the spatio-temporal pattern of AN responses. Cross-frequency coincidence detection can process such spatio-temporal

information in complex sounds and convey this information to the CNS in the form of a rate code.

Cross-frequency coincidence detection relies on the relative timing of responses across different auditory filters. While most studies of hearing impairment have focused on the broad tuning and elevated thresholds of auditory filters due to the loss of cochlear compression, the temporal (phase) changes of the responses of the auditory periphery are ignored. These phase changes may be a critical factor in understanding functional implications of hearing loss. Damage to the auditory periphery affects both frequency selectivity and group delay (slope of the phase responses) of the auditory filter, and thus will change the spatio-temporal pattern of AN responses. Such changes will influence the responses of neurons that are involved in spatio-temporal encoding schemes and may explain the difficulties of hearing-impaired listeners in noisy environments (Moore, 1995).

3.6 Conclusions

The present study provides fundamental information about the responses properties of model AVCN cells that receive convergent AN inputs based on a coincidence-detection mechanism. These model cells were sensitive to the spatio-temporal pattern of AN inputs. The responses of model cells that received mixed-amplitude inputs (model cell *B*) were similar to AN fibers at low stimulus levels and had responses that were affected by the spatio-temporal pattern of their input AN fibers at high stimulus levels. For the tone-in-noise stimulus, the model cell response changed systematically depending on the tone frequency, making the cell capable of processing changes in spectrum shape which were encoded in the spatio-temporal patterns of the AN responses. The response properties of model cell *A* were not observed in CN cells from physiological experiments. However, the responses of such a model cell represented the maximum capability of the coincidence-detection mechanism to encode spatio-temporal information; several stages of processing may be required for the nervous system to extract temporal information completely based on a more realistic coincidence-detection mechanism. In general, the cross-correlation model was a reasonable and simple operation that could account for how the spatio-temporal pattern of AN inputs could be processed by neurons that act as coincidence detectors.

Chapter 4 Detection performance using a Coincidence-Detection Model: Identifying model cells sensitive to different cues

4.1 Abstract

A temporal model based on a monaural, across-frequency coincidence-detection mechanism was used previously to successfully predict simple psychophysical experiments, such as level discrimination (Heinz et al., 2001c) and tone-in-noise detection (Carney et al., 2002). This study provides a detailed evaluation of how the stimulus parameters in masked-detection tasks affect the predictions of such a temporal model. The model receives inputs from auditory-nerve (AN) fibers tuned to different frequencies, and its response changes dramatically when the phase relationship (temporal correlation) between the input AN fibers changed. The temporal pattern of AN fiber responses exploited by the model is determined by the signal-to-noise ratio, thus the model cell response is robust to changes in the overall level of the stimulus. As the bandwidth of noise decreases, the most sensitive model coincidence-detecting cell responds to both rate and temporal information in the input AN fiber responses. The model naturally combines rate and temporal information and predicts performance for masked detection under various conditions. The performance of other model detectors are also evaluated for masked detection of tones in noise. The other model detectors are based on the output energy (rate) of several auditory filters tuned to different frequencies and the envelope statistics (peakiness) of the auditory-filter outputs. These detectors are based on different aspects (cues) of the stimulus and thus predictions for the same experiment provided direct comparisons of how each cue could be used in masked detection.

4.2 Introduction

It is widely believed that data from psychoacoustical experiments have their basis in the physiological mechanisms of the human auditory system. It is thus a fundamental problem of auditory theory to relate perception to the underlying physiological properties of the auditory system. One useful approach to bridge our interpretation of the data from psychophysics and our understanding of the underlying physiological mechanisms or function has been to construct models that mimic (at least at some stage of processing) the physiological mechanism and to use them to predict human performance.

One successful example of this approach is the “power spectrum model” based on the critical-band (CB) hypothesis for the detection of a pure tone in masker noise (Moore 1995). In a classic experiment, Fletcher (1940) demonstrated that the detection threshold increases as noise bandwidth is

increased up to a certain value, known as the CB; the threshold is essentially constant with addition of noise energy outside the CB. The results were interpreted as an application of signal detection theory based on the energy of the output of the auditory filter centered at the tone frequency (Green and Swets, 1966, Chapter 8). The behaviorally measured auditory filter is widely believed to be a direct reflection of the filtering that occurs on the basilar membrane (BM) in the cochlea (Moore, 1986; Evans et al., 1989). These assumptions have been the foundation of many other masking experiments and have led to numerous quantitative studies of the auditory filter and related physiological properties of cochlear function (e.g. Patterson, 1976; Moore, 1978; Oxenham *et al.*, 2003).

However, the power spectrum model fails to predict human performance reported in several other studies. Studies from profile analysis (Green, 1988) and comodulation masking release (CMR, Hall *et al.*, 1984; Hall, 1986) have demonstrated that acoustic information outside the CB can improve signal detection[8], and human performance for detection in wide-band noise is only slightly disrupted when the overall level of the stimulus is varied from interval to interval (Kidd *et al.*, 1989). These results can be partly reconciled by applying channel theory to the responses of a population of auditory filters tuned to different characteristic frequencies (CF). When the different auditory filters are correlated in response to noise and uncorrelated (or have a different correlation) in response to the signal, the information in the responses of different filters can be combined optimally to decrease the uncertainty caused by the noise in the channel centered at the signal frequency to improve performance (Durlach et al., 1986; Kidd et al., 1991).

The power spectrum model generally ignores temporal information in the stimulus, which is apparently important for perception of various aspects of complex sounds and has been studied extensively in modulation discrimination and pitch perception (Eddins and Green, 1988; Houtsma, 1988). Several studies have demonstrated that temporal information in the stimulus, including the stimulus envelope and fine-structure, provide important information for masked detection in narrowband noise without the presence of reliable level differences (Kidd *et al*, 1992; Richards 1992; Richards and Nekrich, 1993). The same temporal information could also contribute to masked detection in wideband noise because narrowband peripheral filtering is widely accepted as the first fundamental processing stage in the auditory system. It is thus of great interest to see if a temporal model can predict performance for masked detection under various conditions and provide an alternative explanation of the underlying mechanisms that account for this simple psychophysical experiment.

Another problem for the "power spectrum model" is that it requires a reliable representation of stimulus energy, presumably in the responses of the peripheral auditory system. Such a representation is not clearly available in the

[8] Studies of information masking (see Watson, 1987; Neff and Green, 1987; Durlach et al., 2003) suggest that threshold could also be elevated due to maskers outside the CB.

discharge rates of AN fibers. The majority of AN fibers have a limited dynamic range and saturate in response to even moderate level sounds[9]. The existence of such fibers suggests that they might be useful in encoding information at sound levels even when their responses are saturated. The inclusion of low-spontaneous-rate, high-threshold, wide-dynamic-range fibers may provide rate information about the stimulus energy over a wide dynamic range (Winslow and Sachs 1988; Heinz *et al.*, 2001c; see review chapter by Delgutte, 1996) but the optimal use of average-rate information in the AN cannot account for human performance on a level discrimination task across a wide range of frequencies (Heinz et al., 2001a). It has been argued that robust performance in level discrimination could be accounted for by a model based on temporal information in AN fiber responses (Heinz et al., 2001a; Colburn et al. 2003). Thus stimulus energy may be represented in the spatio-temporal pattern of the population AN fiber responses.

In the present study, a temporal model based on a monaural, cross-frequency coincidence-detection mechanism (Carney, 1994; Heinz *et al.*, 2001c; Carney *et al.*, 2002) was evaluated in detail. The coincidence-detecting model cell responded to the synchronized inputs from model AN fibers tuned to different frequencies, and was very sensitive to changes of the phase relationship (temporal correlation) between the input AN fiber responses. The effects of the stimulus parameters and model parameters on the response properties of this temporal model were explored. The benefits of combining outputs of a population of model cells were illustrated, and model performance was examined by simulating the same tracking algorithms as used in psychophysical experiments. Specifically, the detection of a tone in noise with different bandwidth and with different roving-level conditions (Kidd et al., 1989) was simulated based on the model outputs. Additionally, the performance of several other models was evaluated in the second phase of the study. The alternative model detectors included an energy detector based on a population of auditory filter responses and an envelope detector. These different model detectors make use of different characteristics or cues available in the stimulus. A comparison of their performance for the same stimuli is helpful in clarifying which mechanisms are appropriate to explain masked detection.

4.3 Methods

4.3.1 Stimuli

Noise stimuli with different bandwidths were created digitally. In each desired bandwidth, a wideband Gaussian noise with a specified spectrum level

[9] Conversational speech is approximately around 60 dB SPL (from http://www.nonoise.org/) and human threshold for tones between 500 Hz and 4k Hz is about 10 dB SPL (Moore, 2003, pg. 129). Many AN fibers (66% in cat) are low threshold with dynamic ranges of approximately about 30 dB SPL (May and Sachs, 1992).

was first generated, and the frequency components outside the desired bandwidth were set to zero in the frequency domain. The noise stimulus parameters were matched to those used in the psychophysical measurements of masked detection that were the focus of this study (Kidd et al., 1989). The bandwidth of the noise was geometrically centered at the tone frequency of 900 Hz and varied from 10 to 3000 Hz. The spectral level of the noise was either fixed at 35 dB SPL or randomly varied across intervals over a maximum range of 32 dB centered at 35 dB SPL. The tone and noise were turned on simultaneously with a total duration of 250 ms, gated with a 20-ms cosine squared function.

4.3.2 Simulation of AN Fiber responses

A computational auditory-nerve (AN) model for the auditory periphery in human (Heinz *et al.*, 2001d (ARLO)) was used to simulate the AN responses. The model was based on a nonlinear model for the responses of AN fibers in cat (Zhang et al., 2001), modified to have bandwidths appropriate for human listeners (Heinz et al., 2001d). Different versions of the AN model can be used to study the role of several response properties associated with the cochlear amplifier, including level-dependent tuning, compression, and suppression, as well as the role of fibers with different spontaneous rates and thresholds (Heinz *et al.*, 2001d, 2002). The model AN responses presented here were produced using nonlinear model fibers with compression and suppression (Model #1 in Heinz *et al.*, 2001d) and with a high spontaneous rate (HSR) of 60 sp/sec. The analysis of models responses from low-spontaneous-rate (LSR) fibers was not included in the present studies because: 1) The temporal information in LSR and HSR fibers is similar (Johnson, 1980), and 2) physiological evidence suggests that compression in the low-frequency region is relatively weak (Cooper and Rhode, 1997), and thus low-CF LSR fibers have rate-level functions with limited dynamic ranges (Sachs and Abbas, 1974; Winter and Palmer, 1991).

Population responses of model AN fibers were generated based on a total of 41 model AN fibers centered at 900 Hz (the 21st model AN fiber had a CF of 900 Hz); model fibers were 0.2mm apart on a cochlear map. The CF of the AN fibers was determined by the human cochlear frequency placemap (Greenwood, 1990) given by

$$CF = A\,(10^{\,ax/L} - K), \qquad (4.1)$$

where x is the distance of the location from the apex and that other variables are constants (A=165, a=2.1, K=1.0, and L=35) chosen for human.

4.3.3 Monaural, across-frequency (correlation) coincidence-detection model

The temporal model used in the present study was a simple coincidence-detection model described in several studies (Colburn, 1973, 1977; Carney et al., 2002; Heinz et al, 2001c). Each model cell received two AN inputs and

discharged only when both input fibers had discharges within a narrow temporal window. When a very small coincidence window (*CW*=20μs in present study) was used, this model is similar to cross-correlation models (Colburn, 1977; Loeb et al. 1983; Deng and Geisler, 1987; Shamma and Klein, 2000), and the model responses are proportional to the multiplication of the instantaneous firing rate of the input AN fibers. To obtain model responses with discharge rates that were consistent with typical auditory neurons, N=10 identically driven inputs from each of two AN CFs converged on each model coincidence cell and the instantaneous firing rate of the model output was approximated by

$$R_{CD}(t) = R_{anf1}(t) \times R_{anf2}(t) \times CW \times N^2, \qquad (4.2)$$

where $R_{anf1}(t)$ and $R_{anf2}(t)$ are the instantaneous firing rate of the input AN fibers at two CFs, *CW* is the width of coincidence window, and *N* is number of identical inputs from each of two AN CFs.

This approximation assumes that the model discharges only when two input spikes from different CF fibers fall within the coincidence window (i.e. coincident spikes from fibers at the same CF are ignored here). In the analysis presented below, the steady state portion (100-200 ms) of the cell responses was used. This simplified operation has performance comparable to that of the coincidence-detection model based on a shot-noise model, which is more physiologically realistic but more computationally intensive (see Chapter 3).

The coincidence-detection model is sensitive to the spatio-temporal pattern of the input AN responses. The temporal AN inputs of a model cell in response to wideband noise only (top panel) and tone-in-noise (bottom panel) are illustrated in Fig. 4-1. The input AN fibers had CFs of 850 Hz and 950 Hz, respectively, and their responses to a 900-Hz pure tone differed by approximately 180 degrees of phase. The model coincidence-detection cell only responded at times when the two AN inputs overlapped (dark regions). The responses of two AN fibers were partially correlated to the noise stimulus, since some energy passed through both filters, resulting in occasional coincidences in their outputs. When the tone was added to the noise, the temporal pattern in both AN responses became dominated by the signal and thus the inputs to the model cell were out-of-phase with each other, resulting in a dramatic reduction in the coincidence-detector model cell responses. The systematic change of the spatio-temporal pattern in the AN responses with the addition of a tone to the noise thus provided information for the coincidence-detecting cell to detect the presence of the signal. Based on

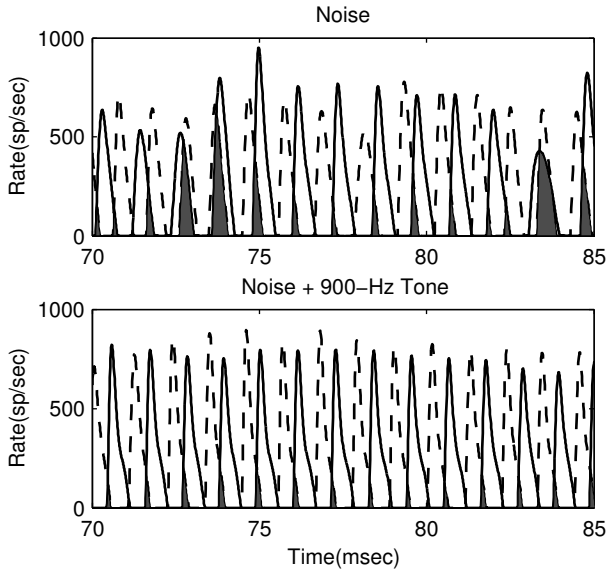


Figure 4-1: Responses of two model AN fibers with CFs at 850 Hz and 950 Hz to wideband noise only (top panel) and to tones-in-noise (bottom panel). The responses of two AN fibers to 900-Hz pure tone were out-of-phase (differed by approximately 180 degree) with each other. The responses of the two AN fibers to the noise stimulus had occasional coincidences in their outputs (dark regions). When the tone was added to the noise, the temporal pattern was dominated by the signal in both AN responses, which were out-of-phase with each other, resulting in a dramatic reduction in the coincidence-detector model cell responses.

the population of AN responses, a matrix of model coincidence-detecting cells was constructed. Each model cell received inputs from AN fibers at two CFs.

4.3.4 Quantitative measure of the detector performance and internal noise

A decision variable C was formed based on model responses to pairs of stimuli. The performance of such a detector was evaluated by calculating the sensitivity measure

$$Q = \frac{E[C(n,L)] - E[C(n)]}{\sqrt{(Var[C(n)] + Var[C(n,L)])/2}} = \frac{E[C(n,L)] - E[C(n)]}{\sqrt{V_{ex}}}, \qquad (4.3)$$

where n is the masker noise, and L is a tone level selected near the performance threshold. A similar metric was used by Heinz et al. (2001c (JASA)); Q corresponds to the commonly used sensitivity index d', which is a complete characterization of sensitivity when the decision variable C has a Gaussian distribution with equal variance under both conditions (noise-only and signal-in-noise). The variability in the decision variable C (or in the model responses) was caused by stochastic fluctuations across different noise samples (i.e. the "external noise", denoted by V_{ex}). A negative sign of the sensitivity measure Q indicated that the detection strategy was to look for a drop in the decision variable when a tone was added to the noise.

A variety of sources can account for internal noise (e.g., imperfect memory or lack of a constant decision criteria), including the stochastic nature of the neural activities. We could include an approximation of “internal noise” in the sensitivity metric with the assumption that individual neural activities could be approximated by Poisson processes (that is, the response variance was equal to the mean activity). If C represented an expected value of the model neural response, then the sensitivity measure of the detector based on M independent identical model cell responses is given by (Heinz *et al.*, 2001a)

$$Q = \frac{E[C(n,L)] - E[C(n)]}{\sqrt{V_{ex} + (E[C(n,L)] + E[C(n)])/(2M)}} = \frac{E[C(n,L)] - E[C(n)]}{\sqrt{V_{ex} + V_{in}/M)}}, \qquad (4.4)$$

where V_{in} represents the variance of "internal noise" from one single model cell.

The detector performance based on a population of responses C_i, where C_i represents the output of the ith channel, can be evaluated by (Durlach *et al.*, 1986)

$$Q^2 = [\Delta_1, ..., \Delta_N] G [\Delta_1, ..., \Delta_N]^T, \qquad (4.5)$$

where $\Delta_i = E[C_i \mid n,L] - E[C_i \mid n]$, N is the total number of the channels, and G is the inverse of the covariance matrix K. The covariance matrix of the population responses, K, which included the neural variability in each channel of M independent model cells was given by

$$k_{ii} = (Var[C_i(n)] + Var[C_i(n,L)])/2 + (E[C_i(n)] + E[C_i(n,L)])/(2M)$$
$$k_{ij} = (Cov[C_i(n), C_j(n)] + Cov[C_i(n,L), C_j(n,L)])/2, \quad (i \neq j) \qquad (4.6)$$

Note k_{ij} ($i \neq j$) has no internal component because cells are statistically independent, and k_{ii} includes the internal noise calculated based on the response count from M independent model cells for each channel[10]. A decision variable L can be formed based on simulated population responses

$$L = W\,[C_1, \ldots C_N]^T, \text{ where } W = [\Delta_1, \ldots, \Delta_N]G = [\Delta_1, \ldots, \Delta_N]K^{-1}, \qquad (4.7)$$

and the expected response difference Δ_i was measured for stimuli having signal-to-noise ratios indicated by human performance thresholds from Kidd *et al.* (1989). The detector constructed above is an optimal detector if C_i was Gaussian with the same variance in both conditions and varied linearly near the threshold level.

For all of the results presented here, 100 independent noise samples were used to compute the means and variances of the model responses. The results with no internal noise represented a lower bound for thresholds that could be obtained with a very large number of identical model cells, because the combination of responses across independent cells reduced the effect of the internal noise.

4.3.5 Population Model Coincidence Detecting Cells

A matrix of *CD* model cells was constructed from a population of model AN fiber responses (see Section 4.3.2). Each model cell in the matrix received different AN fiber inputs based on its position within the matrix. The sensitivity measure *Q* for cells at each location in response to different stimulus conditions was calculated for stimuli at the human detection threshold (Kidd *et al.*, 1989). For all the computations fifty identical and independent cells (M=50 in Eq. 4.4) were used[11], and *Q* actually represented the sensitivity of these 50 identical cells at each location in the matrix. We refer to cells with negative *Q* as "negative" cells, because the responses of these cells decreased with addition of the signal. The cells with positive *Q* are referred to as "positive" cells, correspondingly. A sensitivity measure of 1 or –1 indicates that the model predictions based on the cell responses was close to threshold, because the sensitivity measure *Q* is very similar to the d-prime measure. The sensitivity matrix was calculated using signal levels that matched human thresholds for

[10] Calculation of the inverse of the cross-covariance matrix without including internal noise for each channel may be impossible and also would not be optimal. By including internal noise for each channel, combining two channels that had the same output would enhance discriminability.

[11] Each location in the cell matrix represents 50 identical cells, which decreased the internal noise and increased the sensitivity for each matrix location. Since the cell responses were highly correlated (as we will see below), this was also similar to combining cells across different locations in the cell matrix, which decreased the internal noise but not the external noise. Combining cells across different locations in the cell matrix may decrease the total number of cells in the matrix, which is constrained by anatomyphysiology, see http://earlab.bu.edu/anatomy/Ratcell.aspx.

each bandwidth in the roving-level or fixed-level condition (Kidd et al., 1989), depending on the simulation.

4.3.6 Simulation of the Psychophysical task

To provide predictions that could be directly compared to experimental results, a two-interval, two-alternative forced-choice (2I2AFC) method with an adaptive 2-down, 1-up tracking algorithm was used (matched to Kidd *et al.*, 1989) to estimate the masked threshold for tones in either fixed-level or roving-level noise. Independent Gaussian noises were generated for each stimulus interval in the simulations. In the roving-level conditions, the spectrum level of each noise presentation was chosen randomly from a rectangular distribution (ranges of 4 dB or 32 dB) of levels centered around 35-dB SPL spectrum level. The noise spectrum level in the fixed-level condition was always 35 dB SPL. The 900-Hz signal was present at the level determined by the adaptive tracking procedure in one of the two presentations (intervals). The tone level in each track started at 25 dB re: N_0 and the initial step size was 4 dB. The step size dropped to 2 dB after 4 reversals, a total of 16 reversals was simulated for each repetition, and threshold was taken as the mean of last 12 reversals for each track. A track was stopped if the tracking level went above 50 dB re: N_0. Forty-two tracks were simulated for each condition to match the number of threshold estimates averaged in the Kidd et al. (1989) study, which included 6 estimates of threshold for each condition for 7 listeners.

A decision variable was computed based on the model responses. For coincidence-detecting model cells, the expected value of the discharge counts with latencies between 100 and 200 ms in response to a given stimulus waveform was computed for each stimulus interval (Eq. 4.2). This value varied across noise samples but did not include "internal noise" associated with the stochastic nature of the neural activity. A simple Poisson distribution was used to describe the spike counts of each model cell output. For all the results presented here, the decision variable used for the task was a sum of responses from 50 identical model cells. The randomness of the decision variable caused by the "internal noise" was approximated by adding a Gaussian[12] random variable to the mean output of total counts. The Gaussian variable has zero mean and variance equal to the mean output. In the case of sensitivity measurement, the variance was applied directly to Eqs. 4.3 and 4.4. The computation of decision variables for the other models will be described in the corresponding sections.

4.3.7 Energy-based detectors

Energy-based detectors typically assume that human listeners have a roughly constant threshold (in dB) for level (intensity) discrimination over a wide dynamic range. Numerous theories that have been developed to address the dynamic-range problem all assume that numerous AN fibers with different

[12] Even though spike counts of each model cell output were described by a Poisson distribution, the sum of N independent model cell outputs could be approximated by a Gaussian variable based on statistical theory.

response properties are involved (whether they have different CFs, or they have different rate-level functions, Viemeister, 1988; Winter and Palmer, 1991; Heinz *et al.*, 2002; reviewed by Delgutte, 1991). In this report, we followed a common practice in psychophysical modeling to avoid this problem. The energy output of the model *BM* filter (Zhang *et al.*, 2001) was calculated without any further processing by IHC or AN models. The *RMS* energy over the stimulus duration from 10-240 ms was converted into dB (referenced to 0 dB output) at the output of each channel (so it was dimensionless). Internal noise was simulated by adding a Gaussian random variable with zero mean and 1.5 standard deviation to the channel output. The value of the standard deviation was chosen so that the detector predictions in the narrowband noise for the fixed-level condition were close to human performance. For the detector based on multiple auditory filter outputs, five auditory filters evenly spaced along the basilar membrane (Greenwood, 1990) were used and were separated by roughly one critical bandwidth. The CFs of the auditory filters were 692, 790, 900, 1022, and 1158 Hz. A decision variable was formed based on the weighted combination of different channels (Eq. 4.7).

4.3.8 Envelope-based detector

The envelope-based detector used in the present study was similar to Viemeister's model (1979) for modulation detection except that the current model consisted of a narrowband basilar-membrane filter, followed by an inner hair cell (*IHC*), which was modeled as a half-wave rectifier followed by a low-pass filter. The first two stages of the processing used the corresponding components in the peripheral auditory model (CF = 900 Hz) described in Zhang *et al.* (2001), with parameters adapted for human (Heinz *et al.*, 2001d). The model *IHC* output was used instead of a model AN fiber response because the modulation response of a single AN fiber is affected by stimulus level (Joris and Yin, 1992), and a model with a wide dynamic range is required to explain the psychophysics. The low-pass filter was a first-order Butterworth with a cutoff frequency at 70 Hz. Before calculating the envelope statistics, the output of the detector was low-pass filtered with a 500-Hz cutoff frequency with an 8^{th}-order Butterworth low-pass filter. This filter was applied to reduce the noise in the calculation due to the carrier frequency (900 Hz); the bandwidth of the auditory filter at 900 Hz was approximately 100 Hz, so the envelope fluctuations at the output of this filter were unlikely to exceed 500 Hz. The average slope of the envelope was used as the decision variable and was calculated over a 170-ms window, omitting 30 ms after the stimulus onset, and normalized by the average of the envelope slope, x: $(1/n)\Sigma|x_i - x_{i-1}|/\Sigma|x_i|$ (Richards, 1992). The simulations were run at a sampling rate of 100 kHz, and internal noise was not included.

4.4 Results

In this section, the properties of a population of CD model cells in response to different stimuli is presented first, the strategy of combining the responses of a population of CD model cells in simulating psychophysical tasks and predictions is discussed next, and performance of the energy- and envelope-based models is evaluated last.

4.4.1 Effects of stimulus and model parameters on model responses

Single-cell model responses to tone-plus-noise stimuli

The model coincidence detector, as illustrated in Fig. 4-1, can predict human performance for tone detection in wideband noise (Carney *et al.*, 2002). To illustrate how different stimulus levels and bandwidths affected the model detector predictions, model cell responses are plotted in Fig. 4-2 as a function of the tone level (ref: N_0) for different stimulus conditions. The plots in the top panel show the rate responses of two input model AN fibers. The responses of both model AN fibers were saturated to noise alone and there was no significant change in the model AN responses with addition of the signal. Thus the change of the model CD cell response was mainly due to the temporal change of the AN fiber response pattern. The model CD cell responses to tone-in-noise

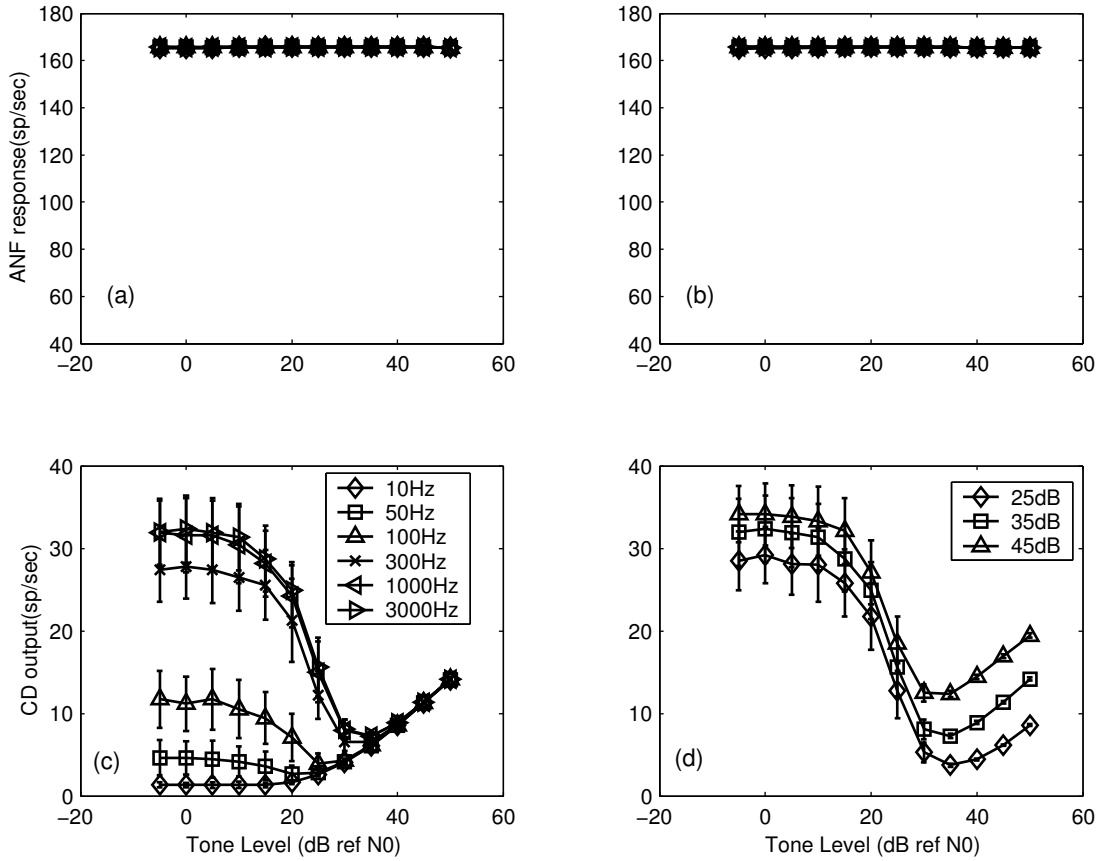


Figure 4-2: Model cell responses as a function of tone level (ref: N_0) for different stimulus conditions. The noise is geometrically centered at the tone frequency (900 Hz). The top panel shows the rate responses of two input model AN fibers with different noise bandwidths at fixed noise spectral level (35 dB SPL): (a) 850-Hz AN fiber responses; (b) 950-Hz AN fiber responses. Both model AN fibers had saturated rates in response to noise alone and there was no significant change in rate with addition of the signal. (c) CD cell responses at different noise bandwidths with fixed spectral levels of 35 dB. The cell responses to the narrowband noise were already negatively correlated, and responses remained low as signal level was increased. The increase in responses at high level was due to the phase changes of the AN fiber responses associated with broadening of the nonlinear auditory filters at high stimulus level. (d) CD cell responses at several noise levels (see legend) in 3000 Hz wideband noise. The changes of the model responses to different noise spectral levels was much smaller than that to different tone levels, and model responses were determined by the signal-to-noise ratio (tone level reference to N_0).

stimuli with various bandwidths are plotted in Fig. 4-2(c) (bottom left panel). The noise spectral level was fixed at 35 dB SPL and the bandwidth of the noise varied from 10 Hz to 3000 Hz (see legend). Generally the model CD response was a reflection of the correlation between the two input AN fiber responses, which depended on the relative energy passing through the overlapping region of the two auditory filters. In the narrowband condition, most energy in the noise stimulus was around 900 Hz and the AN fiber responses to the narrowband noise were already negatively correlated. Adding a signal to the noise would not change the temporal pattern in the AN fiber responses, thus the model cell response did not change much as the signal level increased. In the wideband noise condition, the input AN fiber temporal responses became more negatively correlated as the signal level increased, and the model CD cell response dropped dramatically at high signal levels. The model coincidence detector also had a limited dynamic range; its response rate did not change significantly at high tone levels. The responses increased slightly at high stimulus levels because the two CFs were no longer out-of-phase with each other due to changes of the phase properties of model AN fibers associated with the broadening of the nonlinear model auditory filters as input level increased.

Figure 4-2(d) shows the cell responses at several noise levels (see legend) with a fixed noise bandwidth of 3000 Hz. The changes of the model cell's response to different noise spectral levels was much smaller than the changes in response to different tone levels. So any significant change of the model response was due to the change of the tone level relative to the noise spectrum level. The model CD cell also responded similarly to similar tone levels referenced to N_0 for different noise spectrum levels, making the threshold for detection consistent across noise spectrum levels.

The responses of another CD cell are illustrated in Fig. 4-3. This model cell received two input AN fibers with CFs of 584 Hz and 959 Hz respectively. For the narrowband noise masker, the 584 Hz AN fiber was not saturated and therefore the AN fiber responses changed with tone level but were always less than the responses to tones in wideband noise. The responses of this model CD cell reflected both the temporal correlation between the input AN fibers and the rate responses of the input AN fibers. For the narrowband noise masker (centered at 900 Hz), the CD cell responses increased dramatically as tone level increased and were greater than the responses to the wideband noise with the same tone level. For this auditory filter with CF away from the signal frequency, the responses to tones in narrowband noise had much higher signal-to-noise ratios than the responses to wideband noise with the same signal level (ref: N_0). The input AN fiber responses were more temporally correlated for the narrowband noise at the same signal level, and this temporal correlation contributed to the higher response rate of the CD cell in the narrowband noise than that in the wideband noise. The responses of the CD cell for 100-Hz narrowband noise at different spectrum levels are plotted in Fig. 4-3(d) with the 100-Hz narrowband noise. The model responses were affected by both the noise spectrum

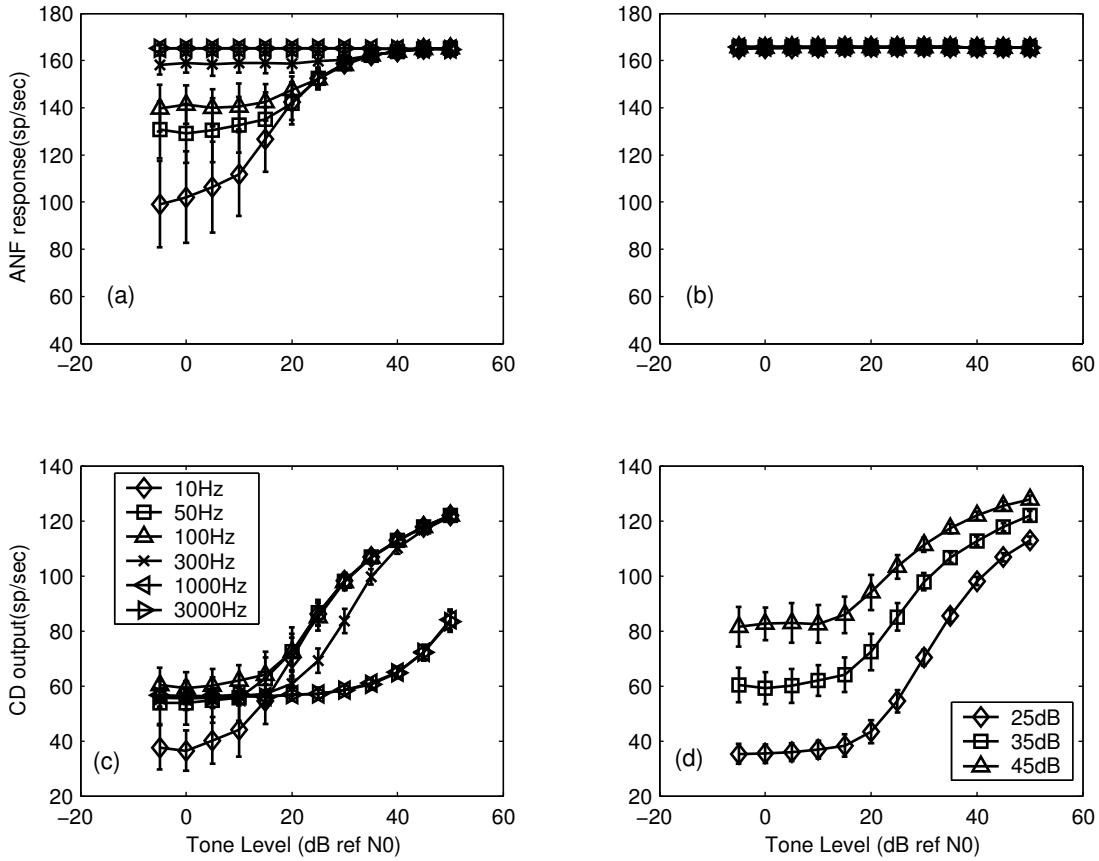


Figure 4-3: Similar plot as in Fig. 4-2 for another model CD cell with same stimulus conditions. The model CD cell received two inputs from 584-Hz and 959-Hz AN fibers. (a) The 584-Hz AN fiber was not saturated in response to narrowband noise and the responses for narrowband noise masker were always less than those to wideband noise masker. (b) 959-Hz AN fiber responses. (c) CD cell responses with different noise bandwidths. The model responses for narrowband maskers increased dramatically as tone level increased and were greater than the responses to the wideband noise with the same tone level, suggesting that changes in temporal correlation between AN fiber responses for narrowband noise masker contributed to the CD cell responses (for wideband masker, this happened at much higher signal levels). (d) CD cell responses at different noise spectral levels with 100-Hz narrowband noise. The responses were affected by both the noise spectral level and the tone level.

level and the tone level, and thus roving the noise level would greatly affect the performance of this model cell.

Responses of a population of model cells to tone-plus-noise stimuli

The plots in Fig. 4-4 are the sensitivity matrices for CD cells in the fixed-level condition for each noise bandwidth. The noise bandwidth increases from left to right and from top to bottom (a)-(f). The CD cell sensitivity changed systematically as the temporal relationship of the two AN inputs changed. In the narrowband noise condition, there were CD cells with negative sensitivity, even though the responses of input AN fibers increased as the tone was added to the noise. A CD cell became more sensitive when both the temporal correlation and the rate responses of the input AN fibers changed positively when the signal was added to the narrowband noise. Thus the most sensitive cell in the narrowband noise condition was a positive cell that received input AN fibers with different CFs (i.e. those cells *not* located on the bottom row in the matrix).

There were several consequences when the AN fiber responses became saturated in response to wideband noise. 1) There was less rate change for the AN fibers with addition of the tone, and sensitivity of some positive CD cells (especially cells with same-CF inputs) dropped; 2) the noise-only responses from AN fibers near signal frequency were more uncorrelated, and thus there was more temporal change when a tone was added to the noise; 3) the AN fibers with CFs away from the signal frequency required higher signal levels to change their temporal patterns. As a result, the most sensitive CD cells in response to wideband noise

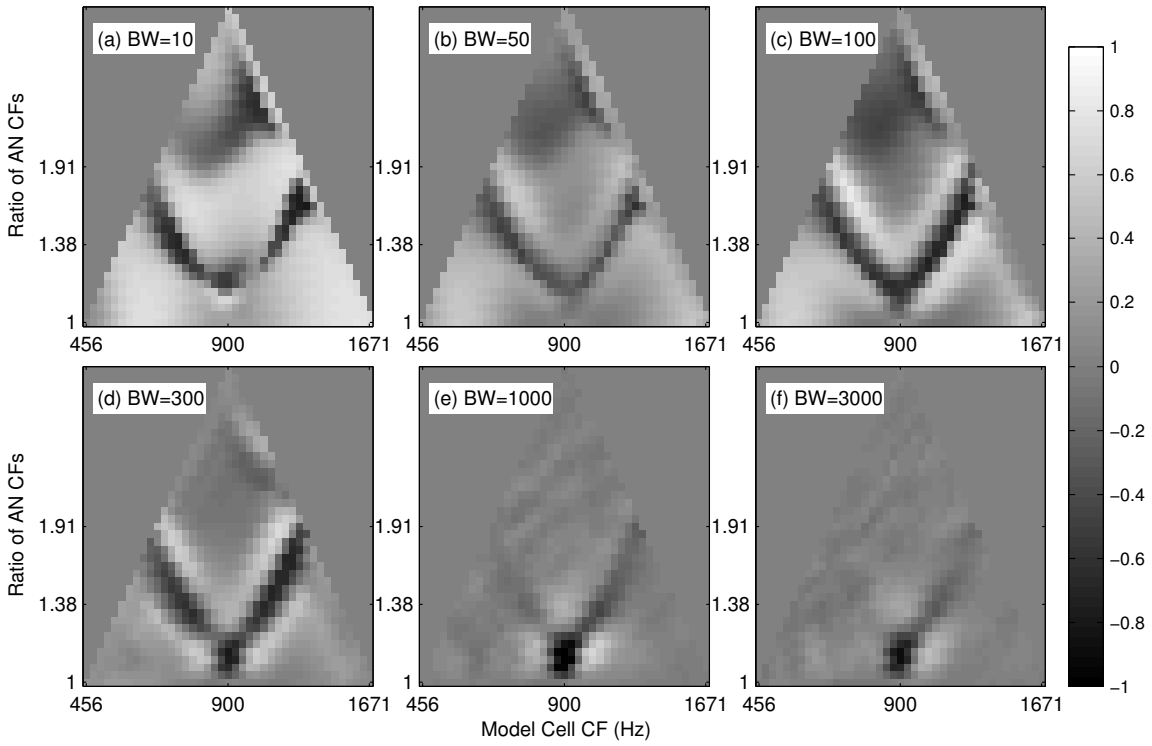


Figure 4-4: Sensitivity matrices for CD cells in the fixed-level condition for different noise bandwidths. The noise bandwidth increases from left to right and from the top to bottom (a)-(f). The abscissa of each plot represents the mean CF of the input AN fibers, and the ordinate represents the CF ratio of two input AN fibers. The sensitivity patterns were mainly determined by the temporal relationship of two AN fiber inputs (there were negative cells even though AN fiber responses increased as the tone was added for the narrowband noise masker). For the narrowband masker, the most sensitive cell was the positive cell that received different AN inputs (i.e. these cells not on the bottom row in the matrix). For the wideband noise masker, the most sensitive cells were the negative cells in the central region, which were surrounded by less sensitive positive cells.

were the negative cells near the center of the matrix, and the nearby positive cells had less sensitivity since they required higher signal levels to change their temporal patterns.

The sensitivity matrices for CD cells in the roving-level condition are plotted in Fig. 4-5. Higher signal levels (matched to human thresholds, Kidd et al., 1989) were used in the calculation of the sensitivity measure in the roving-level conditions. The threshold difference used between fixed-level and roving-level conditions was highest for the narrowband noise, but the sensitivities of positive cells still decreased in the roving-level condition compared to the sensitivities in the fixed-level conditions (Fig. 4-4). The negative cells in the sensitivity matrices had sensitivities that were comparable or even enhanced with respect to that in the fixed-level conditions (Fig. 4-4) because these cells were less dependent on the level changes of the stimuli. For the wideband noise masker, the threshold difference between fixed-level and roving-level conditions was small, and the pattern of sensitivity matrix was very similar for roving (Fig. 4-5) vs. non-roving (Fig. 4-4) conditions.

4.4.2 Psychophysical predictions based on the population of CD cell responses

Population Coding and correlation between the population CD cells

There are several ways to construct a detector based on the population of CD cell responses. The simplest strategy was to construct a detector based on a single cell that was most sensitive in each stimulus condition. The optimum strategy was to construct a detector that

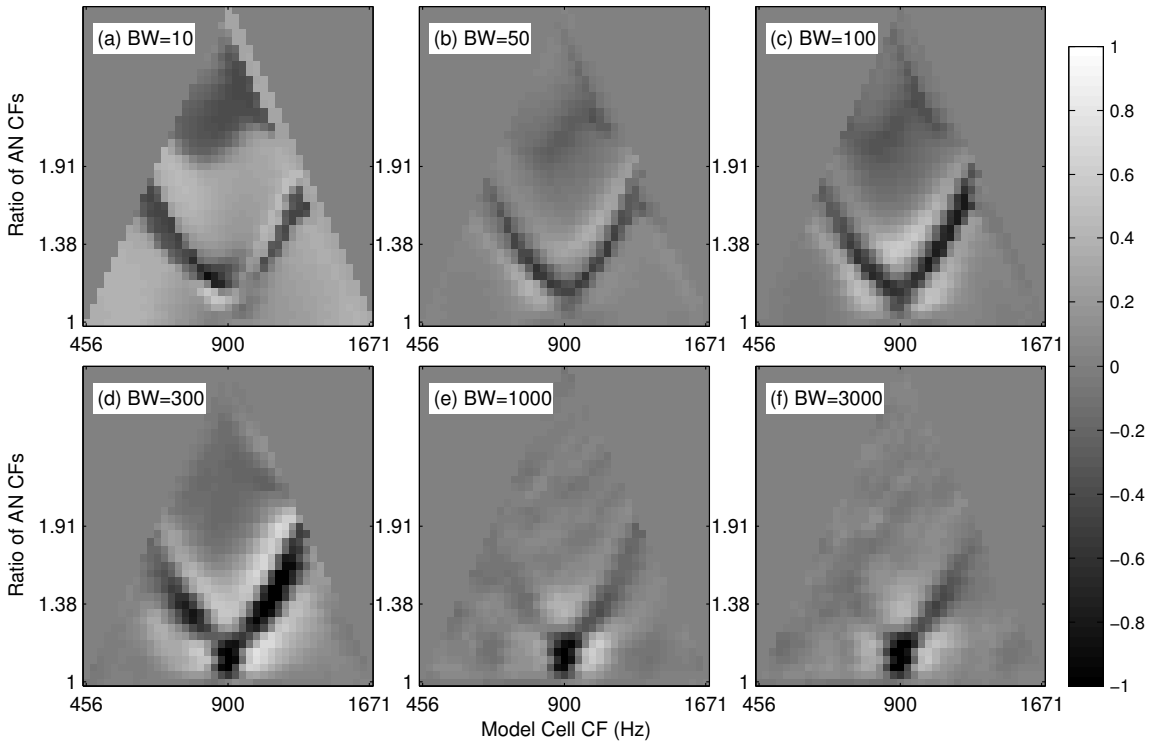


Figure 4-5: Sensitivity matrices for CD cells in the 32 dB roving-level condition. Even though higher signal-to-noise ratios (matched to human thresholds) were used in the computations for narrowband noise conditions, the sensitivities of positive cells decreased. The negative cells in the sensitivity matrices had comparable or even enhanced sensitivities because these cells were less dependent on the level changes of the stimuli. The sensitivity matrices were very similar under different conditions for wideband noise maskers.

combined the population responses optimally based on the covariance matrix K across CD cells (See Eq. 4.5-4.7). If all CD cells had independent responses, the sensitivity measure of the optimal process would be $Q_{sum}^2 = \sum_i Q_i^2$, where Q_i is the sensitivity measure of each individual model cell. This was only true when the response variance was caused solely by the internal noise of each individual CD cell, in which case the combination of population cell responses was similar to increasing the number of independent cells.

Another mechanism that can improve the performance by combining population responses applies if there were cells in the population that responded only to the noise, and if the cell responses to the noise were highly correlated. By combining different channel outputs optimally, the response variance caused by the noise stimuli (i.e. the external noise) could be effectively decreased (or suppressed). This is essentially the same mechanism (see Moore, 1988 for review of "dynamic aspects of auditory masking") that has been hypothesized to improve human performance in *CMR* or profile analysis, where outputs of different auditory filters are correlated either in the temporal domain (e.g. the envelope in CMR) or in terms of their power spectra, and only one auditory filter is assumed to change its response when the signal is added to the noise. In this case, the model cells can generally be classified into two types: those that respond to the signal and those that do not.

To test if such a mechanism could improve the performance of the population CD model, the optimum sensitivity measure was calculated based on the outputs of two CD cells: one was always the most sensitive cell (representing a cell that responded to the signal), and the other cell was chosen from the population of cells. The enhancement of the sensitivity measure was calculated as the difference between the optimum sensitivity based on the combination of the two cells and the sensitivity of the best cell. A matrix illustrating the enhancement of sensitivity by addition of the second cell can thus be constructed (Fig. 4-6); the enhancement of sensitivity was plotted at the location of the second model cell. The maximum enhancement of the sensitivity, calculated for the roving-level condition in which there was maximum external noise, was less than 0.2 for different bandwidths (Fig. 4-6). This value was much less than the maximum sensitivity of the best cell (around 1.0), so we could conclude that the advantage of combining outputs from different cells was small, and that the predictions based on a detector using the most sensitive cell would be similar to those of the optimum detector that combines information from multiple cells.

Model Predictions based on individual cells

Predictions based on individual cells in a set of model CD cells for fixed-level conditions are plotted in Fig. 4-7. The most sensitive cells (both positive and negative) at each noise bandwidth were chosen to be included in the simulation. The performance for each CD cell across different noise bandwidths

are connected with dotted lines; the solid line connects the best thresholds across this set of model cells for each

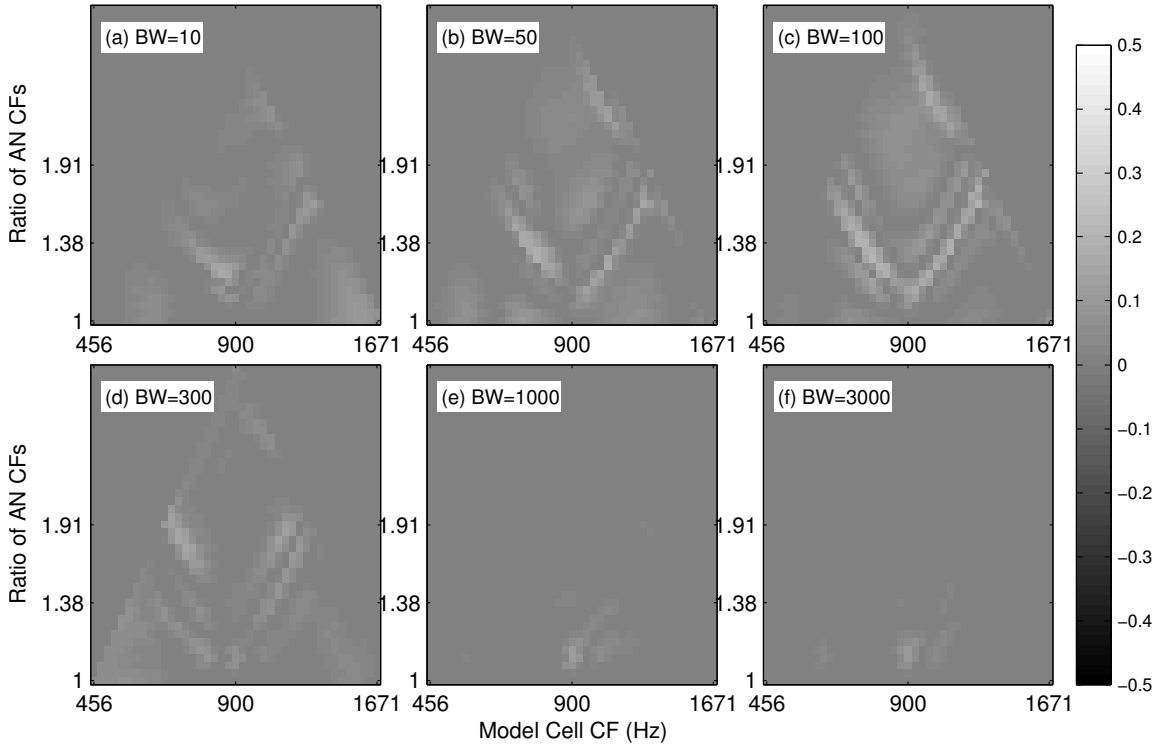


Figure 4-6: The enhancement of sensitivity in predictions based on a second cell combined with the most sensitive cell for each bandwidth. The enhancement is plotted at the location of the second cell. The most sensitive cell was selected separately for each bandwidth condition. The results were computed for the 32-dB roving-level condition. The maximum enhancement of the sensitivity was less than 0.2, which was much less than the maximum sensitivity of the best cell (around 1.0). (Notice that a different gray scale was used for this figure as compared to Figs. 4-4 and 4-5.)

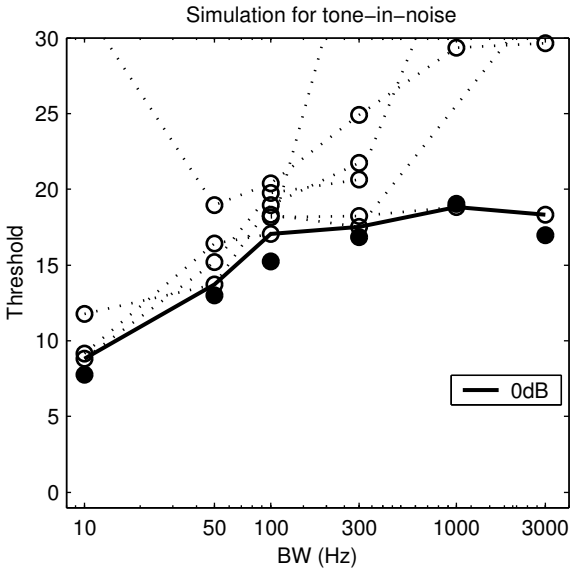


Figure 4-7: Predictions from a population of model CD cells for fixed-level conditions. The most sensitive cells (both positive and negative) in each noise bandwidth were used in the simulation. Each dotted line shows the performance for one CD cell; different cells had different thresholds for each noise bandwidth. The solid line connects the thresholds for the best model cells for each bandwidth, and human performance (replotted from Kidd et al., 1989) is plotted with filled symbols.

bandwidth. The predictions for different roving-level conditions are plotted in Fig. 4-8. The model predictions generally agreed with human performance: the threshold in the wideband noise was not affected by the roving-level paradigm, whereas the thresholds for narrowband noise were elevated by the roving-level paradigm. The location of the best cell used in each condition is plotted in Fig. 4-9. The same model cell was used in the wideband noise conditions for both roving-level experiments and fixed-level experiments. In the narrowband noise conditions, different cells were used: a positive cell had the best performance in the fixed-level experiment, and a negative cell was most sensitive in the roving-level experiment.

4.4.3 Onset responses of the CD model cells

The results presented above were all based on sustained responses of the model cells. In this section, we investigated the potential contribution of the onset responses. Figure 4-10 shows the sensitivity matrices based on the first 50 ms of the model CD responses in roving-level condtion. Including the onset responses of the model CD cells increased the sensitivity of positive cells in narrowband noise and degraded the performance of negative cells in wideband noise. This was because 1) the onset responses of AN fibers had larger dynamic ranges than sustained responses, and onset responses of CD model cells were more sensitive than the sustained responses to the level cues of the sitmuli; 2) Onset adaptation affected the timing of the AN fiber responses, the temporal information in CD cell onset responses was less reliable than that for sustained responses under roving-level condtions.

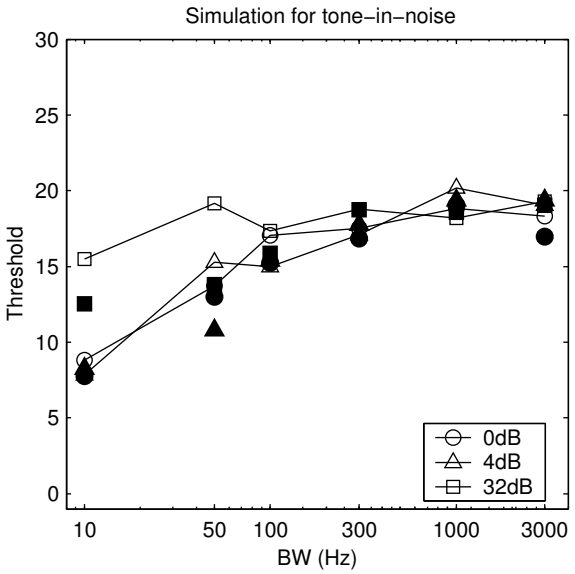


Figure 4-8: The predictions for different roving-level conditions from a population of model CD cells. Only the best predictions from the population CD cells are plotted and connected with solid lines. Human performance is plotted in filled symbols (replotted from Kidd et al., 1989).

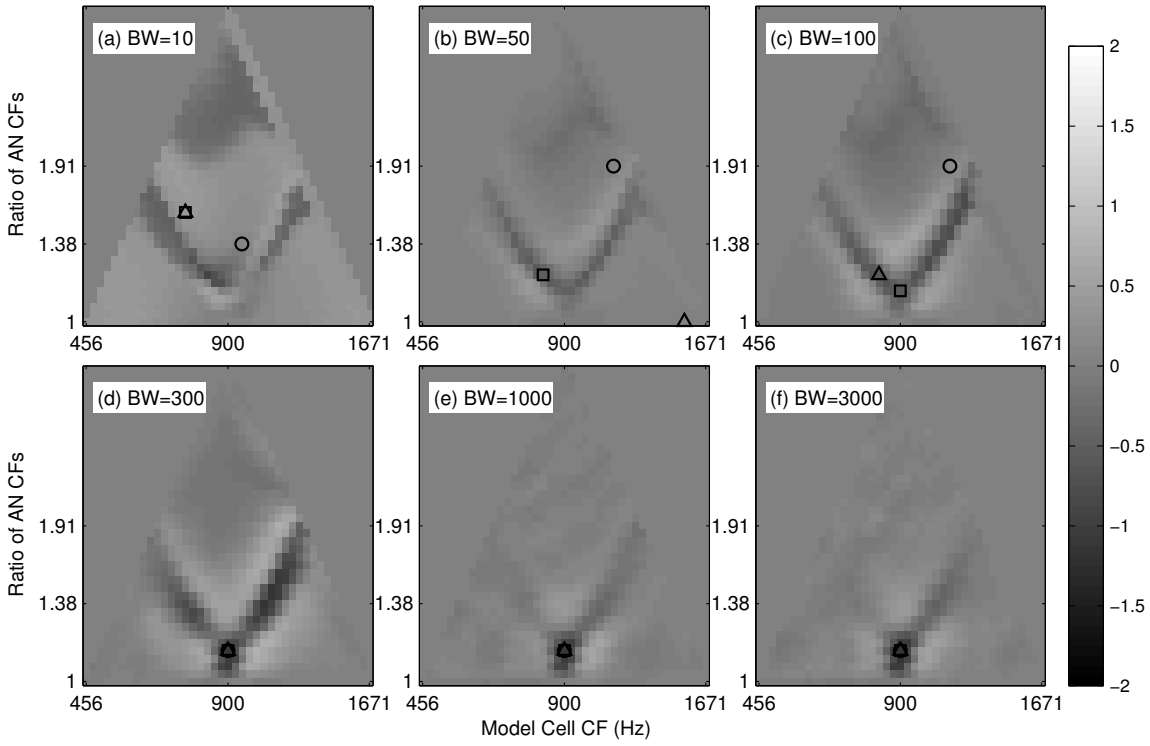


Figure 4-9: The location of the best cell used in each condition; symbols match those used in Fig. 4-8 (circles: 0 dB; triangles: 4 dB; squares: 32 dB). The sensitivity matrices for roving-level condition are plotted as the background to indicate the sensitivity of the cells. The same model cell was used in the wideband noise conditions for both roving level experiments and fixed-level experiments. In the narrow band noise conditions, different cells were used for different roving-level conditions: the cell that had the best performance in the fixed-level experiments was a positive cell, and the cell that was most sensitive in the roving-level experiments (except for 10 Hz narrowband noise) was a negative cell.

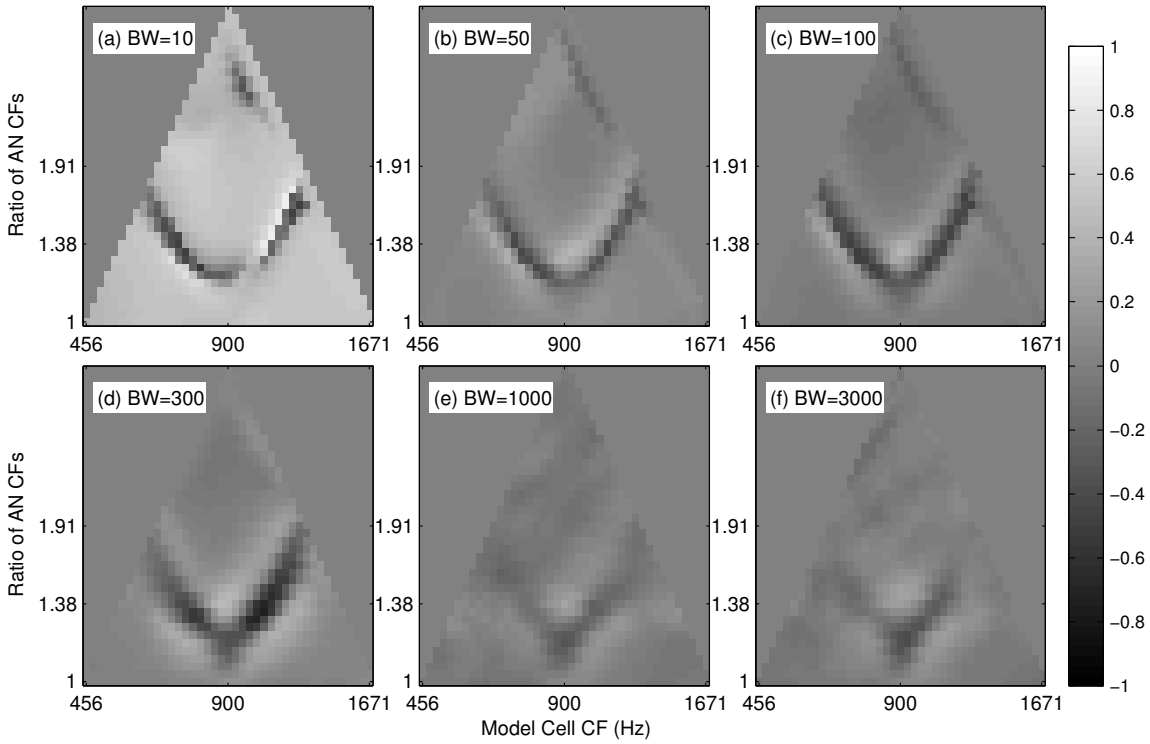


Figure 4-10: Sensitivity matrices for CD cells in the roving-level condition for different noise bandwidths. The first 50 ms of the model CD responses were used in these computations. Including the onset responses of the model CD cells increased the sensitivity of positive cells in narrowband noise and degraded the performance of negative cells in wideband noise.

Several studies support the assumptions that onset responses may carry more level cues than the sustained responses. Gilkey (1987) measured fixed-level and roving-level tone-in-noise detection thresholds using a much briefer stimulus (50 ms). He reported a small (about 2 dB) difference in threshold at a noise bandwidth of 1900 Hz between roving-level and fixed-level condtions. Richards (1992) measured the performance for detection of tones in noise with various durations and noise bandwidths; her results support the hypothesis that energy-based detection dominates only when the masking noise is relatively narrow and of a relatively short duration.

4.4.4 Predictions from other models

Results from an energy detector based on a population of auditory filter outputs

The energy detector combined the output from different auditory filters; some general properties of the energy detector are illustrated in Fig. 4-11. The sensitivity Q for each individual filter output is plotted as a function of filter CF in Fig. 4-11(a) for different roving-level conditions. The sensitivities were calculated for a 3000-Hz bandwidth noise. The filter centered at the signal frequency had the greatest sensitivity, and the sensitivity decreased quickly as the CF of the auditory filter moved away from the signal frequency. The maximum sensitivity measure (based on the 900 Hz-filter output) for the 32-dB roving-level condition was much smaller than the value of 1, suggesting that human performance in the roving-level condition cannot be explained based on the energy output of a single BM filter. The best sensitivity measured for

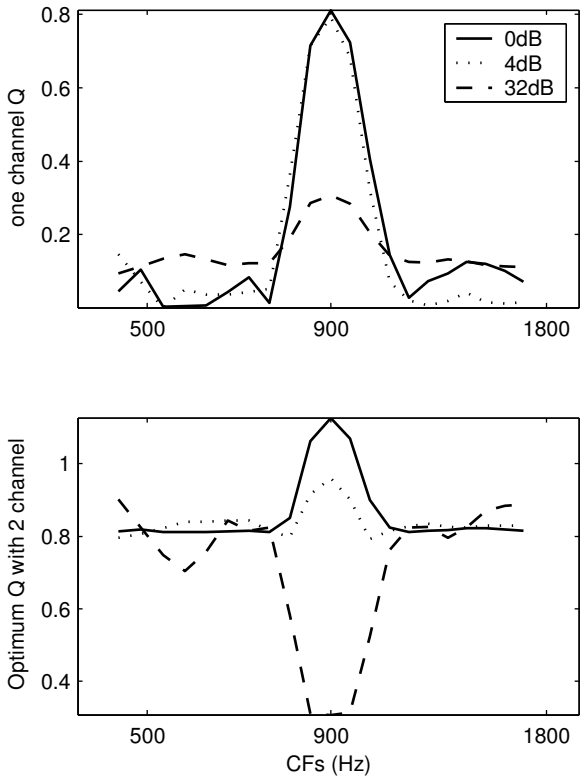


Figure 4-11: The sensitivity measure Q for the energy detector calculated using signal-to-noise levels matched to human threshold (Kidd et al., 1989) for 3000-Hz bandwidth noise masker with different roving conditions (see legend). (a) The sensitivity *Q* for each individual filter as a function of filter CF. The best sensitivity for the 4-dB roving-level condition was slightly better than that of fixed-level condition since a higher signal-to-noise ratio was used in the calculation (based on higher human thresholds for the 4-dB roving-level condition). (b) The optimum sensitivity measure based on two filter outputs. One filter was always centered at 900 Hz and the other varied across the abscissa. For the 32-dB roving-level condition, the optimum sensitivity increased dramatically as the second filter moved away from 900 Hz. In the fixed-level conditions, combining two filters both centered at 900 Hz decreased the internal noise, and the optimum sensitivity was larger than the sensitivity based on a single filter.

the 4-dB roving-level condition was slightly better than that of the fixed-level condition since a stimulus with a higher signal-to-noise ratio was used in the calculation (based on human thresholds). The optimum sensitivity measure based on two filter outputs is plotted in bottom panel of the figure. One filter was always centered at 900 Hz and the other varied across the abscissa. For the 32-dB roving-level condition, the optimum sensitivity increased dramatically as the second filter moved away from the 900 Hz. When two filters overlapped near 900 Hz, the second filter also responded to the signal and thus did not provide independent information about noise level (Durlach, 1986), so the performance of the detector was not improved much. In the fixed-level conditions, combining two filters both centered at 900 Hz decreased the internal noise and the optimum sensitivity was larger than the sensitivity based on a single filter. For the 4-dB roving-level condition, a combined effect of internal noise and external noise was observed. The sensitivity dropped and then increased a small amount as the second filter moved away from the signal frequency.

Figure 4-12 shows the optimum weights (see Eq. 4.7) of five auditory filters that were evenly spaced along the basilar membrane and were separated by roughly one ERB (see Section 4.3.6). The weights are plotted as a function of CF of the auditory filters for different roving-level conditions for each noise bandwidth in (a)-(f). Stimulus levels at the signal-to-noise ratio indicated by human thresholds in the corresponding conditions were used in these calculations. For narrowband noise, the weights were nearly flat since all filter outputs had similar signal-to-noise

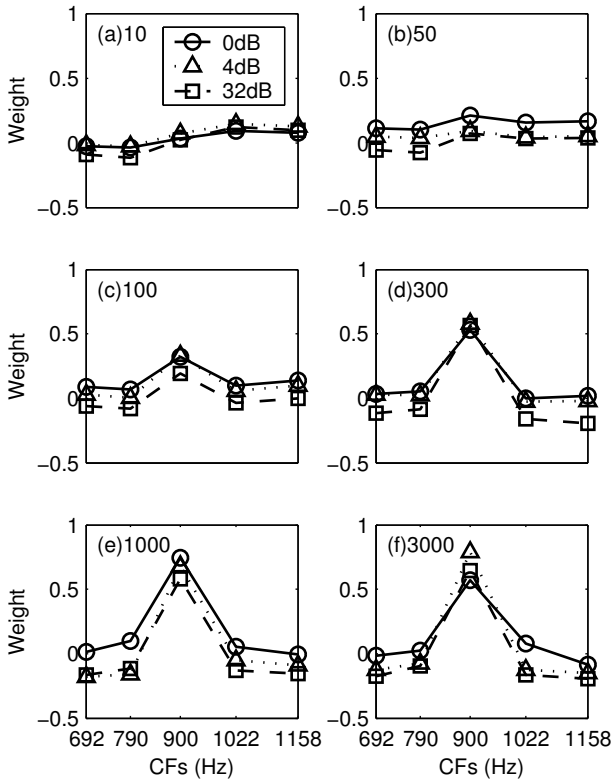


Figure 4-12: The optimum weights (see Eq. 4.7) of five auditory filters which were used as the optimal population energy detector. The weights are plotted as CFs of auditory filters for different roving conditions for each noise bandwidth in (a)-(f). Stimulus levels matched to the human thresholds in corresponding condition were used in these calculations. For narrowband noise, the weights were nearly flat since all filter outputs had similar signal-to-noise ratios and were correlated with each other. The weight of the center filter increased as noise bandwidth increased since this filter had a better signal-to-noise ratio at its output. In the wideband noise condition, the weight of the filters away from signal frequency were approximately zero for the fixed-level condition, suggesting that they contributed little information to the detector; the weights of these filters were negative for the roving-level condition, suggesting they were used as the reference of the noise energy output.

ratios and were correlated with each other. The weight of the center filter increased as noise bandwidth increased since this filter had a better signal-to-noise ratio at its output. In the wideband noise condition, the weight of the filters away from signal frequency were approximately zero for the fixed-level condition, suggesting that they contributed little information to the detector; the weights of these filters were negative for the roving-level condition, suggesting they were used as the reference of the noise energy output.

Threshold predictions based on a single filter centered at the signal frequency (900 Hz) are plotted in Fig. 4-13. For the fixed-level and 4 dB-roving condition, the model predicted human performance successfully, and the predicted threshold increased as the noise bandwidth increased up to the critical bandwidth value. For the 32 dB-roving condition, the model threshold was elevated across all bandwidths, and the detector based on a single filter output obviously could not account for human performance for the roving-level condition except at the narrowest bandwidth (10 Hz).

Predictions of the optimal population energy detector that combined responses of the five auditory filters are plotted in Fig. 4-14. For the fixed-level condition, the model predictions across noise bandwidths were similar to the predictions based on a single filter output. Combining information across different filter outputs decreased the predicted threshold for the roving-level condition in wideband noise. The optimal population detector did not improve performance for a narrowband noise

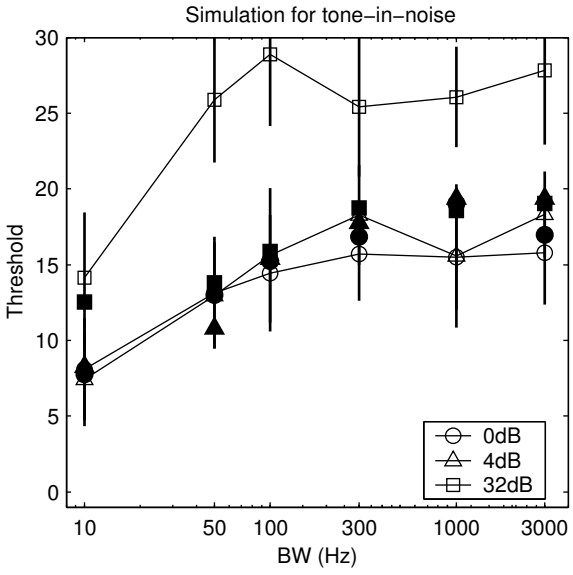


Figure 4-13: The predictions based on the energy in the response of a single filter centered at the signal frequency (900 Hz). The model predictions were connected with lines and the human performance were represented by solid symbols (replotted from Kidd et al., 1989). For the fixed-level and 4-dB roving-level condition, the model predicted human performance successfully, and the

predicted threshold increased as the noise bandwidth increased up to the critical bandwidth value. For the 32-dB roving-level condition, the model threshold was elevated across all bandwidths, and the detector based on a single filter output obviously could not account for human performance for the roving-level conditions except at the narrowest bandwidth (10 Hz).

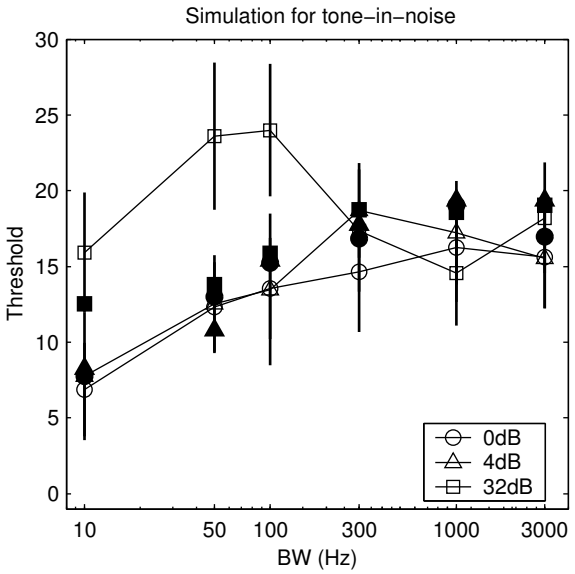


Figure 4-14: Predictions of the optimal population energy detector. For the fixed-level condition, the model predictions across noise bandwidth were similar to the predictions based on a single filter output. Combining information across different filter outputs decreased the predicted threshold to near the human threshold (filled symbols) for the roving-level condition in wideband noise. The optimal population detector did not improve performance for a narrowband noise masker, and thus failed to predict human performance for the roving-level condition in narrowband noise (50Hz, 100Hz).

masker, and thus failed to predict human performance for the roving-level condition in narrowband noise (50 Hz, 100 Hz).

Predictions from an envelope detector

The performance of an envelope-based detector for masked detection is plotted in Fig. 4-15. No internal noise was used in this simulation, and the predictions therefore represent the best performance of the detector. The model thresholds were near or below human thresholds in most cases, except for the fixed-level condition in the narrowest noise bandwidth (10 Hz). The roving-level condition slightly affected the model predictions due to the nonlinear *BM* filter used in the model. The model had slightly better predictions than human performance for the wideband noise masker. Experimental and modeling studies from amplitude modulation detection (Viemeister, 1979; Dau and Kollmeier, 1997b) suggest that envelope cue is interfered with by the envelope fluctuations from other auditory filters, thus the performance of the envelope detector would be expected to decrease when outputs of other auditory filters were considered in wideband noise condition.

4.5 Discussion

Critical band theory and the detection of tones in wideband noise

Critical band theory holds for both temporally based models and for the "power spectrum model." That is, the addition of a tone in wideband noise causes local changes (in either the temporal properties or amplitudes of the responses) of the basilar membrane, and the detection

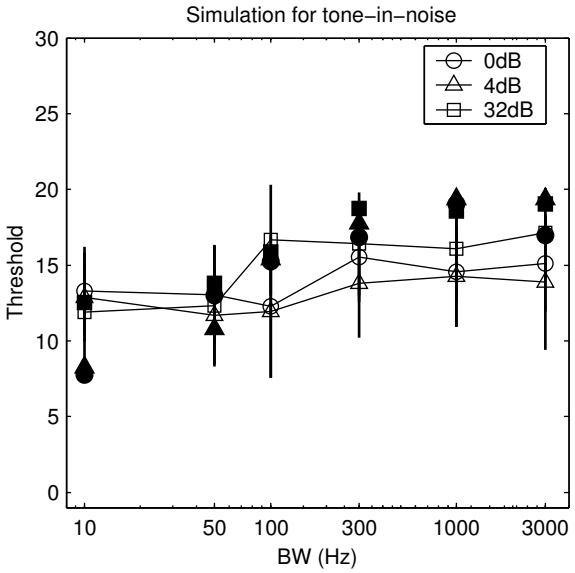


Figure 4-15: The predictions of an envelope-based detector. The envelope statistic was computed based on output of single auditory filter centered at signal frequency (the envelope detectors used in other studies usually integrate outputs across auditory filters; Viemeister, 1979; Dau et al., 1997). No internal noise was added to the envelope statistics and the predictions represented the best performance of the detector. The model predictions were near or below human performance (filled symbols) in most cases except for the fixed-level condition in the narrowest noise bandwidth (10 Hz). The roving-level condition slightly affected the model predictions due to the nonlinear *BM* filter used in the model.

threshold does not change when the noise bandwidth is larger than the critical bandwidth. The temporal changes in the population AN fiber responses are only determined by the signal-to-noise ratio, and such changes are robust to the roving level and not constrained by the limited dynamic range of AN fibers. The predictions based on the energy model are also robust to the roving-level conditions for wideband noise when the outputs from different auditory filters are combined together.

Envelope information was most effective for narrowband noise conditions because integration across different auditory filters made the change of envelope fluctuation in a single channel imperceptible in the case of wideband noise. Of course, it is possible that changes of the envelope fluctuation in a single filter would be perceived as changes in other qualities of the sound. The envelope cues could also be important for detection of a high-frequency tone in noise when fine timing cues are not available.

The detection of tones in narrowband noise

Human performance in detection in narrowband noise suggests that both energy cues and temporal cues are important to the subject. The detection threshold in narrowband noise is elevated by the roving-level condition, indicating that energy cues affect performance. The energy-based model failed to predict the threshold difference between the fixed-level condition and roving-level condition in the narrowband noises. Envelope cues can be used to account for human performance in narrowband noise (Kidd et al., 1989; Kidd et al., 1992; Richards, 1992; Richards and Nekrich, 1993), indicating that temporal cues are important in this condition. The monaural cross-frequency coincidence detection combines the energy and temporal cues, and its predictions agree with the human performance for both fixed-level and roving-level conditions.

Implication of the location of the most sensitive cell in monaural cross-frequency CD model

The most sensitive model cells used for detection in wideband and narrowband noise were different. This suggests that different strategies may be used by the subject for detection of a tones in wideband and narrowband noises. The perceptual quality of a narrowband noise is more tonal and is quite different from that of a wideband noise (Kidd et al., 1989). In the wideband noise, that fact that the same model cell was most sensitive in the roving-level and non-roving-level conditions suggests that subject might focus on the same "cues" in both detection tasks. However, in the narrowband noise, different model cells were most sensitive in the roving and non-roving conditions. This result suggests that subjects might shift their strategy (pay attention to different "cues" of the stimulus) in different conditions (Richards, 1991).

Neural implementation of the Cross-frequency coincidence-detection model

Several studies suggest that cells in AVCN respond like coincidence detectors and are sensitive to changes of the temporal pattern across different AN fibers (Joris et al., 1994; Carney, 1990). In the gerbil, some cells in the AVCN show a decrease in response rate when a tone is added to the wideband noise (Carney et al., in preparation), which is similar to the responses properties of "negative cells" in the present study. These studies also suggest that AVCN cells usually do not have response properties predicted by strong phase-opponency (PO) cell (e.g. model cell *A* in Chapter 3). The optimum cross-frequency coincidence detector may represent an abstract operation to process the temporal correlation across different auditory filters; this operation may actually involve neurons at several levels of the auditory pathway. Coincidence-detecting cells may be involved in other neural processing mechanisms at the same time. How the coincidence detection cells interact with each other, and how the nervous system combine different processing mechanisms is a topic of future study.

It is worth pointing out there are two distinct questions relating to the temporal and energy based models. The first is whether the temporal aspects of a stimulus or just the power spectrum of the stimulus underlies the behavior we observed. The second question is whether the neural system processes or encodes sound information based on the timing of the spikes or on the average rate of the spikes. Whereas these two questions are usually separated in studies of the central nervous system, they are typically combined in studies of the peripheral auditory system. While it is straightforward to find empirical evidence that temporal information could be encoded in the rate responses of some neurons (such as cells in the inferior colliculus that are sensitive to envelope cues; reviewed by Schreiner and Langer, 1988) and in the responses of cells in MSO that are sensitive to the ITD cues (Yin and Chan, 1990), it is also possible that the stimulus level (energy) is encoded in the timing of the neural spikes (Heinz et al., 2001a; Colburn et al., 2003). Under these circumstances the relative timing of discharges in different channels (or perhaps in the same channel) is more important than the precise timing of individual discharges; cross-frequency coincidence detection is a fundamental operation to process temporal cues across different channels.

Chapter 5 Evaluating performance of several models for level-discrimination in noise

5.1 Abstract

Performance in level discrimination of tones in noise is enhanced as the bandwidth of the noise increases in both fixed-level and roving-level conditions (Kidd et al., 1989). These data can not be explained by energy-based channel theory since either 1) there is no additional filter output that could be used to aid the signal detection in the narrowband noise condition; or 2) there is no across-frequency temporal or spectral correlation cues available for the channel theory in the fixed-level condition. The present study explored the ability of several models to explain these experimental results. Models based on detection of different cues in the stimulus were used, including 1) cross-frequency coincidence detectors, which are sensitive to temporal cues across different auditory filters; 2) a multi-channel detector that uses the energy output of several filters to discriminate spectral shape; 3) envelope detectors that are based on the envelope statistics of the filter outputs. The results suggest that there is more temporal information available for level discrimination in noise as the noise bandwidth increases. The enhancement of level discrimination in noise as bandwidth increases can be explained based on models that detect these temporal cues. The results also suggest that different types of temporal cues must be combined together to successfully predict human performance.

5.2 Introduction

Psychophysical studies of auditory masking are usually interpreted using the “power spectrum model” (Moore, 1995). This model assumes that the subject “listens” through a single auditory filter that provides the highest signal-to-masker ratio, and that the detection strategy is to measure the long-term power spectrum of the auditory filter output. This classical view of masked detection has been challenged, and there is convincing evidence that across-frequency comparisons are involved in signal detection. Two well-researched examples are comodulation masking release (*CMR*, Hall et al., 1984; Hall and Grose, 1988) and profile analysis (Green, 1988; Green et al., 1983). In *CMR*, enhancement of signal detection is achieved by making envelope fluctuations coherent across different frequency bands. The listener may use temporal coherence across different bands to aid detection, either by comparison of temporal responses from different channels or by detecting the signal during a dip of the envelope (Buus, 1985).

In a typical profile analysis experiment, the overall intensity is randomized within trials to ensure that absolute intensity is not a reliable cue, and the task is to detect an intensity increment to a pedestal tone in a multitone masker. Performance is improved when nonsignal tones are added to the multitone spectrum, even when they are placed well outside the bandwidth of a single auditory filter centered at the pedestal frequency (Green et al., 1983). The interpretation of the results is that filters outside the critical band that is centered on the pedestal frequency respond to noise only, and can be used to estimate the noise level. Listeners can base their discrimination on the difference between the pedestal plus increment and the background masker outside the auditory filter. The use of across-channel correlation to improve performance is well addressed by channel theory (Durlach, 1986). The masking caused by noise fluctuations can be effectively decreased by optimally combining outputs of multiple channels.

The detection of an increment to the pedestal tone in the presence of noise is also improved when the noise outside the auditory filter bandwidth centered on the pedestal frequency is added without any temporal or spectral correlation across different auditory filter outputs (Kidd et al., 1989; Plack and Viemeister 1992; Plack, 1998). Kidd et al. (1989) measured the threshold for level discrimination (detecting an increment of the signal added in phase to the pedestal tone) in noise with different bandwidths. They found that the threshold decreases as the noise bandwidth increases within the critical bandwidth for both fixed-level and roving-level conditions. The thresholds in wideband noise are constant and lower than the threshold in narrowband noise and are only slightly affected by the roving-level conditions. The spectral level of the noise is fixed in the fixed-level experiments, and there are no temporal or spectral correlations across-frequencies. Both channel theory and the power spectrum model fail to explain human performance in these conditions (Kidd et al., 1989). The only possible explanation is that cross-frequency comparisons generate a “categorical” representation of intensity that may be more reliable than “absolute” representations of intensity (Green, 1983; Plack 1998). This general hypothesis has not been explicitly stated and quantitatively tested using computational models, and the question still remains as to what is the neural mechanism underlying the enhanced performance for level discrimination in noise with increasing bandwidth.

In this chapter, level discrimination in noise was studied using the same temporal model proposed in Chapter 4. The coincidence-detecting model cells were sensitive to changes in temporal information when the signal was added to the pedestal, but the pattern of the sensitivity matrix was different from that for tone-in-noise detection. The enhancement of human performance as noise bandwidth increased could be predicted based on model cells that had similar level-discrimination resolution to each other. The performance of several other models, including an energy detector and an envelope detector, were also evaluated for comparison.

5.3 Methods - Stimuli and simulations

Noise stimuli were created in the same way as discussed in Chapter 4. The level of the 900-Hz pedestal tone was always fixed at 30 dB above the spectrum level of the noise. The signal was added in phase to the pedestal, and threshold was measured as the signal level referenced to the pedestal level. For example, a threshold of 0 dB corresponds to level difference of 3 dB between the pedestal and pedestal plus signal.

The model structures, parameters, and simulations used in the present study were modeled after Kidd et al. (1989), and were generally the same as in Chapter 4, in which they were discussed extensively (See Chapter 4.3). Any changes of methods for each model are discussed in detail in the following section.

5.4 Results

5.4.1 Sensitivity matrix of coincidence detection (CD) cells

In tone-in-noise detection experiments, the strategy is to compare the responses of a detector to noise-alone and to tone-in-noise stimuli. In level-discrimination-in-noise experiments, noise is present in both intervals but the tone level varies with respect to the noise spectrum level (N_0). The sensitivity measure for a detector that compares these two responses was defined as response difference divided by mean of the standard deviation of the two responses, and was given by

$$Q = \frac{E[C(n, L+\Delta L)] - E[C(n, L)]}{\sqrt{(Var[C(n, L+\Delta L)] + Var[C(n, L)])/2 + (E[C(n, L+\Delta L)] + E[C(n, L)])/2M}}, \quad (5.1)$$

where {C | n, L} represents the CD cell response to the pedestal tone at level L with noise masker n, ΔL is the level difference between the pedestal and the pedestal plus signal, and M is the number of identical CD cells used in the computation. The first term in the square root of the denominator represents the mean response variance caused by the external noise (n), and the second term represents the mean variance caused by neural stochastic discharges ("internal noise"), which decreases as the number of identical CD cells (M) increases. The value of Q^2 indicates the detector performance, similar to sensitivity index $(d')^2$ used in detection theory (e.g., Green and Swets, 1966).

The sensitivity measures for a population of CD cells for fixed-level conditions and roving-level conditions are plotted in Figs. 5-1 and 5-2. The value of Q at each location was computed using a signal level that matched human thresholds (from Kidd et al., 1989) and was based on the responses of 50 identical cells (M=50). Cells with positive values of sensitivity were referred to as "positive cells" since they increased their response when the signal was

added to the pedestal, and accordingly the cells with negative sensitivity values were referred to as “negative cells”. The distributions of the positive and negative cells for both conditions were very similar across different noise bandwidths, suggesting that the temporal correlation across the population of AN fiber responses was invariant under these different conditions.

In the presence of wideband noise, the location of the negative cells was shifted up, towards cells with AN inputs that were more separated in terms of CF (i.e. larger CF ratios), as compared to the sensitivity pattern from the tone-in noise detection experiments (c.f. Figs. 4-4 and 4-5). This is because the coincidence detectors responded to the temporal correlation across their AN inputs, and thus the pedestal tone at the signal frequency was a more effective *temporal masker* than the noise away from the signal frequency. The input AN fibers tuned further away (compared to the tone-in-noise condition) from the signal frequency attenuated the temporal contribution of the pedestal input, and addition of the signal to the pedestal resulted in larger changes in the temporal responses of these AN fibers. The CD cells that received such inputs were more sensitive than the CD cells that received AN fibers near the signal frequency, which changed little with addition of the signal due to the domination of the input AN responses by the pedestal tone. There was a consistent difference in human performance between fixed-level and roving-level conditions in the wideband noise. The higher signal levels used for simulations for the roving-level condition (chosen to match the higher thresholds for human listeners) yielded a higher sensitivity measure in the roving-level condition for wideband noise, suggesting that elevation of the human threshold in wideband noise was not predicted by the coincidence-detection model.

In the narrowband noise conditions, cell sensitivities measured at signal levels matching the lower threshold for the fixed-level condition (Fig. 5-1) were much higher than the sensitivities measured at higher levels for the roving-level condition (Fig. 5-2). The most sensitive cells were located at the edge of the sensitivity matrix; these cells were more sensitive because the AN inputs were less saturated at the lower stimulus levels. Also, for the narrowband noise condition, the pedestal tone dominated the cell responses, and there were small temporal changes when the signal was added. In the narrowband roving-level conditions, the sensitivity of these cells was greatly reduced, especially for the 50 and 100 Hz bandwidths, where using a higher signal level (to match human threshold) did not overcome the roving-level effect. As a result, a larger threshold for the model predictions was expected in the roving-level conditions.

Variations of cell responses due to external noise were different for narrowband and wideband noises. The pedestal tone was always 30 dB

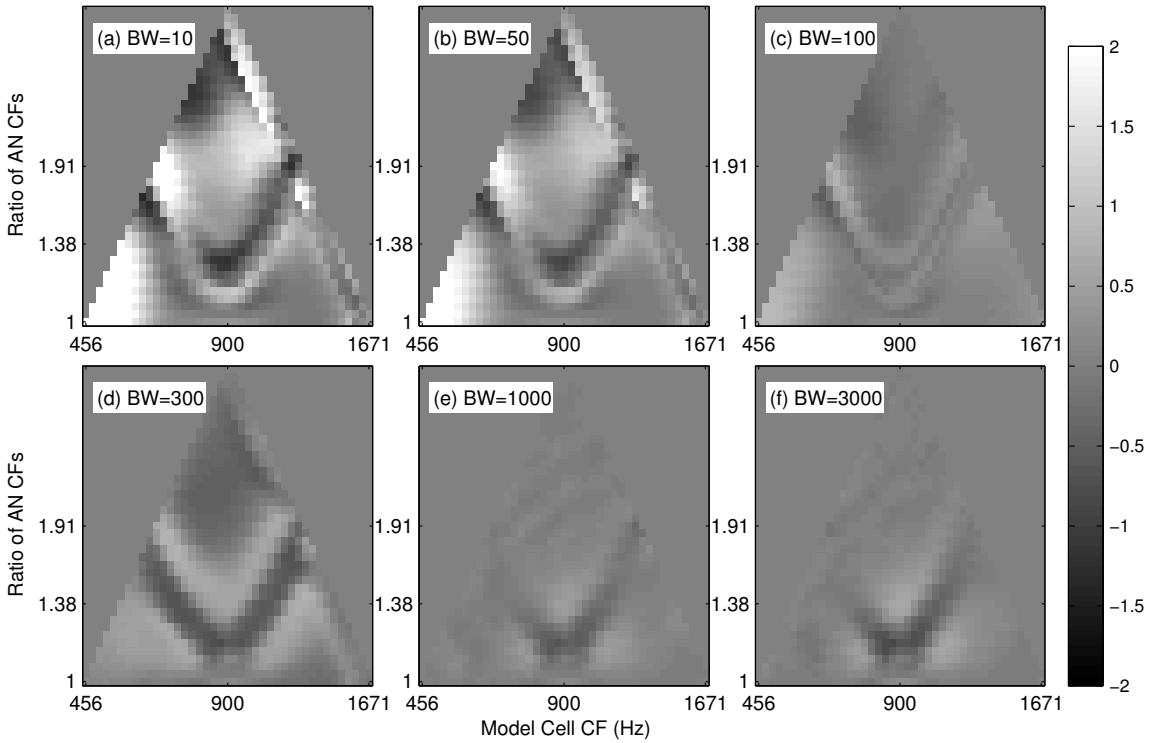


Figure 5-1: The sensitivity matrix of a population of CD cells for the fixed-level condition. Each sensitivity was computed based on 50 identical cells. The noise bandwidth increased from left to right and from top to bottom (a)-(f). The abscissa of each plot represents the mean CF of the input AN fibers, and the ordinate represents the CF ratio of the two input AN fibers.

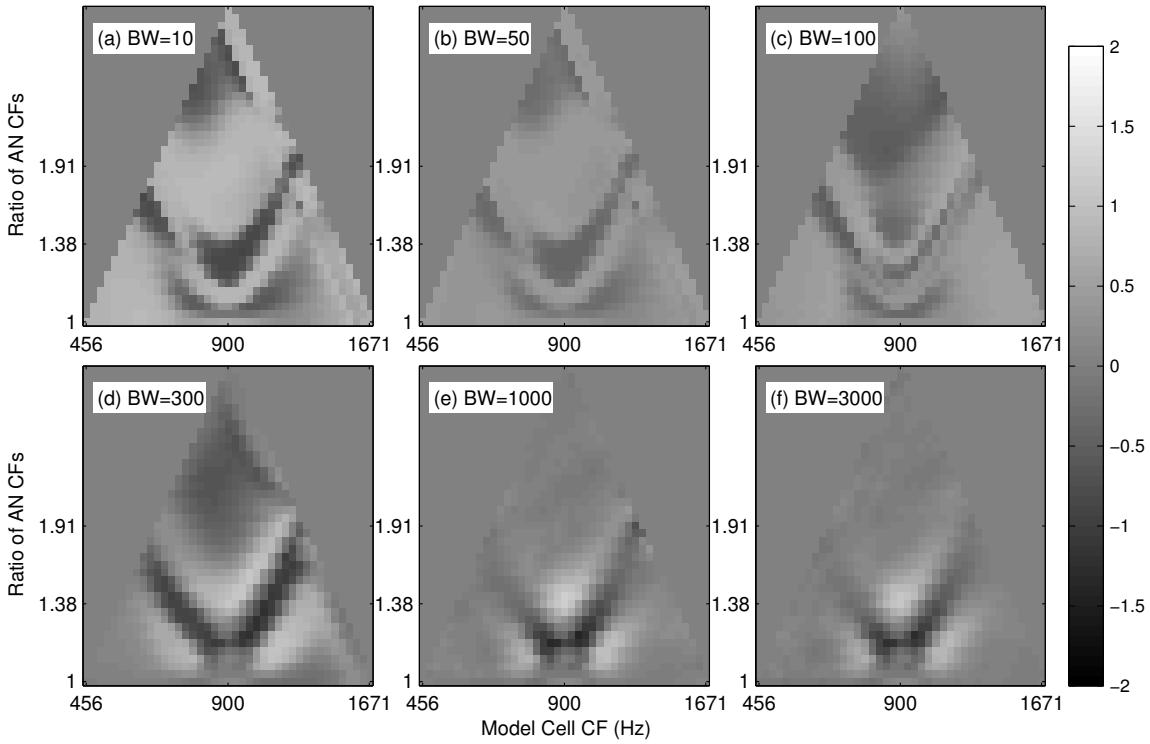


Figure 5-2: The sensitivity matrices for the 32 dB roving-level condition. Same format as for Fig. 5-1.

above the noise spectrum level. In the narrowband fixed-level condition, there was little fluctuation in the responses of the CD cells across noise samples, even when input AN fibers were tuned away from the pedestal frequency. This is illustrated in Fig. 5-1; in narrowband noise, the cells at the edge of the sensitivity matrix (which responded to the level change of the stimulus) had sensitivity much higher than 1.0 due to the small variance across stimuli. In the wideband noise, the most sensitive cells receive input AN fibers tuned away from the pedestal frequency. These AN fibers attenuated the pedestal level and boosted the contribution of the noise. In this case, small changes in the stimulus level could not be detected because of the response variation caused by the external noise.

5.4.2 Limiting the sensitivity of CD cells to changes in stimulus level

In the calculation of the sensitivity measure, each model cell was approximated by a Poisson process, and 50 identical model cells were used to estimate the internal noise caused by the stochastic nature of the neural activity. The internal noise estimated with this method was not enough to account for the limited resolution of human listeners to detect the level change of the whole stimulus. For the narrowband fixed-level condition, the pedestal was always 30 dB above the spectral level (which was 65 dB SPL) and there was little variance in the stimulus (both temporal and amplitude properties) for different noises. The CD cell responses to the pedestal plus noise were almost deterministic with different tokens of noise. This could be observed from the response properties of CD cells discussed in Chapter 4 (see Figs. 4-3c and 4-3d): the response variance in narrowband noise, for both AN fibers and CD cell, decreased dramatically as tone level increased; any small change in the signal (tone) at high levels could be detected unless sufficient internal noise was included to limit the threshold. For the model CD cells with two 490-Hz AN inputs (and with 50 identical CD cells), the predicted threshold for level discrimination was about 0.45 dB, which was well below human performance.

To address this problem, another central noise was introduced that limited the ability of the model cell in a level-detection task without external noise (i.e., the total level of the stimulus was varied and the same waveform was used across intervals). The performance of the detector for level-detection without external noise in our stimulus condition was evaluated as

$$Q = \frac{E[C_{\Delta L}(n, L) - C(n, L)]}{\sqrt{\sigma_L^{\,2}}}, \qquad (5.2)$$

where $C(n, L)$ was the model response to masker n with tone level at L, $C_{\Delta L}(n, L)$ was the model response to the same stimulus increased by ΔL dB SPL, and σ_L^2 was the variance of a Gaussian variable used to simulate the internal noise of level detection. If it were further assumed that the human threshold for level discrimination was Δ_{Lthr} (Q=1), the internal noise for the model that performed

the level-discrimination (not of the tone level, but of the total level of the stimulus) without external noise could be estimated as

$$\sigma^2_{\Delta_{Lthr}} = E[C_{\Delta_{Lthr}}(n,L) - C(n,L)]^2 . \qquad (5.3)$$

5.4.3 Predictions of CD model cells with limited level-detection performance

The decision variable was computed by adding a random deviate to the combined responses of 50 identical CD cells (see Chapter 4.3.6). The random deviate represented the internal noise associated with level-detection, and the variance was computed based on Eq. 5.3. For all the simulations, the resolution of level detection was limited to 1.5 dB (Δ_{Lthr} = 1.5 dB in Eq. 5.3 for tone-in-noise stimuli; Florentine et al., 1986, 1987), and the variance of the internal noise was estimated for CD cells at each location in the matrix under each noise band condition. The simulation was based on the pedestal in noise stimulus, and 300 independent noises were generated at a fixed spectrum level (35 dB SPL). The estimated internal noise varied from cell to cell and also varied at different noise bandwidths[13], but the same internal noise was used in different roving-level conditions.

The most sensitive cells (both positive and negative) in each noise bandwidth were used in the simulation. The best threshold for each bandwidth and roving-level condition from the population of model cells was selected as the model's threshold prediction for that condition. The model cells with the best thresholds were different in narrowband and wideband noise but were similar for each noise bandwidth in different roving-level conditions. Figure 5-3 plots the best thresholds across different conditions (symbols with lines) along with human performance (solid symbols). The predicted threshold in the wideband noise was lower than the threshold in the narrowband noise for both fixed-level and roving-level conditions, suggesting that the temporal information in the wideband noise was a more reliable cue[14] than the stimulus level change in both conditions. The coincidence-detecting cell with limited level-detection resolution was able to account for human performance in the wideband noise condition. In narrowband noise, the temporal cues that could be detected by the cross-frequency coincidence detection mechanism were not sufficient to account for human performance. The predictions of CD cells

[13] If the same internal noise for level-detection was applied for all the coincidence cells, this internal noise greatly degraded the performance of the cells that were sensitive to temporal cues. We cannot use "power spectrum" models with different internal noises for different channels to account for the improved performance in level discrimination in wideband noise, because the same model used in the wideband noise condition can then be used in the narrowband noise condition to provide performance that is better than that of human subjects. For the different values of internal noise used in present study, model cells with small amounts of internal noise did not provide more information about the level change of the stimulus than the cells with large internal noise value. The fact that different cells have different amounts of internal noise but similar level discrimination resolution suggest that these cells were in fact representative of different processes in the auditory system.

[14] Since we already limited the model resolution for level detection, so a better resolution (lower threshold) of pedestal level change in wideband noise is due to the temporal information carried by input AN fibers.

depended on the level difference of the stimulus and failed to predict the increased human thresholds in roving-level conditions for noise bandwidths of 50 Hz and 100 Hz.

5.4.4 Predictions of the "power spectrum model" based on multiple channel theory

General models of spectrum shape discrimination based on multiple comparisons of level across different auditory filter outputs have been discussed in several studies (Durlach, 1986; Green, 1988; Kidd et. al., 1991). The decision variable of the model is a weighted sum of multiple auditory filter outputs. The predictions from these studies are analytical and rather limited due to simplifying assumptions concerning each

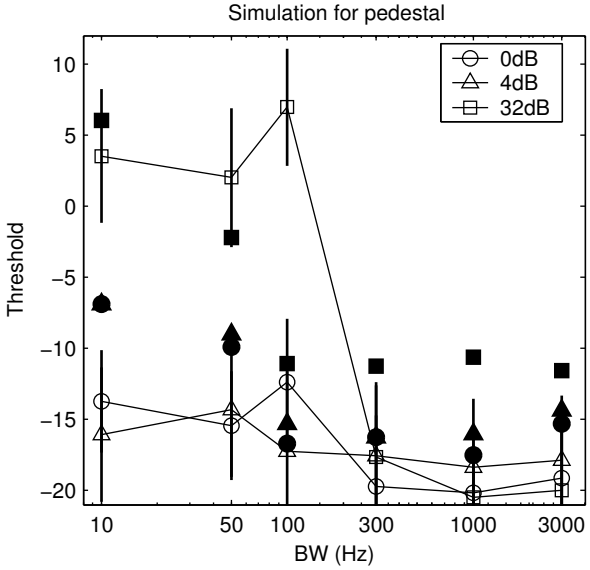


Figure 5-3: Predictions of level discrimination thresholds in noise from a population of CD cells. The most sensitive cells (both positive and negative) in each noise bandwidth were used in the simulation. The best threshold for each bandwidth and roving-level condition was selected from a population of model cells, based on the model prediction for that condition. Only the best thresholds for each condition are plotted (symbols with lines) along with the human performance (solid symbols, replotted from Kidd et al., 1989). The cell that gave the best threshold was different in narrowband noise and wideband noise, and the same cell gave the best predictions for each noise bandwidth in different roving-level conditions.

auditory filter output (i.e. they assume that only one auditory filter responds to the signal, and that all other auditory filters have the same responses to noise). In this section, we describe the predictions of a simple model based on the average energetic outputs of a population of auditory filters (see Chapter 4). Optimal weights of auditory filters were derived computationally, at signal levels matching human thresholds, based on the responses to 300 noise tokens[15] (see Chapter 4). This computational method avoided several limitations that were required for the optimal detector in previous studies (see Kidd et. al., 1991).

The optimal weights derived computationally for each noise bandwidth are plotted in Fig. 5-4. The optimal detector consisted of five filters and the weight for each filter was calculated for each roving-level condition (plotted in different symbols). Since the weights for the auditory filters could be scaled arbitrarily, the relative values of the weights across channels were more important than the absolute values. When the noise bandwidth was less than 100 Hz, the signal and noise were all within one critical band, and the optimal weights were flat across channels because the signal-to-noise ratio in the output was the same for each filter. When the noise bandwidth was greater than 100 Hz, the filters tuned to frequencies distant from the signal frequency had lower signal-to-noise ratios, so the weights of these auditory filters decreased. The filter that was dominated by the noise (the 1st and 5th filter, which were most distant from the signal/pedestal frequency) had weights of zero in the

[15] The optimal weights were derived based on the molar approach rather than on the molecular approach proposed by Gilkey and Robinson (1986).

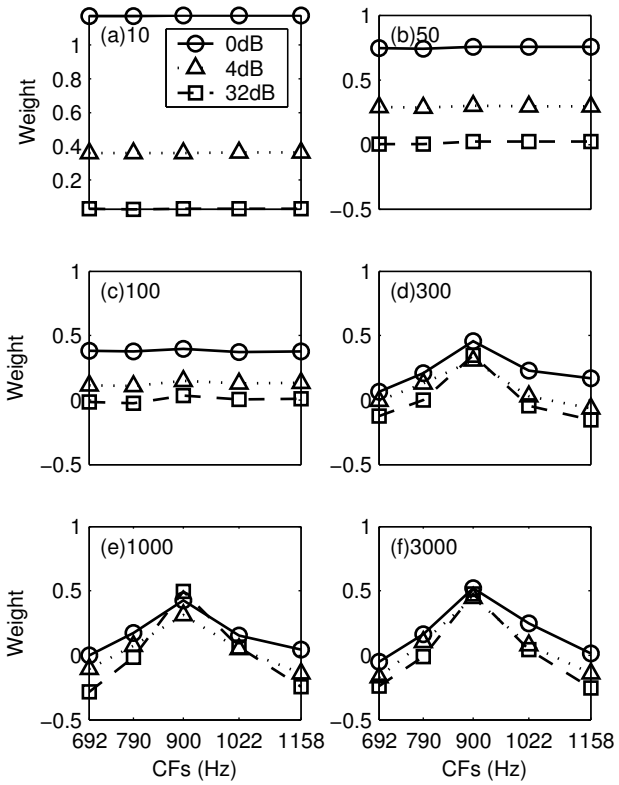


Figure 5-4: The optimum weights (see Chapter 4, Eq. 4.7) for five auditory filters with CFs of 692, 790, 900, 1022, and 1158 Hz. The optimal population energy detector was constructed based on the weighted sum of these filter responses. The weights are plotted as a function of the CF of the auditory filter in different roving conditions for each noise bandwidth in (a)-(f). Human thresholds in a corresponding condition were used in the calculation. For narrowband noise, the weight was nearly flat since all filters had similar signal-to-noise ratio and they were correlated each other. The weight of the center filter increased as noise bandwidth increased since it had better signal-to-noise ratio at its output. In the wideband noise, the weight of the filters away from signal frequency was about zero for the fixed-level condition, suggesting they did not change their response upon addition of the signal; the weight of these filters was negative for the roving-level condition, suggesting that they these cells may provide a reference for the estimation of the noise energy output.

non-roving conditions, suggesting that these filters did not respond to the signal. In the roving-level condition, the filters that did not respond to the signal could be used as a reference for the noise spectrum level, and thus could improve performance in the roving-level conditions. As a result, the weights of these filters were negative in the roving-level conditions. Another interesting observation is that for the filters with CFs near the signal frequency (but outside the critical bandwidth), there was still not enough noise *away from* the signal frequency to dominate the auditory responses (since the pedestal was 30 dB above the noise level); thus these filters had positive weights in the non-roving condition and had near zero weights in the roving-level conditions.

Figure 5-5 plots the predictions of the power spectrum model along with the corresponding results from experiments for level discrimination in noise (Kidd et al., 1989). In general, the predicted values did not fit the data very well. In the fixed-level conditions, the predictions from the model increased slightly as the noise bandwidth increased. The output of the auditory filter centered at 900 Hz was dominated by the pedestal tone and changed very little as noise bandwidth increased (the pedestal was 30 dB above the noise spectrum level, and the critical bandwidth at 900 Hz was about 100 Hz), thus the level of the signal that was required to maintain a constant signal-to-noise ratio as noise bandwidth increased was small.

In the roving-level conditions, the power spectrum model did predict trends in human performance to some extent. The predictions decreased as the noise bandwidth increased, and the threshold approached that of

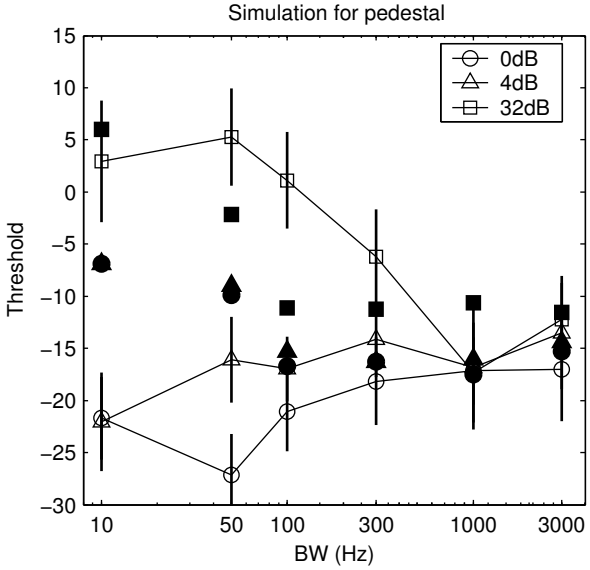


Figure 5-5: Model predictions based on the energy in the output of multiple auditory filters. The model prediction is connected with lines and human performance is plotted with solid symbols (replotted from Kidd et al., 1989). Each prediction was made using the optimum weights calculated for each condition from Fig. 5-5. For the fixed-level and 4 dB-roving condition, the model predictions increased slightly as a function of the noise bandwidth, which was not consistent with human performance. For the 32 dB-roving condition, the model predictions generally decreased as noise bandwidth increased. The predicted threshold was elevated across all of the narrowband noise bandwidths, and the performance in 300 Hz was still much worse than the corresponding human performance and the model prediction for the fixed-level conditions. When the noise bandwidth was larger than 1000 Hz, the model benefited from the multiple channel comparisons and thus the performance was robust in the roving-level conditions.

the non-roving conditions in the wideband noise. However, the model predictions did not decrease as quickly as observed in human performance. The model predictions for narrowband noises were all affected in the roving-level conditions and the threshold was elevated, since the multiple channel comparisons did not benefit the model at all for the narrowband noise. The model also failed to predict human performance in 300 Hz wideband noises, which had a similar threshold as for the other wideband noise conditions. The model filters adjacent to the filter centered at the pedestal frequency still responded to the pedestal at the signal frequency and could not be used as a reliable reference for the noise level. Thus, the power spectrum model predictions for the roving-level condition in 300-Hz bandwidth noise were much worse than in the fixed-level condition, as expected.

5.4.5 Predictions based on the envelope model

A simple envelope-based model was described in Chapter 4 (see 4.3.7) to extract the envelope cues from a single auditory filter output centered at the signal frequency. The normalized average absolute slope s of the envelope x: $s=(1/n)\Sigma|x_i - x_{i-1}|/\Sigma|x_i|$ (Richards, 1992), was used as the decision variable in the simulations, and the detection strategy was to detect a drop of the average slope when a tone was added to the pedestal in the noise.

The model predictions without any internal noise (i.e. the decision variable was deterministic for a given stimulus waveform) are plotted in Fig. 5-6 along with human performance (solid symbols). Generally the predictions outperformed human listeners across all noise bandwidths,

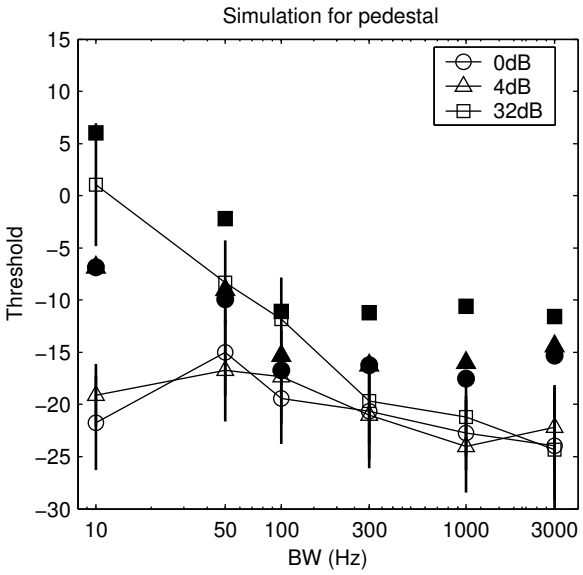


Figure 5-6: Predictions based on the envelope detector. Details of the envelope detector were discussed in Chapter 4 (see 4.3.7). The decision variable used was the normalized average slope of envelope (Richards, 1992). No internal noise was added (the decision variable was deterministic for a given stimulus waveform) so the predictions represented the best performance that the model could achieve. Results for different roving-level conditions are plotted with different symbols. Generally the predictions (symbols with lines) outperformed human performance (solid symbols) across all noise bandwidths. The model results predicted the increase in threshold as noise bandwidth increased. The super-performance for the fixed-level condition in narrowband noise was due to the nonlinear filter (cochlear compression) and can be avoided by introducing some small internal noise to the decision variable (See Figs 5-8 and -9).

and the performance generally improved as noise bandwidth increased. The performance in the fixed-level condition in narrowband noise is much better than the human performance and can be attributed to the following two factors: 1) For the pedestal tone in a narrowband noise, the fluctuation of envelope statistics (here the average envelope slope) was extremely small and the decision variable varied very little across different noise waveforms (i.e. external noise was very small). A small change (less than 1%) in the decision variable could be detected when the signal was added to the pedestal in noise. 2) The envelope of the stimulus was compressed more by the nonlinear filter (or cochlear compression) at high levels than that at low levels (Moore, 1995; Oxenham and Bacon, 2003). This change in the envelope due to the cochlear non-linearity was detectable by the envelope detector. In the roving-level condition, the envelope change due to cochlear compression was not reliable and thus the model predictions were greatly elevated. In the wideband noise condition, the envelope change due to cochlear compression was masked by envelope fluctuations (i.e. the variance of the envelope statistics) caused by the external noise, so the performance of the envelope detector was similar for both fixed-level and roving-level conditions.

To illustrate how the nonlinearity of the auditory filter affected the model predictions, envelope-detector predictions based on a linear auditory filter (Heinz et al. 2001d, model number 3) are plotted in Fig. 5-7. Since the normalized envelope of the linear filter was the same across different stimulus levels, the model predictions for the fixed-level

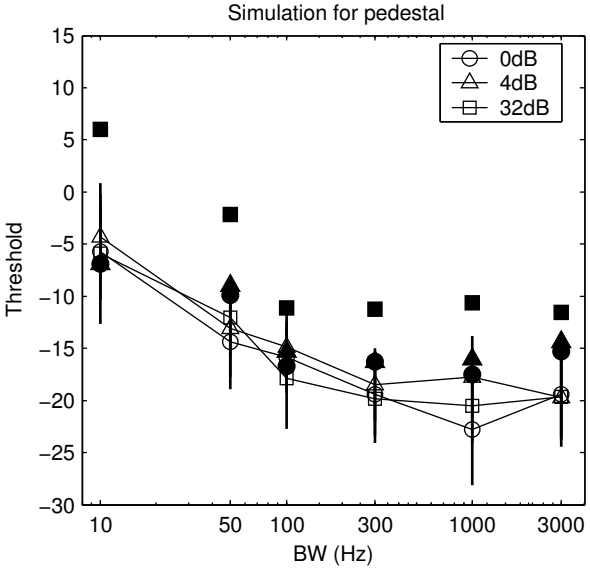


Figure 5-7: Predictions based on the envelope detector with a linear filter (Heinz et al., 2001d, model number 3). The model without the cochlear non-linearity had similar predictions in the fixed level condition as in the roving-level conditions.

condition were very similar to the results in the roving-level conditions. The results were also similar to the results with the nonlinear filter (Fig. 5-6) under roving-level conditions, where the envelope cues due to the nonlinear filter were not reliable.

Internal noise had different effects on model performance depending on differences in the variation of envelope statistics for different bandwidths. Fig. 5-8 plots the predictions based on the envelope detector (with the nonlinear filter) with internal noise added. The decision variable was computed by adding a Gaussian random variable with zero mean to the envelope statistic. The variance of the random number was always 0.01 (1%) of the envelope statistic. In the narrowband noise condition, the model predictions were greatly affected by the internal noise since external noise was very small; the change of envelope statistics due to the nonlinear filter was masked by the internal noise, and the predictions were similar across different roving-level conditions. When the noise bandwidth approached the critical bandwidth[16], there was more variance in the envelope statistics caused by external noise, and the addition of the internal noise had little effect on model performance.

[16] It is also worthwhile to point out that in the wideband noise, the predictions of the envelope detector out-performed humans. Usually envelope changes in a single channel are not considered to be a perceptual cue in wideband noise since various studies (Viemeister, 1979) suggest that perception of envelope fluctuations is based on the integration of information across a wide range of auditory filters.

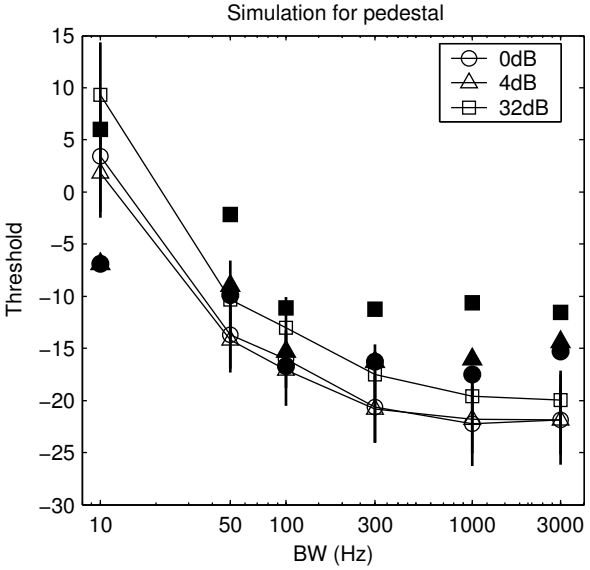


Figure 5-8: Predictions based on the envelope detector with internal noise. The internal noise was approximated by adding a Gaussian deviate to the decision variable. The variance of the random deviate was always %1 of the decision variable. The envelope detector used the same nonlinear filter as in Fig. 5-7. The performance in the narrowband noise was elevated by the internal noise since the variation in envelope statistics caused by the external noise was small. Also the envelope changes caused by the nonlinear filter were masked by the internal noise, and thus the results were similar across different roving-level conditions.

5.5 Discussion

Multiple cues, multiple processes, and strategies used in level-discrimination in the presence of noise

The results in the present study show that the fact that there is more temporal cues available when a signal was added to the pedestal in noise as noise bandwidth increased could be attributed to at least two observations from studies of level discrimination of tones in noise: First, human performance improves as noise bandwidth is increased within a critical bandwidth (Kidd et al, 1989; Kidd et al., 1992); and second, human performance is best in wideband noise, even when there is no correlation (either temporal or spectral) across different auditory filters (Kidd et al., 1989; Plack 1998[17]). In narrowband noise, changes in envelope fluctuations upon addition of the tone were an important cue for level discrimination. In wideband noise, changes in spectral shape upon addition of the signal to the pedestal resulted in systematic changes of the temporal response pattern across the population of AN fibers. Model CD cells responded to the temporal pattern of AN responses and could be used as the underlying mechanism to detect different spectral shapes in wideband noise. Compared to tone-in-noise detection, there was a small but consistent elevation of the level-discrimination threshold for the wider bandwidths. It is possible that when the tone at the signal frequency was always perceived, the decision process was more affected by the perception of the loudness/level of the signal. The effect of the roving-level condition in narrowband noise could also be due to the same perturbation. Thus, it may not be a question of whether multiple cues are used in the psychophysical experiments, but rather a question of whether there is a common process[18] to combine these cues. Otherwise the subject must switch to different processes for different conditions, or even for different trials based on the molecular structure of the waveform. Since different levels of internal noise were required for the different models to prediction human performance, it is possible that these cues were evaluated by different processes.

The strategy of how these different cues are combined is an interesting topic for future study. Richards and Nekrich (1993) concluded that human subjects combine these cues non-linearly[19]. Performance could either be worse

[17] Most studies related to this phenomena have been done in the context of the reduced mid-level hump in the presence of noise (e.g. Plack and Viemeister, 1992).

[18] This does not include the situation that after each cue is evaluated, these cues are used altogether in some decision strategy. In this situation, different processes are still involved to detect different cues.

[19] The nonlinear combination could be algorithmic. For example, a "sequential" strategy could first attempt to detect a tone based on CD model cell, and if it fails as is "not sure", the decision process could continue to look for other cues (energetic) to determine which interval has a tone. By examining the performance of other models where the CD model fails, it should be possible to clarify which strategy is used in the decision process.

or better than the linear combination rule for different cues. Molecular psychophysics (e.g. Gilkey and Robinson, 1986) may provide a better situation to explore the detailed strategy involved in the decision process. Most studies (Gilkey and Robinson, 1986; Berg, 1992; Southworth and Berg, 1995; Richards and Buss, 1996) use simple models based on channel theory and focus on the different patterns of decision weights. Southworth and Berg (1995) suggested that different weights correspond to differences in the cue used to make the discriminations (i.e., pitch, loundness, roughness). It is feasible that the output of different models sensitive to different cues were combined instead of a simple channel model with different weights being involved in the decision strategy.

Neural mechanisms for level discrimination in wideband noise (lateral inhibition and cross-correlation)

In wideband noise, the underlying perceptual processing mechanism that explained human performance was to detect spectral contrast rather than to detect a difference in overall sound level or the level in a single critical band. These two processes could be distinctly different (Green et al., 1983). The spectral shape discrimination stores a classification of each spectrum in memory and compares the two stored spectra. Intensity discrimination in an across-trial roving-level paradigm is based on a short-term comparison, and performance is degraded if the inter-stimulus duration is increased (see review by Braida and Durlach, 1995). So spectral shape discrimination in wideband noise could be explained based on channel theory if it is assumed that the internal noise is different from that required for intensity discrimination in narrowband noise (thus we can predict the enhanced performance in wideband noise).

Lateral inhibition (Shamma, 1985; Rhode and Greenberg, 1994) has been discussed as a possible underlying neural mechanism for discrimination of spectral shape. The lateral inhibition model output is the weighted sum of the different input channels, where the weight of each channel is derived based on channel theory, as described above. The negative value of weight indicates an inhibitory input from the corresponding channel. The lateral inhibition scheme does not address questions concerning the duration over which the spectrum should be integrated and then compared. Several studies show that short-term spectral shape discrimination provides a possible cue to account for human performance (Hall and Grose, 1988; Richards et al., 1997), and human subjects can make quite fine (< 2ms) across-channel timing discriminations between broadband stimuli (Patterson and Green, 1970). In some sense spectral shape discrimination over a very short time duration has to rely on a coincidence-detection mechanism in a neural implementation; because the time constant must be short to avoid temporal summation, the temporal correlation across channels will affect the model response.

Deng and Geisler (1987) discussed the difference between the cross-correlation and lateral-inhibition mechanisms. Lateral inhibition is essentially a differential operation across different channels, whereas cross-correlation

involves a multiplication across channels. In the low-frequency region, AN fiber responses are halfway rectified, so multiplication is more sensitive to temporal changes across the channels. In the high frequency region, AN responses usually follow the envelope of the stimulus and have lower synchronization indices (there is a strong DC component in responses that are phaselocked to the envelope, which reduces the synchronization coefficient), and the differential operation is more effective to eliminate the DC component (mean rate) in the responses. It is possible that the nature of the temporal information in the AN fiber responses across frequencies leads to the use of different neural mechanisms at different frequencies.

5.6 Summary

The results from the present study suggested that there was more temporal information available for level discrimination in noise as the noise bandwidth increased. The enhancement of level discrimination in noise for wider bandwidths could be explained based on models detecting these temporal cues. In narrowband noise, the model based on an envelope detector could be used to account for human performance in roving-level conditions. In wideband noise, both the energy-based population model and the cross-frequency coincidence-detection model could be used to explain the constant threshold-bandwidth functions and robust performance in roving-level conditions. The coincidence-detection model responded to temporal changes across different AN fibers and did not require a mechanism to avoid the limited dynamic response range of the AN fibers. To predict human performance more accurately, different internal noises were introduced for different bandwidths to limit the coincidence-detection model's sensitivity to level changes and to spectro-temporal changes of the stimulus. The differences in internal noise that were required to adjust the model's sensitivity to level changes may reflect the fact there are more neurons involved in spectro-temporal processing than in coding the spectrum level of each critical band. The required internal noise decreased as noise bandwidth increased, and this decrease could be achieved by increasing the number of cells involved in processing (see Eq. 5.1). In summary, both temporal changes in the population of AN fiber responses and temporally related processing mechanisms should be considered in developing physiologically realistic models for experiments in psychophysics related to complex sounds.

Chapter 6 Summary and Comments

This dissertation was motivated by the fact it is still not well known how neurons at high levels of the auditory pathway respond to the spatio-temporal patterns in population responses of auditory-nerve (AN) fibers, despite the spectacular ability of AN fibers to encode the fine structure, or temporal details, of sounds over a wide range of low frequencies. Without a better understanding of the temporal response properties and the processing mechanisms of cells that are sensitive to changes in the relative timing across AN fibers, improved physiological descriptions of the temporal response properties of AN fibers are sometimes thought to be a distraction from the basic structure of information processing. These physiological details are often ignored in efforts to relate neural mechanism to psychophysical performance, without considering their potential fundamental importance. Thus, the primary goal of this dissertation was to investigate response properties of model coincidence-detecting cells that receive convergent inputs from AN fibers. The study also focused on the capability of these model cells in coding of complex sounds, for which temporal cues are believed to be important for related psychophysical tasks. Quantitative predictions of model cell response properties and psychophysical performance in this dissertation were made based on computational models for AN fibers and coincidence-detecting cells.

6.1 Response properties of an I&F model that receives sub-threshold inputs

Chapter 2 describes a computational method for calculating the output statistics (*PST* and *ISI* histograms) of an integrate-and-fire (I&F) model with arbitrary input waveforms. Several studies (Stein, 1965; Molnar and Pfeiffer, 1968; Kempter *et al.*, 1998; Burkitt and Clark, 2001) have provided analytical descriptions of the response statistics of the I&F model, but these studies were tied to specific assumptions that either the inputs were stationary or that there were large number of weak inputs. These assumptions are not appropriate for the auditory neurons in brain stem nuclei because most of them (e.g. AVCN bushy cells) receive a limited number of inputs and respond to the temporal structure of acoustic stimuli. The methods developed in Chapter 2 provide a new way to calculate the statistics of the neural model with more accuracy and efficiency than Monte Carlo simulations, and without the compromise of using only stationary inputs or limiting the model parameter space.

This method was used to analyze the response properties of an I&F model with different model parameters, in particular with different synapse configurations. These response properties, including the regularity measure, synchronization index, and output-input rate function, are thought to be important for the estimation of physiological and anatomical parameters of different neurons in the CN. The results provided insights into the relationship between these model parameters and response properties. For example, the

regularity of the cell responses were mainly determined by the EPSP time constant, and both the linear output-input rate function and enhanced phase-locking were observed for model cells with mixed-amplitude inputs.

6.2 Response properties of model coincidence-detecting cells that receive convergent AN fiber inputs

Recently, several studies (Deng and Geisler, 1987; Carney, 1994; Shamma and Klein, 2000; Carney et al., 2002) have proposed monaural cross-channel coincidence detection as a physiologically realistic mechanism to decode the spatio-temporal discharge patterns in AN responses for several specific psychophysical tasks. The distinctive response properties of a coincidence-detecting model have not been studied systematically, and thus it is difficult to identify the cells that behave as coincidence detectors in physiological studies.

Chapter 3 provided a survey of the general response properties of these model coincidence-detecting cells. The model results showed that the coincidence-detecting cell could have "normal" tuning (i.e. a sharp tuning curve with one tip) and a "normal" response area in response to tones at low levels (i.e. single-peaked rate vs. frequency functions), but have distinctive properties in response to the complex sounds. This observation had several important implications: 1) coincidence-detection mechanisms were more important in processing complex sounds than simple stimuli; 2) responses to simple stimuli do not provide good criteria to identify coincidence-detecting cells; 3) to explain the physiological responses of cells to complex sounds, the role of the coincidence-detection mechanism should be considered if a cell receives convergent inputs and has a limited integration window.

Chapter 3 also compared the response sensitivities of the shot-noise and cross-correlation models for detection of a tone added to wideband noise. The cross-correlation model could be described as an explicit physiological coincidence-detecting cell that received two AN inputs with a very short coincidence-detecting window, and the simple cross-correlation based CD cell had very low response rates. The shot-noise model included more AN inputs and had higher, and thus more realistic response rates, which were comparable to physiological observations, but its coincidence response could be masked by temporal integration from single-channel inputs. The results in Chapter 3 showed that with proper choices of model parameters, realistic response rates and temporal coding capabilities based on the coincidence-detection mechanism could be achieved by a single model cell. With few free parameters, the cross-correlation model is an efficient way of representing the coincidence-detection mechanism, and it is capable of processing information from input spatio-temporal response patterns that are of fundamental importance for the psychophysical tasks considered.

Chapter 3 also highlights the conclusion that the spectral shape of sound stimuli could be represented by the spatio-temporal pattern of AN responses

and decoded by the cross-channel coincidence-detection mechanism. A common mistake in developing auditory theory is to assume that each auditory filter represents the spectro-temporal information in its own narrow frequency region, and that the interactions between channels are relatively slow, and require delay lines or auto-correlation operations to extract the fine temporal structure of responses from each single channel. This is not true, as Chapter 3 illustrated, because auditory filters typically overlap each other, and changes in the spectrum shape of sound stimuli cause coherent temporal changes across several filters, even when such changes in temporal information were not obvious in single auditory filters.

The model responses to complex sounds predicted in Chapter 3 have been observed in physiological studies of AVCN cells in gerbil. In addition, other studies (Carney, 1990) have reported several neural response types in AVCN that have properties that are consistent with the monaural cross-frequency coincidence-detection mechanism. These observations suggest that cells in AVCN act like coincidence detectors to extract spatio-temporal cues. It is also worth pointing out that the sharp tuning and phase-locking to the fine temporal structure of AN responses are maintained or even enhanced along the auditory pathway up to the inferior colliculus (IC). Thus, the cross-channel coincidence matrix could happen anywhere along the auditory pathway at or prior to the IC (e.g. in the superior olive or the lemniscal nuclei). The simulations in Chapter 3 provide insights into the response patterns of coincidence detectors that should help the search for physiological evidence for the presence of cross-channel coincidence detection.

6.3 Analysis of psychophysical experiments with a coincidence-detection model

The goal of Chapters 4 and 5 was to examine the role of the coincidence-detection mechanism in the processing of spatio-temporal patterns of population AN fiber responses in experiments commonly associated with spectral masking. Two masking experiments, tone in noise detection and level discrimination in noise (Kidd et al., 1989), were considered. Experiments on tone-in-noise detection have been a foundation for the development of modern auditory theories. Experiments of level discrimination in noise are similar in paradigm to profile analysis, which has challenged the power spectrum hypothesis used to explain tone-in-noise detection experiments.

Chapter 4 presented a detailed analysis of response properties of the model coincidence-detecting cell with different stimulus conditions relevant to the masking experiments. The results supported the conclusion from Chapter 3 that the coincidence-detection mechanism was a robust mechanism for processing spatio-temporal cues in complex sound stimuli. In wideband noise, the most sensitive cell received inputs that had out-of-phase responses to the signal (the phase-opponency mechanism, see Carney et al., 2002). This was because the temporal changes of two adjacent input filters were easily detectable when their responses to the signal were out-of-phase with each

other. The input AN fibers could not be tuned too close to each other because their responses were already temporally correlated with the noise, and the fibers could not be tuned too far away from each other because that would require a higher signal-to-noise ratio to have the signal dominate the temporal responses of both auditory filters. The location of the most sensitive cell was also frequency specific, because correlation between two AN fiber responses depends on the signal frequency. The sensitivity of the coincidence-detecting cells in wideband noise was determined by the signal-to-noise ratio of the inputs and was not affected by the total level of the stimulus. In narrowband noise, the coincidence-detecting cell still responded to the spatio-temporal change of the inputs, but such changes were less dominant in the AN responses compared to the changes caused by the stimulus level. The coincidence-detecting cell that was sensitive in the narrowband-stimulus case had input AN fibers tuned further away from the signal frequency. In a neural implementation of a matrix, or population, of coincidence-detecting cells, such a cell could be replaced by cells that received low-spontaneous-rate AN inputs (which had high thresholds and wider dynamic response ranges); such a coincidence matrix could be constructed based on cells that received inputs that were more closely tuned.

Predictions from Chapters 4 and 5 based on a coincidence-detection matrix shared many similarities: 1) In wideband noise, the same model cell was dominant for both roving-level and fixed-level conditions, and predictions were robust to the roving-level conditions and were comparable to the human performance. These results suggest that the same process may be involved in both of these experiments. 2) In narrowband noise, the coincidence-detection model was affected by stimulus level in both experiments. Different cells were used in fixed-level and roving-level conditions, suggesting that different cues may be involved in the narrowband noise masking experiments. There were also some differences between tone-in-noise detection and level discrimination in noise: 1) in wideband noise, the most sensitive cell for level-discrimination still received out-of-phase inputs, but the inputs were tuned further away from the signal. This phenomenon was similar to the “spread of excitation” in the energy-based model, but it happened in the temporal domain and was mainly determined by the signal-to-noise ratio instead of limited neural response dynamic ranges. Also, auditory filters are broader in response to the high-level tone due to the compressive nonlinearity, and thus the 180-degree phase difference will occur for CFs that are further apart from each other; 2) In narrowband noise, the AN temporal responses were already dominated by the signal tone for level discrimination because of the presence of the pedestal tone, and there were smaller *changes* in the fine-temporal information associated with changes in the high-level tone than for the detection case. It was possible that the envelope fluctuations were more detectable in the narrowband level-discrimination in noise task because the envelope statistics were less variable in the presence of the high-level tone.

The overall performance of the coincidence model together with several other models discussed in Chapters 4 and 5 are summarized in Table 6.1.

Coincidence detection is a useful mechanism for robust encoding of spatio-temporal patterns of AN responses, especially in wideband noise for both fixed-level and roving-level conditions. The coincidence-detecting cell that is sensitive to the level and temporal

		Cross-Frequency Coincidence Detector	Multiple-channel based Energy Detector	Envelope Detector
Physiological Mechanism		Coincidence Detection (Small membrane time constant, Sub-threshold, convergent inputs)	Lateral inhibition (Large membrane time constant, Non-saturated responses)	Modulation filter (Rate-based modulation transfer function)
Tone-in-noise Detection Task	NB	Fixed-level Roving-level	Fixed-level	Roving-level
	WB	Fixed-level Roving-level	Fixed-level Roving-level	
Level-Discrimination Task	NB	Fixed-level (partially) Roving-level (partially)	Fixed-level (partially)	Roving-level
	WB	Fixed-level Roving-level	Fixed-level Roving-level	

Table 6.1: Summary of the different models in psychophysical experiments. The second row shows the possible underlying neural mechanism for each models, and the other rows list the different stimulus conditions for which the model successfully explained psychophysical results. The conditions for which the model is useful but cannot explain the human performance alone are indicated by the word "partially" in the table.

information of the stimulus can also be used to account for human performance in narrowband noise. Most cells in the auditory CNS (below the IC) receive sub-threshold convergent inputs and have short membrane time constants, making coincidence detection a realistic neural implementation of a wide variety of psychophysical auditory functions.

The energy-based multiple channel model predicted human performance in wideband noise for both fixed-level and roving-level conditions but failed to predict human performance in narrowband noise in the roving-level condition. A lateral inhibition network (LIN, Shamma, 1985) is a candidate for neural implementation of the energy-based channel model. The LIN is similar to cross-channel correlation if the integration time of the model cell is short. Slow across-channel interactions based on LIN can only extract information from AN fibers that change their rate responses when the stimulus level changes; this mechanism discards the spatio-temporal information from AN fibers that have saturated responses.

The results from an envelope-detector model illustrated that the envelope fluctuations in narrowband noise were an important cue for both detection of tones in noise and level discrimination of tones in noise. There were more envelope cues available when the bandwidth of the noise increased, consistent with the enhancement of human performance in experiments of level discrimination in noise. The performance of an envelope detector is not affected by the roving-level paradigm in narrowband noise and is worse than human performance in fixed-level experiments, suggesting that the envelope cue may not be used alone in fixed-level experiments. It is also reported (Wakefield and Viemeister, 1990) that human performance for modulation-depth discrimination is much worse than predictions from a model envelope detector at high modulation depths, suggesting that envelope cues may not be a reliable cue in detection experiments for which the envelope fluctuations are large.

6.4 Limitations and Directions for Future Studies

The central issue in the investigation of nervous system function is to understand how information is encoded and processed by neurons. The fact that we often have to ignore physiological details, even though they are potentially important, reveals a lack of insight into the dynamics of the underlying system. While the present study used computational nonlinear AN models that are based on physiological data from various species, it is difficult to test whether this model describes human AN physiological responses in a quantitatively accurate manner. Also, the present study did not include AN fibers with low spontaneous rates; coincidence-detecting cells that receive these AN inputs may have different response properties from those that receive high-spontaneous-rate fibers. The computational nonlinear AN model used in the present study could simulate many effects of cochlear nonlinearity, but the detailed effects of cochlear nonlinearity on coincidence-detecting cell responses and model performance have not been addressed.

The physiological responses of neurons involved in coincidence detection in binaural processing have been well characterized (see Yin and Chan, 1988; Irvine, 1991); however, there is still much to be clarified about the response patterns of monaural cross-frequency coincidence-detecting cells. Coincidence detection may be just one aspect of cell responses together with other response dynamics (e.g. on-frequency inhibition), thus the performance of coincidence detection should be evaluated with a model that incorporates other processing mechanisms. It is also possible that coincidence detection happens at several stages along the auditory pathway. A more detailed understanding of the neural circuits will improve the accuracy of computational modeling and thus help gain insights about neural functions.

Coincidence detection (or cross-correlation) is a fundamental operation that underlies almost all models related to binaural processing as well as many models in monaural processing (e.g. models in pitch perception). Essentially, this operation transforms the spatio-temporal pattern of AN responses into another domain (space). Many questions remain for future studies: which features of sounds are more clearly represented in this new domain? What psychophysical tasks can benefit most from such transformations? How is such a transformation affected by hearing impairment, and what information we can get from studying hearing impairment in this context? Generally, for psychophysical tasks that potentially benefit from coincidence detection, this operation should result in response patterns that are easier to identify and are more tightly clustered with respect to the sound information to be extracted.

The present study provides potential suggestions for future physiological experiments, and new insights into the physiological response properties of neurons. Specifically, cross-frequency coincidence cells should have complex tuning (e.g. multiple peaks) at high stimulus levels, and these cells must exhibit predictable response properties to temporal changes across AN fibers. The changes of temporal correlation across AN fibers could be manipulated by adding a tone to the noise stimulus, and coincidence-detecting cells should decrease their response rate when a tone near CF is added to noise. There are also several possible ways to distinguish the contributions of coincidence detection and lateral inhibition to decrements in response rate: the coincidence detection cell should generate a decrease in response rate when a tone is added at CF, and such a decrement should be observed at noise levels for which the cell response is saturated. In contrast, a LIN cell should have a decrease in response rate when a tone is added away from CF, and inhibition should be less obvious at noise levels for which the cell response is saturated.

As advances in physiology continue to provide a more detailed description of the underlying neural system, the use of more complex and computational models that incorporate realistic neural circuits to predict psychophysical experiments is inevitable. Molecular psychophysical (Gilkey and Robison, 1986) techniques offer a valuable approach for the demanding tests of these complex models. Molecular psychophysics studies the relationship between human performance on individual waveform and the microstructure of that particular waveform. Molecular psychophysical measures consider the stimulus-response

relationship with little regard to “correctness” and provide descriptions and insights about listening behavior in the context of the microstructure of the stimulus. It is often the case that models providing good predictions of average threshold data fail to explain the microstructure of the data. Regarding tone-in-noise detection, for instance, the violation of the “power spectrum model”, which successfully describes several masking phenomena, has been confirmed in a number of molecular based studies (Gilkey and Robinson, 1986; Richards and Buss, 1996). It is possible that individuals attend to very different aspects of the available acoustic information (i.e., pitch, loudness, roughness), yet show similar thresholds. Also, molecular techniques should not only pay attention to the different model parameters for each microstructural aspect of the data, but also to possible strategies that combine different models (or models with different parameters) in a single task.

Appendix A. Analysis of models for the synapse between the inner hair cell and the auditory nerve

A.1 Introduction

The inner hair cell (IHC) and auditory nerve (AN) complex is a critical element in the peripheral auditory system that converts a mechanical signal (the response of the cochlea to sound) into a neural signal. Physiological studies (Westerman and Smith, 1984; Westerman 1985; Rhode and Smith, 1985) have provided insight into the temporal dynamics of IHC-AN synaptic processing. The response of an AN fiber to a constant-intensity tone burst is typified by very rapid firing at the onset that declines with time, rapidly at first, then more slowly over a period of tens of milliseconds; the AN response at the offset of the stimulus is greatly reduced and slowly recovers over several tens of milliseconds (Harris and Dallos, 1979; Westerman, 1985). The likely source of these effects can be inferred from other physiological studies. Since the amplitude of the receptor potential of the inner hair cell (IHC) produced by a constant amplitude stimulus (for example, a high-frequency tone) is essentially constant (Russell and Sellick, 1978), adaptation is believed to occur at the level of the transmitter release process in the IHC-AN synapse. The adaptation process in the IHC-AN synapse results in greater sensitivity to transient stimuli than to steady-state stimuli and underlies other temporal aspects of AN responses. These properties also limit some aspects of temporal coding of stimulus envelopes. For example, offset adaptation may be responsible in part for the psychophysical phenomenon of forward masking (Harris and Dallos, 1979).

IHC-AN synaptic adaptation is very complex. Its characteristics depend on stimulus intensity, duration, and previous stimulation history. Synaptic adaptation at the onset of AN responses to tone bursts is usually characterized functionally by two exponential components (Westerman, 1985; Westerman and Smith, 1984; Rhode and Smith, 1985); adaptation at the offset of the stimulus can also be described as an exponential recovery component, with a different time constant than the onset, after a deadtime period (Fig. 1; Harris and Dallos, 1979). Other measurements that have been made to help characterize AN adaptation processes include responses to increments or decrements in stimulus intensity with an ongoing tonal background (Smith and Zwislocki, 1975; Smith et al., 1985). Adaptation characteristics also differ for fiber types with different spontaneous rates (Rhode and Smith, 1985). The diversity and complexity of IHC-AN synaptic adaptation provide a challenge for successful modeling of synapse dynamics (Hewitt and Meddis, 1991).

The mechanisms that give rise to synaptic adaptation have often been modeled using multiple reservoirs of neurotransmitter, with diffusion out of the cell and between reservoirs within the cell (Furukawa and Matsuura, 1978;

Schwid and Geisler, 1982; Meddis, 1986; Westerman and Smith, 1988). Each diffusion step is controlled by a permeability parameter, and at least one of the permeabilities in these models is determined by the stimulus (presumed to be controlled by IHC calcium concentration, intracellular IHC voltage, or equivalently stimulus

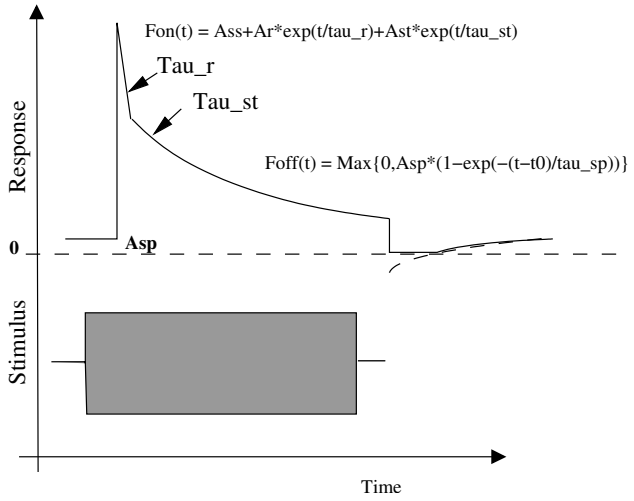


Figure 1: Schematic diagram of the response of an AN fiber to a tone-burst (Harris and Dallos, 1979; Westerman, 1985). The onset response of the AN fiber can be described as a sum of two exponential components (see Eq. A19), and the offset response can be fit with a single exponential recovery function after a deadtime (Eq. A22).

intensity). Adaptation of the synaptic output depends on the reduction of the driving force for the diffusion of the synaptic material (transmitter) from the cell into the synaptic cleft (Fig. 2). Mathematically, low pass filters are used to implement the replenishment and diffusion mechanisms between different transmitter reservoirs. These models can be implemented using either a cascade of low-pass filters or parallel low-pass filters (depending on the interconnection of the reservoirs).

The variation of adaptation characteristics across AN fibers makes it desirable to have different sets of model parameters to predict individual AN fiber responses more accurately, but determining these parameters is always a tedious job. Depending on the model structure, simulation results may or may not give insight into the model; that is, quantitative relationships between the model parameters and adaptation properties may not be revealed by the exercise of fitting the parameters to a single fiber response. For the model proposed by Westerman and Smith (1988), equations were derived to determine the model parameters from the desired shape of the onset PST histogram (Westerman and Smith, 1988; Zhang et al., 2001), but a detailed exploration of the relationship between model parameters and other adaptation properties is still needed. The studies of model adaptation characteristics are limited by the sets of parameters provided by the authors (e.g. Westerman and Smith, 1988; Hewitt and Meddis, 1991; Sumner et al., 2003), which makes it difficult to understand the relationship between the parameters and different adaptation properties. A systematic study of the adaptation characteristics of these models has never been completed, perhaps due to the lack of insight concerning the parameters of these models.

In this article, we will use a unified approach to consider two models, that of Meddis (1986) and of Westerman and Smith (1988; we will refer to this as the Westerman model). In our approach, both models can be described as circuits of interconnected low-pass filters. In this framework, the responses of each model to a constant-intensity stimulus can be determined analytically, and relationships between the response characteristics and model parameters can be established. This mathematical method makes the difference and similarities between the two models more clear, leading us to the conclusion that, despite their different structures, the two models are essentially the same. In addition, because the model equations provide better insight into the model structure, we can easily derive the model parameters based on the desired adaptation properties.

We propose a modified version of Westerman's model that has the same onset adaptation but improved offset adaptation. The modified offset adaptation enhances the modulation gain of the model AN fiber responses to modulated stimuli, consistent with AN physiology (Joris and Yin, 1992).

Finally, we consider an efficient way to include refractoriness in a phenomenological AN model. The refractoriness observed in AN fiber responses is generally associated with the mechanism that generates the action potential. A method developed by Gaumond et al. (1982, 1983) can be used to remove the effect of refractoriness from measured PST histograms and thus estimate the

underlying synapse output (transmitter release rate or instantaneous firing rate) from AN fiber recordings (Westerman and Smith, 1988). When simulating AN responses, the model synapse output can be used to generate a model PST by including the characteristics of the refractoriness (Edwards and Wakefield, 1990). A computationally efficient, modified version of the Edwards and Wakefield PST approximation procedure is introduced in this study.

A.2 Analysis of the Meddis synapse model

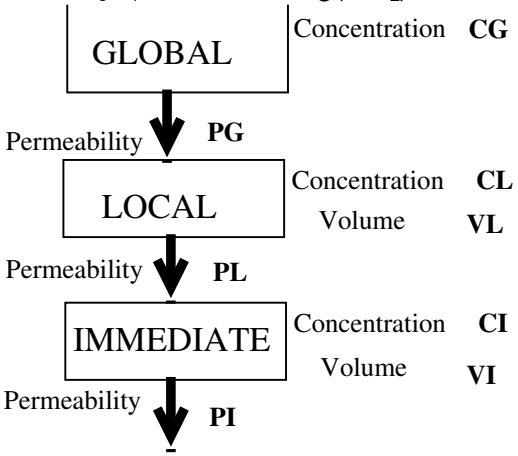
The model proposed by Meddis (1986, 1988) has three neurotransmitter reservoirs that are arranged in a cycle: the immediate store (q); the synaptic cleft (c); and the reprocessing store (ω) (see Fig. 2). The transmitter replenishment and release in the model can be described by the following equations:

$$\frac{dq}{dt} = y(M - q(t)) + xw(t) - k(t)q(t)\,, \tag{A1}$$

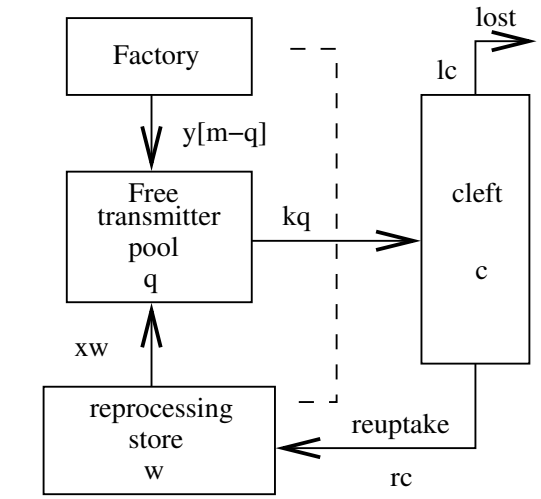
$$\frac{dc}{dt} = k(t)q(t) - (l + r)c(t)\,, \tag{A2}$$

$$\frac{dw}{dt} = rc(t) - xw(t)\,; \tag{A3}$$

The first equation describes the dynamics of the amount of transmitter in the immediate store: new transmitter in the immediate store is either manufactured from the factory (at a rate of $y(M-q)$, where M is the



(a) Westerman's Model: $S_{out}(t) = P_I(t)\, C_I(t)$



(b) Meddis' Model: Sout(t)=k(t) q(t)

Figure 2: Schematic diagram of two synapse models (Meddis, 1988, and Westerman and Smith, 1988). (a) In the Westerman model (see Eqs. A14 and 15 below), three transmitter reservoirs are cascaded, each with their own concentration and volume. The transmitter diffuses from global to local and then from local to immediate reservoirs. The output of the synapse model, S_{out}, is determined by the product of the concentration in the immediate reservoir and diffusion permeability from the immediate reservoir into the synaptic cleft. (b) In the Meddis model (Eqs. A1-3), the transmitter in the cleft is either lost or reprocessed by the reprocessing store (w). The free transmitter pool receives transmitter either from the reprocessing store or from a global factory and releases the transmitter into the cleft (with instantaneous release probability density *k*). The model synapse output is a value proportional to the amount of transmitter left in cleft (c) or the rate of transmitter released into the cleft (Meddis, 1988; Sumner et al, 2003).

concentration of neurotransmitter in the global store), provided by a reprocessing store (at a rate of $x\omega$, see Eq. A3), or released from the immediate store into the synaptic cleft at a rate of kq. Transmitter in the cleft is either lost at a rate of lc or recycled at a rate of rc (Eq. A2). The instantaneous release probability of a vesicle from the immediate store, $k(t)$, is usually a function of intracellular IHC voltage (defined as V_{ihc}, referenced to the IHC resting potential) which is determined by the input sound stimulus. For a high-frequency tone burst, the IHC receptor potential V_{ihc} is dominated by the "DC" component (Cheatham and Dallos, 1993; Russell and Sellick, 1978) and it is reasonable to assume that $k(t)$ is a constant after the onset (denoted as $k2$; the value of $k(t)$ before the onset is denoted as $k1$). For further analysis, Eqs. A1-3 can be transformed into the Laplace (complex frequency) domain as follows for t > 0:

$$sQ(s) - q(0-) = yM/s - yQ(s) + xW(s) - k2Q(s), \tag{A4}$$

$$sC(s) - c(0-) = k2Q(s) - (l+r)C(s), \tag{A5}$$

$$sW(s) - w(0-) = rC(s) - xW(s), \tag{A6}$$

where $q(0\text{-})$, $c(0\text{-})$ and $\omega(0\text{-})$ are the reservoir concentrations before the onset.

After solving for $Q(s)$,

$$Q(s) = \frac{(sq(0-) + ym)(s+x)(s+l+r) + c(0-)rxs + w(0-)xs(s+l+r)}{s(s+x)(s+y+k2)(s+l+r) - k2rxs}, \tag{A7}$$

the characteristic function of $q(t)$ can be represented by

$$q(t) = \Phi_0 + \Phi_1 e^{-t/t_1} + \Phi_2 e^{-t/t_2} + \Phi_3 e^{-t/t_3}, \tag{A8}$$

where $-1/t_i$ are pole of $Q(s)$. The values of t_1, t_2, t_3 and Φ_0, Φ_1, Φ_2 can be calculated from Q(s) directly.

Based on the parameters given in Meddis (1986, 1988), we can assume that l and $r \to \infty$ (or $l,r >> x,y,k$), and we will let $u = r/(l+r)$. Thus, the following equations can be used to describe the above results:

$$\frac{dq}{dt} = y(M - q(t)) - kq(t) + xw(t), \tag{A9}$$

$$\frac{dw}{dt} = kuq(t) - xw(t), \tag{A10}$$

$$C(s) = \frac{k(0+)Q(s) + c(0-)}{s + l + r}, \tag{A11}$$

$$Q(s) = \frac{(sq(0-) + ym)(s + x) + w(0-)xs}{s(s + x)(s + y + k2) - k(0+)uxs}, \tag{A12}$$

$$q(t) = \Phi'_0 + \Phi'_1 e^{-t/t'_2} + \Phi'_2 e^{-t/t'_2}. \tag{A13}$$

Figure 3a shows the comparison of the analytical solution of k(t)q(t) for both the original model (Eq. A8) and the simplified model (Eq. A13) using the parameters given in Meddis (1988). In Fig. 3b, simulation results from both models with the same input k(t) (a sinusoid waveform at 1kHz) are compared. Both plots show that solutions from the simplified equations and the original equations are essentially identical.

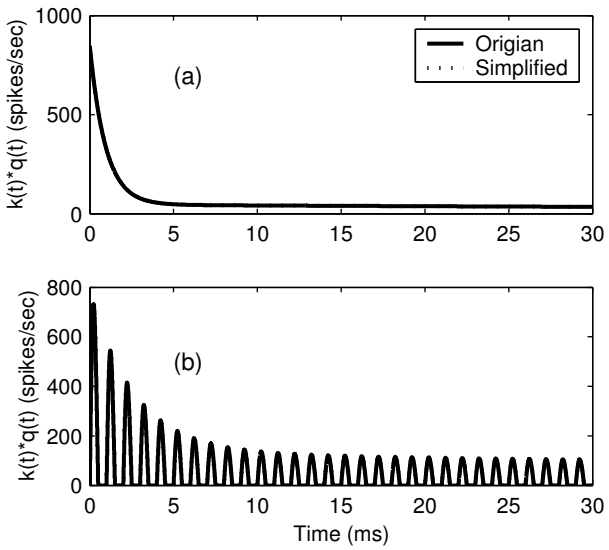


Figure 3: Comparison of synapse output for the original Meddis model (solid line) and the simplified model (dotted line) described in the current study. Both models use the parameter values given in the Meddis study (1988): l = 2500, r = 6580, x = 66.31, M = 1, y = 5.05. The responses from two models are essentially identical, as illustrated by the overlapping curves. (a) The analytical solution of $k(t)q(t)$ for both models for a constant stimulus, with k_1 = 40.49 and k_2 = 1660. (b) Simulation results of $k(t)q(t)$ for both models with a time-varying input $k(t)$ (a half-wave rectified sinusoidal waveform at 1 kHz with amplitude of k_2: $k(t)=0.5k_2$ [$\sin(2000\pi t)+|\sin(2000\pi t)|$]).

These results also hold for other parameter sets given in various studies (Meddis, 1986; Meddis 1988; Sumner et al., 2003; not shown). Thus, the simplified equations provide a good description of the Meddis model, and the solution (Eq. A13) has two components with different time constants that are similar to the characteristics of onset adaptation in AN fibers.

The accuracy of the simplified equations describing the Meddis model across different values of the sum $l+r$ is illustrated in Fig. 4. Each individual component of $q(t)$ (t_i and Φ_i) are plotted as a function of $l+r$ for both models (Fig. 4). The values of these components are constant for the simplified model because u is kept the same for different values of $l+r$. The values of corresponding components in the original model reach the values in the simplified model as the value of $l+r$ increases, and the values are indistinguishable from each other when $(l+r) > 5000$ (the values of $l+r$ used in the Meddis (1986, 1988) were usually greater than 15000). The amplitude (Φ_3) of the additional component in the original model can be neglected compared to the other components when the value of $l+r$ is large.

A.3 Comparison of Meddis and Westerman models

Westerman and Smith (1988) proposed a model with three transmitter release reservoirs (global, local and immediate, see Fig. 2). The governing equations for the transmitter release between these reservoirs are

$$V_I \frac{dC_I(t)}{dt} = -P_L C_I(t) - P_I(t) C_I(t) + P_L C_L(t), \qquad \text{(A14)}$$

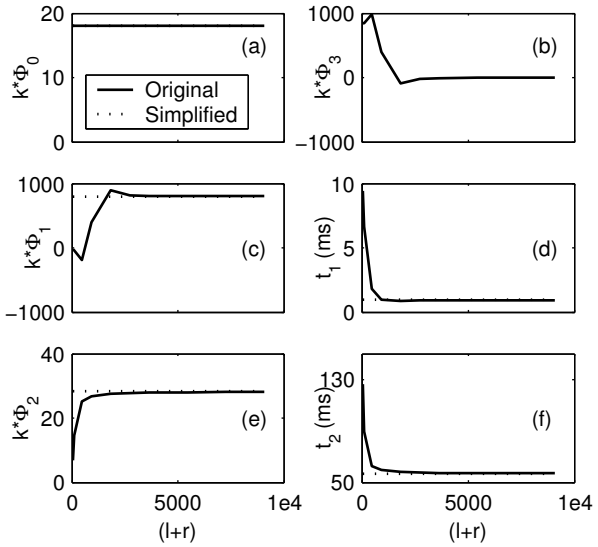


Figure 4: The difference of each individual component (t_i and Φ_i) of $q(t)$ between the Meddis (1986) model and the simplified model as a function of $l+r$. The values of these components are constant for the simplified model (dotted line) because $u=r/(l+r)$ is kept the same for different values of $l+r$. The amplitude of the component (Φ_i) is scaled by k so it can be interpreted as the instantaneous firing rate of the synapse output. The values of corresponding components in the original model (solid line) all converge quickly to the values in the simplified model as $l+r$ increases, and the values are indistinguishable from each other when $l+r > 5000$. The amplitude (Φ_3) of the additional component doesn't exist for the simplified model; in the original model, this amplitude converges to zero as $l+r$ increases and it can be neglected when $l+r$ is large.

and

$$\frac{dC_L(t)}{dt} = \frac{P_G C_G}{V_L} + \frac{P_L}{V_L} C_I(t) - (\frac{P_L + P_G}{V_L}) C_L(t) , \qquad \text{(A15)}$$

where C_I, C_L, and C_G are the immediate, local and global reservoir concentrations; V_I and V_L are the immediate and local reservoir volumes; and P_L, P_G are the permeabilities between the local and immediate and between global and local reservoirs. The release permeability $P_I(t)$, which is the same as $k(t)$ in the Meddis model, is a function of intracellular IHC voltage ($Vihc$), which is determined by the input sound stimulus. The model has the same characteristic function as the simplified Meddis model (Eq. A13):

$$C_I(t) = \Phi'_0 + \Phi'_1 e^{-t/t'_2} + \Phi'_2 e^{-t/t'_2} . \qquad \text{(A16)}$$

Because $k(t)$ and $P_I(t)$ are the only stimulus dependent variables in these two models, to compare the two models more directly we substitute $k(t)$ with $P_I^M(t)$ and $q(t)$ with $C_I^M(t)$ in Eqs. A9 and A10, and the constant yM in Eq. A9 can be removed by substituting $w(t)$ with $\frac{y+xu}{x} C_L^M(t) - yM/x - uC_I^M(t)$. We can then get the following equations describing the Meddis model:

$$\frac{dC_I^M}{dt} = -(y + xu) C_I^M(t) - P_I^M(t) C_I^M(t) + (y + xu) C_L^M(t) , \qquad \text{(A17)}$$

$$\frac{dC_L^M}{dt} = \frac{xyM}{y + xu} + xu(\frac{x}{y + xu} - 1) C_I^M(t) - x(1 - u) C_L^M(t) . \qquad \text{(A18)}$$

Since there is one free parameter in Westerman's model, we can always assume V_I=1 (that is, we can decrease V_I, V_L, P_I, P_L, P_G, and increase C_I, C_L, C_G by same scale V_I, and the synapse output $P_I C_I$ remains the same). It is obvious that the equations describing the two models (Eqs. A14-15 and Eqs. A17-18) are directly comparable with each other. If we set the corresponding parameters to the same values (e.g. PL=y+xu), then the synaptic response of the Meddis model (kq) and that of the Westerman model ($P_I C_I$) to arbitrary inputs should be exactly the same.

A.4 Deriving Model Parameter Values from PST properties -The relationship between the model parameters and adaptation characteristics

Extensive measurements of adaptation characteristics (Rhode and Smith, 1985; Westerman and Smith, 1984; Westerman, 1985) show that the onset adaptation of AN fiber can be described as a sum of two exponential functions:

$$R_{on}(t) = A_{ss} + A_r e^{-t/t_R} + A_{st} e^{-t/t_{ST}} , \qquad \text{(A19)}$$

where A_r and A_{st} are the components of rapid and short-term adaptation, t_r and t_{st} are the respective decay time constants, and A_{ss} is a steady-state component. Fitting the characteristic equation (Eq. A19) to the PST histogram will determine the values of these parameters; the refractoriness from the PST histograms can be removed before fitting (Westerman and Smith, 1988) so the IHC-AN synapse output (instantaneous firing rate) can be more accurately represented by the characteristic function.

The parameters in the Westerman model can be derived (Westerman and Smith, 1988; Zhang et al., 2001) from the adaptation parameters (A_{ss}, A_r, A_{st}, t_R, t_{ST}) and the spontaneous rate (A_{sp}) of the AN fibers. It is also possible to derive the values of the parameters in the Meddis model from these characteristic parameters (See Appendix B), since both models can be described with essentially the same set of differential equations (see above).

Though the same desired adaptation responses could be obtained from both models, the dependence of the different parameters in each model upon different adaptation characteristics is different. Also, because the models have different structures, these parameters have different physical meanings in each model. Figures 5 and 6 show the changes of the parameters in the two models for different model AN fiber types. The peak-to-sustained ratio ($PTS=(A_r+A_{st}+A_{ss})/A_{ss}$) of the onset adaptation changes with spontaneous rate (A_{sp}), as described by the following equation:

$$PTS = 1 + 9 \times A_{sp} / (9 + A_{sp}) . \qquad \text{(A20)}$$

The other adaptation parameters are held constant across model AN fibers (t_R = 2ms, t_{ST} = 60ms, A_r/A_{st}=6, A_{ss}=350). Figure 5a shows that the value of M in the Meddis model changes nonmonotonically with increasing spontaneous rate; it drops initially, and then increases from 3 to 10 for spontaneous rates increasing from 1 to 100 spikes/sec. All the

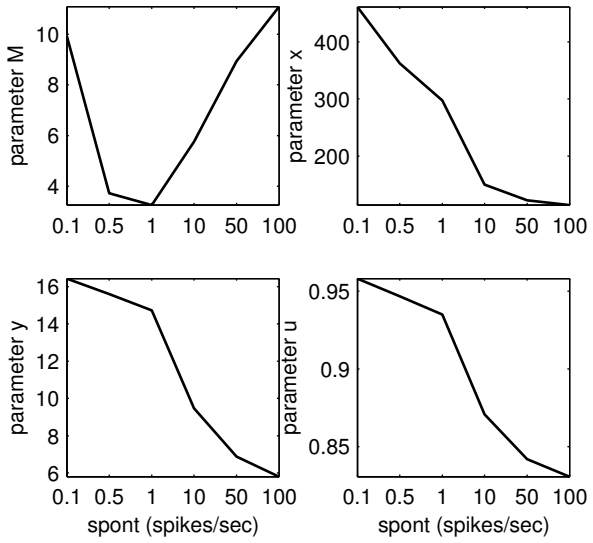


Figure 5: The relationship between parameter values (M, x, y and u) and spontaneous rate for the Meddis model. Other adaptation parameters used to derive the model parameters are: $PTS=(A_r+A_{st}+A_{ss})/A_{ss}=1+9A_{sp}/(9+A_{sp})$, $t_R=2$ ms, $t_{ST}=60$ ms, $A_r/A_{st}=6$, and $A_{ss}=350$. The value of M in the Meddis model changes nonmonotonically with increasing spontaneous rate; all other parameters decrease as spontaneous rate increases.

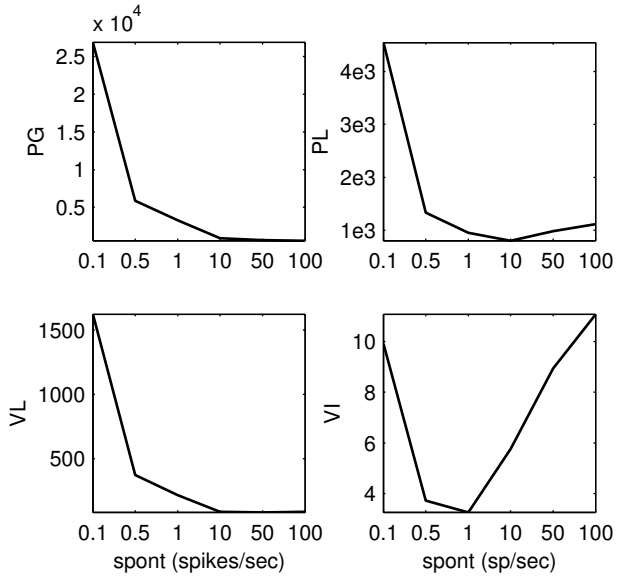


Figure 6: The relationship between parameters (P_L, P_G, V_L, V_I) and the spontaneous rate in the Westerman model. The other adaptation parameters are the same as in Fig. 5. The parameters were derived with global concentration (C_G) set equal to 1.

other parameters (x, y, u) in the Meddis model decrease continuously as the model's spontaneous rate increases. The changes of parameters in Westerman's model are plotted in Fig. 6. The global concentration (C_G) was always set to 1 before determining the other parameter values. The immediate volume, V_I (Fig. 6d), has similar trends as M observed in Fig. 5a. Other parameters (P_G, P_L, and V_L) decrease at first and then are insensitive to the spontaneous rate above 5 spikes/sec.

The A_{ss} defined in the adaptation parameters is the sustained rate at a certain stimulus level (the level is not specified here, but is presumably a high level which will be determine by the level-permeability function in the rest of the composite model for the AN response). When the permeability increases, the sustained rate in both models saturates at a value

$$A_{ss\_MAX} = \lim_{Pi \to \infty} P_I C_I = \frac{P_L P_G C_G}{P_L + P_G} = \lim_{k \to \infty} kq = \frac{yM}{1-u}. \qquad \text{(A21)}$$

The value of A_{ss_MAX} also changes with AN fiber spontaneous rate (see Fig. 7): the value for high-spontaneous model fibers is comparable to A_{ss}, whereas A_{ss_MAX} increases rapidly with decreasing of the spontaneous rate. The results above may suggest there is a fundamental difference between the low and high spontaneous rate AN fibers. When spontaneous rate is low, the onset adaptation is not sensitive to the values of the parameters t_R, t_{ST}, A_r/A_{st}, which reduces confidence in analysis based solely on these parameters.

It is also possible to study the effects of the model parameters on adaptation characteristics analytically with the solution given above (Eq. A12, for example, gives the analytical solution for the simplified Meddis model). Figure 8 shows the change of components of the model onset adaptation as a function of the value of immediate permeability P_I (or k in the Meddis model) for a model IHC-AN synapse with spontaneous rate of 50 sp/sec. Since P_I is level dependent, the figure also represents the effect of level on model onset characteristics. (A full composite model for the auditory periphery is needed to study the effects of level on onset adaptation characteristics, e.g. Sumner et al., 2003.) The model has a spontaneous response with permeability of P_{I1} when there is no stimulus input. The sustained rate (A_{ss}) is roughly saturated at permeability P_{I2} which generates the desired onset adaptation while the contribution of short-term and rapid components continues to change as the permeability P_I increases (Fig. 8a). Both the rapid decay time constant and short-term decay time constant decrease as P_I increases (Fig. 8b, note that t_R is multiplied by 10 in this plot). The short-term time constant is about 120 ms at rest and decreases to 50 ms at high levels; and the rapid time constant is 8 ms at rest and decreases continuously as the input sound level increases (the value we choose to derive the model parameters was t_{ST}=60 ms, t_R=2 ms). The prediction here does not fit the physiological data quantitatively, which may suggest that the model is a poor description of the IHC-AN-fiber (it is possible

that other model parameters are also stimulus dependent, see Westerman and Smith, 1988).

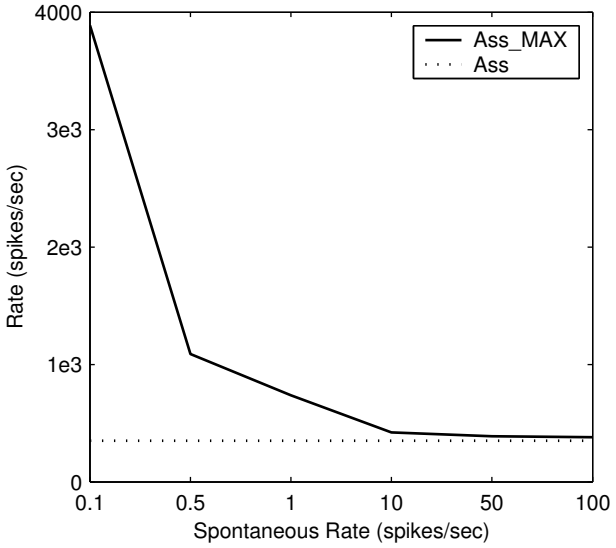


Figure 7: The maximum sustained rate (A_{ss_MAX}, solid line; see Eq. A21) (for both Meddis and Westerman models) as a function of model spontaneous rate. The other adaptation parameters are set as in Fig. 5 (thus A_{ss} has a constant value, as shown by the dotted line). The value of A_{ss_MAX} for high-spontaneous rate model fibers is comparable to A_{ss}, whereas the value increases rapidly as spontaneous rate decreases.

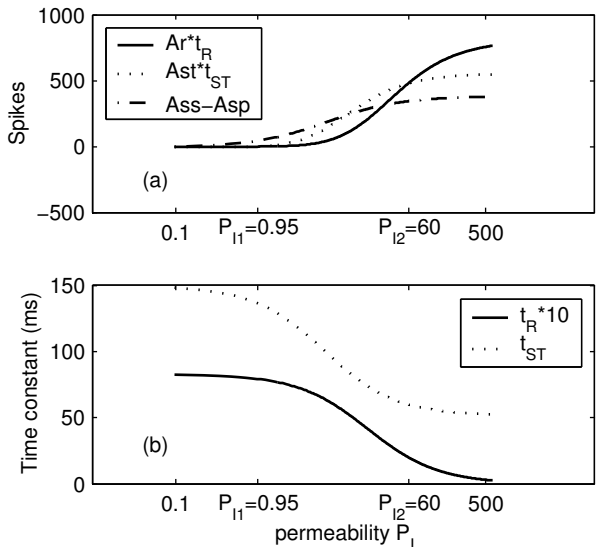


Figure 8: Effects of the model parameter, P_I, (for Westerman model) on onset adaptation characteristics. The model parameters were derived from adaptation properties specified in Fig. 5 with a spontaneous rate of 50 spikes/sec. The values of P_{I1} and P_{I2} are the values of the immediate permeability before and after the stimulus onset. (a) The contribution of each PST onset component. The contribution of the rapid component (solid line) and short-term component (dotted line) are the integrals of their exponential functions (and thus have units of spikes). The contribution of the constant component (sustained rate – spontaneous rate; dashed line) saturates as P_I increases. (b) Both rapid and short-term time constants decrease as the permeability increases. The rapid time constant (t_R, solid line) is multiplied by 10 to allow it to be plotted with the short-term time constant (t_{ST}, dotted line).

It is important to note that the relationship between V_{ihc} and P_I should change together with the fiber adaptation characteristics: different adaptation characteristics need different sets of the model parameters (including the permeability at rest and permeability at any other desired level); the spontaneous rate, rate threshold, and rate-level function curve are different for different fiber types that require different V_{ihc}-P_I relationships. An appropriate V_{ihc}-P_I representation is critical for a composite model that successfully describes AN responses and will be discussed in more detail below.

A study of the effect of adaptation characteristics on model parameters and of the model parameters on adaptation characteristics can be accomplished with this analytical method (we only show the effects of spontaneous rate and P_I). For example, it is also possible to derive the model parameters from adaptation characteristics for different stimulus levels. When more than one model parameter (P_L, P_G, V_L, etc) varies with level (Westerman and Smith, 1988), which invalidates the underlying assumption of the method used here to generate the model parameters, it is more difficult to build up the relationship between adaptation characteristics and model parameters. On the other hand, when the model parameters are derived to produce the desired adaptation at each level, the effects of the model parameters on adaptation characteristics can be studied in detail and more information about the synapse dynamics can be retrieved.

A.5 Modifying the synapse model to obtain the desired offset adaptation response

The above analysis shows that the model parameters can be determined based on the onset adaptation characteristics (at a certain stimulus level). When the model parameters are set, the other adaptation characteristics of the model synapse are automatically determined. If we set the level-dependent permeabilities P_{I1} and P_{I2} (k1 and k2 in the Meddis model) to the values before and after the offset, the analytical solution from Eq. A16 provides the offset responses of the model synapse. A direct conclusion is that the offset adaptation of the model synapse is also a sum of two exponential functions. The time constants of these two components are determined by P_I after offset and limited by the time constant at the onset (since only P_{I2} affects the time-constant, see Westerman and Smith, 1988, for Westerman's model, Eq. A12 for the Meddis model, Fig. 8b also shows how the time constant in the offset adaptation changes with the permeability after offset). While physiological studies (Smith, 1977; Smith and Zwislocki, 1975; Westerman, 1985) suggest that AN fibers with medium or high spontaneous rate usually stop responding right after the offset (deadtime period) and recover slowly with a time constant longer than that of short-term adaptation, the rapid component of the model recovery function causes the synapse to recover quickly after the stimulus offset. Since $P_I(t)$ and $C_I(t)$ can not be negative, the magnitude of these recovery

components (Φ_1, Φ_2 in Eqs. A13, A16; the sign of these components is negative at stimulus offset) is also limited by the value of Φ_0 (determined by the spontaneous rate). These limitations make the offset adaptation of the model response unrealistic as compared to the observed physiological data.

The physiological response of AN fibers following the tone-burst offset can be fit (Westerman, 1985) with one exponential recovery function with a desired deadtime period

$$R_{off}(t) = \begin{cases} 0; & t < t_0 \\ A_{sp}\,(1 - e^{-(t-t_0)/t_{sp}}) \end{cases}, \qquad \text{(A22)}$$

where A_{sp} is the spontaneous rate, t_0 is the dead time period and t_{SP} is the recovery time constant of the offset responses. The fitting function has another general form (Harris and Dallos, 1979)

$$\begin{aligned} R_{off}(t) &= (A_{sp} - A_{\min})(1 - e^{-t/t_0}) + A_{\min} \\ &\qquad \text{for R(t)>0}, \\ &= 0 \qquad \text{for R(t)<0} \end{aligned} \qquad \text{(A23)}$$

where A_{min} is a negative value which accounts for the deadtime period of the offset response.

Since an AN fiber doesn't respond during the deadtime period, the recovery process during this period is not well studied, and it is possible there exists a rapid recovery component corresponding to the onset rapid adaptation during this period. The offset characteristic function thus can be described as

$$\begin{aligned} R_{off}(t) &= (A_{sp} - A_{\min})(1 - e^{-t/t_0}) + A_{\min} + A_r^{off} e^{-t/t_r^{off}} \\ &\qquad \text{for } R_{off}(t)>0 \\ &= 0 \qquad \text{for } R_{off}(t)<0 \end{aligned} \qquad \text{(A24)}$$

where A_r^{off} is the magnitude (with a negative value) and t_r^{off} is the time constant of rapid recovery component. Since the model synapse output $P_I C_I$ in the Westerman model cannot be negative, we can assume that the measured synapse output is

$$\begin{aligned} R_{off}(t) &= P_I(t) C_I(t) - A_{shift} \\ &\qquad \text{for } R_{off}(t)>0 \\ &= 0 \qquad \text{for } R_{off}(t)<0 \end{aligned} \qquad \text{(A25)}$$

where A_{shift} is a shift value to guarantee $P_I C_I > 0$ while making $R_{\text{off}}(t)$ the same offset adaptation function as in Eq. A24 (which means $A_{shift} > \left|A_r^{off}\right| + \left|A_{\min}\right|$). The onset of the model synapse output now becomes

$$\begin{aligned} P_I(t)C_I(t) &= R_{on}(t) + A_{shift} \\ &= A_{shift} + A_{ss} + A_r e^{-t/t_r} + A_s e^{-t/t_{st}}. \end{aligned} \tag{A26}$$

Thus, by including this shift, the same equations can be used to describe both the onset and offset adaptation.

Figure 9 shows the onset and offset responses of the synapse model with or without the shift added. The value of the model parameters was derived to have the same onset adaptation. The permeability $P_I(t)$ jumps from P_{I1} to P_{I2} at 5 ms and drops back to P_{I1} at 55 ms. When the shift ($A_{\text{shift}} = 50$) was added, the model parameters (e.g. P_{I1} in Figs. 9c and 9d) were changed to keep the same onset response. The original model and modified model have the same onset responses (Fig. 9a) while the modified model has a greatly improved offset adaptation response (Fig. 9b).

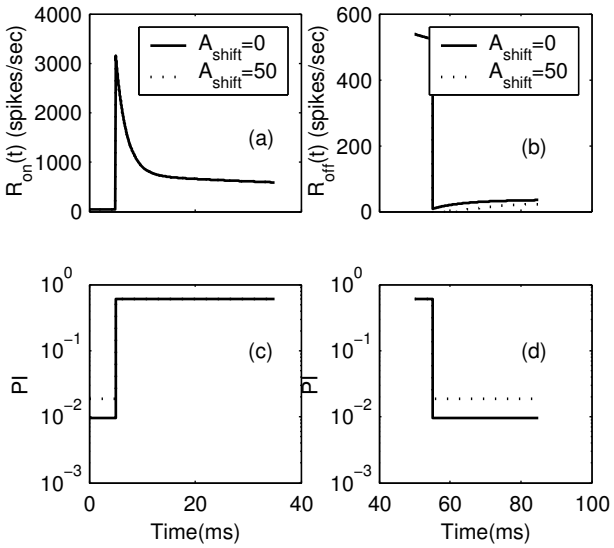


Figure 9: Onset and offset responses of the Westerman model with (dotted line) or without (solid line) the shift value added. The shift value is set to 50 spikes/sec. (a) The model onset responses are unchanged because the model parameters were derived to have the same onset adaptation. (b) The model offset response with the shift recovers more slowly than the model offset response without the shift and has the desired deadtime period. (c), (d) The permeability P_l(t) jumps from P_{l1} to P_{l2} at the onset (5 ms) and drops back to P_{l1} at the offset (55 ms). The model parameters (e.g. P_{l1}) were changed to keep the same onset response.

The effect of A_{shift} on the offset components is illustrated in Fig. 10. Since the rapid component recovers quickly during the deadtime period, the offset adaptation process is dominated by the short-term component. The magnitude of the short-term component (A_{min}-A_{sp} in Eq. A23) determines the duration of the deadtime period and the time-constant (t_{ST}) determines the value of recovery time constant t_{SP}. The recovery time constant is always greater than that of onset short-term adaptation (60 ms), though the value decreases as the shift is increased. The value of A_{min} (Eq. A23) becomes more negative with a larger shift, which produces the desired deadtime period.

The offset adaptation properties of IHC-AN synapse could be used to account for the enhanced phase locking to the stimulus envelope in AN fibers. The AN fibers were less responsive during the dip of envelope due to the offset adaptation, and thus the fibers were more synchronized to the peak of the envelope. Figure 11[20] illustrates the effects of offset adaptation on the modulation transfer function (MTF, see Joris and Yin, 1992 for a comprehensive experimental study) for a model AN fiber (Zhang et al., 2001) to sinusoidally amplitude-modulated (AM) stimulus. The IHC-AN synapse component of the model was replaced with the synapse model presented here (the onset adaptation parameter were the same as in previous study), and different shift value was used to represent different offset adaptation (a shift value of 0 means the synapse model was the same as the original one). The stimulus was always fully

[20] The model results presented here are just illustrative. More extensive explorations have been conducted in another study (Nelson and Carney, submitted).

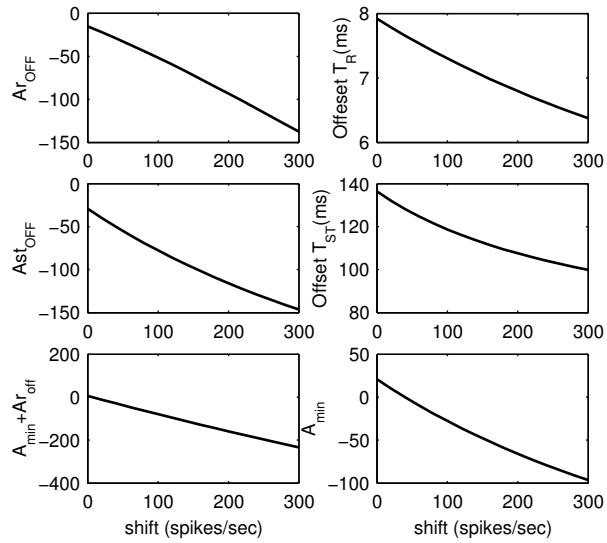


Figure 10: Effects of the parameter A_{shift} in Eq. A26 on the model offset components. The model parameters are derived so they have the same onset adaptation with different A_{shift} values (spontaneous rate A_{sp}=50, A_{ss}=350, PTS=(A_r+A_{st}+A_{ss})/A_{ss}=9, t_R=2 ms, t_{ST}=60 ms, A_r/A_{st}=6.0; see Eq. A19). The upper plots illustrate how the magnitude and time constant of the offset rapid component changes as a function of shift; the middle plots show changes of the offset short-term component. The time constants for both components decrease as shift increases and are larger than the corresponding onset time constants. The rapid component recovers quickly (usually during the deadtime period) and the recovery time constant is thus determined by the short-term component. The magnitude of this component is always negative and decreases (becomes more negative) as shift increases. A_{min} (bottom right; see Eqs. A23, A24) changes from positive to the desired negative as the shift value increases; A_{min} and the time constant of the short-term component determine the duration of the deadtime period. The plot at the bottom left shows the most negative value (including the contribution of the rapid component, which recovers quickly) of the model response without rectification (R_{off}(t=0) in Eq. A24) as a function of the shift value.

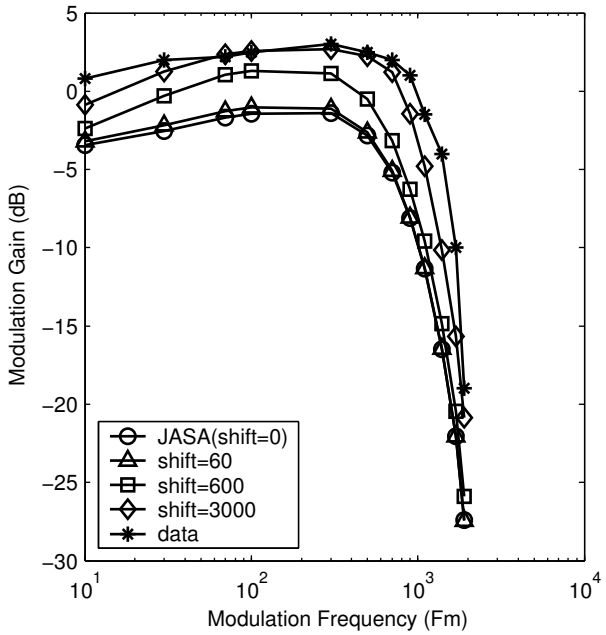


Figure 11: Modulation transfer function for the model AN fibers. The stimulus was always fully modulated at 10 dB SPL with carrier frequency at AN CF of 21 kHz, gated with 10 ms cosine-square function. The modulation gain was defined as 20 log[(modulation of responses)/(modulation of stimulus)] (i.e., modulation gain = 20 log(2R/m), where R was the vector strength of the AN responses and m was the modulation depth; see Joris and Yin, 1992). The model AN fibers with larger shift in synapse had higher modulation gain, and were more consistent with the data from physiological experiments (asterisks in the figure, from Joris and Yin, 1992).

modulated (modulation index m = 1) at 10 dB SPL, with carrier frequency set equal to the AN CF of 21 kHz. The model AN fibers with offset adaptation (e.g. with larger shift value) had increased modulation gain, and an enhanced modulation transfer function more consistent with the data observed from physiological experiments (Joris and Yin, 1992).

A.6 Including the refractoriness in the synapse output to approximate the PST response of AN fibers

As for many neurons, auditory-nerve (AN) activity can be modeled as a renewal process described by its instantaneous rate function λ(t) (also referred to as a hazard function in renewal process theory; see Cox, 1962). The synapse output $s(t)$ is equal to $k(t)q(t)$ for the Meddis model and $P_I(t)C_I(t)$ for the Westerman model. The synapse output usually represents the instantaneous discharge rate of AN fibers without considering the effects of refractoriness. To generate realistic AN discharge spike trains, refractory characteristics must be incorporated into the stimulus-dependent synapse output $s(t)$. Model AN discharge generators (e.g. Carney, 1993) often use a multiplicative relationship between the synapse output and refractoriness to model the instantaneous firing rate $\lambda(t)=s(t)(1-r(t-\mathbf{t}_L))$, where $r(t)$ is the refractoriness factor (0~1) determined by the time since the most recently discharge ($\mathbf{t}_L$). The refractoriness function is represented by a sum of two exponentials (Westerman and Smith, 1985; Carney, 1993; Zhang et al. 2001):

$$r(t-t_L) = c_0 e^{-(t-t_L-R_A)/s0} + c_1 e^{-(t-t_L-R_A)/s1}, \quad \text{for } t-t_L \geq R_A, \qquad = 1, \text{ for } t-t_L < R_A \tag{A27}$$

where R_A is the absolute refractoriness period, and c_0, c_1, s_0, s_1 ($c_0 + c_1 = 1$) are the parameters for two exponentials.

The unconditional firing rate function of a renewal process can be represented by (Cox, 1962)

$$p(t) = \int_{-\infty}^{t} p(x) P_{sur}(t,x) \rho(t,x) dx, \tag{A28}$$

where ρ(t, x) and $P_{sur}(t, x)$ are the hazard and survival functions, respectively, for the renewal process given the previous spike time at x. The survival function $P_{sur}(t, x)$ can be calculated from the hazard function directly (Cox, 1962)

$$P_{sur}(t,x) = e^{-\int_x^t \rho(s,x) ds}. \tag{A29}$$

For the model AN spike generator with refractoriness factor given in Eq. A27, we can described the unconditional firing rate function of the model AN fiber as

$$\begin{aligned} p(t) &= \int_{-\infty}^{t} p(x)P_{sur}(t,x)s(t)\left[1-r(t-x)\right]dx \\ &\approx s(t)\int_{t-T}^{t} p(x)P_{sur}(t,x)\left[1-r(t-x)\right]dx + s(t)\int_{-\infty}^{t-T} p(x)P_{sur}(t,x)dx \end{aligned}, \qquad \text{(A30)}$$

where T is the time period over which the refractoriness can be neglected (r(t, x)≈0 when t-x>T).

One popular technique used to characterize the neural response during the time course of a stimulus, known as the post-stimulus time (PST) histogram, is a discrete form of the above representation and can be described as

$$\begin{aligned} \tilde{p}(k) &= \tilde{s}(k)\Delta \sum_{i=-\infty}^{k-n_T-1} \tilde{p}(i)\tilde{P}_{sur}(k,i) + \tilde{s}(k)\Delta \sum_{i=k-n_T}^{k} \tilde{p}(i)\tilde{P}_{sur}(k,i)\left[1-\tilde{r}(k-i)\right] \\ &= \tilde{s}(k)\Delta\left(\tilde{P}_{TAIL}(k) + \sum_{i=k-n_T}^{k} \tilde{p}(i)\tilde{P}_{sur}(k,i)\left[1-\tilde{r}(k-i)\right]\right) \end{aligned}, \qquad \text{(A31)}$$

where $\tilde{p}(k)$ is the unconditional probability of firing during the discrete time bin, of length Δ, from t_k to t_{k+1}; n_T is the number of time bins during time period T ($n_T = T/\Delta$); $\tilde{r}(\cdot)$, $\tilde{s}(\cdot)$, and $\tilde{P}_{sur}(\cdot)$ are the discrete functions corresponding to their own continuous version; and $\tilde{P}_{TAIL}(k)$ represents the first summation in the first line of the equation. For the small time bin Δ (e.g. $\tilde{s}(k)\Delta<0.1$, for all k) the Poisson process can be approximated by a Bernoulli process (Edwards and Wakefield, 1990), so $\tilde{s}(k)\Delta$ and $1-\tilde{s}(k)\Delta$ represent the probability of one discharge or no discharge during one bin, respectively. The firing probability including the refractoriness, given a previous discharge at time t_i, can be represented as $s(k)\Delta[1-r^i(k)]$ and the survival probability can be represented by

$$\tilde{P}_{sur}(k+1, i) = \tilde{P}_{sur}(k, i)\,(1-\tilde{s}(k+1)\Delta[1-\tilde{r}(k+1-i)]) \qquad \text{(A32)}$$

When refractoriness due to the previous discharge can be ignored (i.e. for $i<k-n_T$), we have $\tilde{P}_{sur}(k+1)= \tilde{P}_{sur}(k)(1-\tilde{s}(k)\Delta)$ and thus $\tilde{P}_{TAIL}(k+1)$ can be calculated recursively:

$$\tilde{P}_{TAIL}(k+1) = (1-\tilde{s}(k)\Delta)\tilde{P}_{TAIL}(k) + \tilde{p}(k-n_T)\tilde{P}_{sur}(k+1, k-n_T). \qquad \text{(A33)}$$

The PST histogram thus can be calculated directly from Eqs. A31-33, given the refractoriness effect r(t) and synapse output s(t). The computational result can precisely represent the PST generated by the discharge generator based on the same input (Fig. 12), and provides an

accurate description of the PST without the need to simulate thousands of repetitions using discharge times.

A.7 Discussion

This study shows that two phenomenological IHC-AN synapse models that were designed to explain the onset adaptation characteristics of AN responses are essentially the same. The Meddis model was simplified based on the assumption of a large value of the sum of two parameters, $l+r$, which is appropriate based on the parameter values used in previous studies of this model (Meddis, 1986, 1988; Sumner et al., 2003). A low value of $l+r$ primarily changes the value of $c(t)$ [$c(t)$ is low-pass filtered version of $q(t)$, see Eq. A11] and choosing $hc(t)$ (Meddis, 1986, 1988) as the synapse output results in a low-pass filtered version of the desired onset adaptation responses (in a recent study by Sumner et al, $kq(t)$ is used as the synapse output instead of $hc(t)$). Also, when $l+r$ is too small, the filter has imaginary roots and the model onset response have undesired oscillations. A high value of $l+r$ is thus desirable in the Meddis model, which can then be simplified to the same form as the Westerman model.

Since the Westerman model and the Meddis model have different structures, they provide different insight to the underlying physical

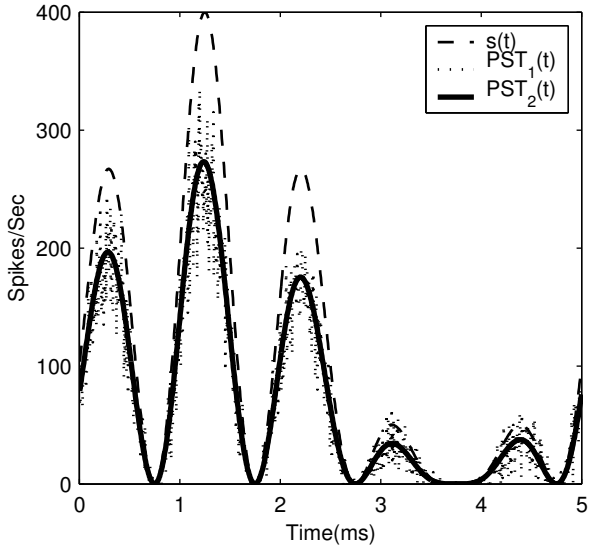


Figure 12: Comparison of PSTs derived from different methods based on the same synapse output $S_{out}(t)$, with and without refractory characteristics. To illustrate how the method applies to arbitrary inputs, the synapse output $S_{out}(t)$ (dashed line) is a periodic amplitude-modulated signal defined as $S_{out}(t) = 100[\sin(400\,\pi\,t)+1][\sin(2000\,\pi\,t)+1]$. The refractoriness is described by Eq. A27 using the parameter values described in Carney (1993; c_0=0.5, c_1=0.5, R_A=0.75 ms, s_0=1 ms, s_1 = 12.5 ms). PST_1 was constructed using a discharge generator (dotted line) and 40,000 repetitions of the $S_{out}(t)$ period, with a bin size of 5μs. PST_2 was calculated directly from $S_{out}(t)$ based on Eq. A31 (solid line) using a sampling rate of 200 kHz.

system. The similarity between the two models comes from the assumption that only one parameter in the model is affected by the stimulus ($P_I(t)$ for the Westerman model and $k(t)$ for the Meddis model). It is possible that other parameters should change with the stimulus and such a change would provide different interpretations of the IHC-AN synapse process for both models. For example, if changing the value of u in the Meddis model (percentage of the reuptake transmitters) upon stimulation gave a more accurate description of the actual IHC-AN synapse responses, it would be easy to gain insight using the Meddis model. However, the same change in Westerman's model would affect several variables: P_L, P_G, V_L would all change with level to obtain the same response properties of the model, and in that case this model would not provide intuition concerning the actual processing of AN-IHC synapse.

The study provides an analysis tool to link the parameters of the adaptation characteristics directly to the model parameters and to determine how the derived parameter values differ from each other. It is possible to pick any one set of parameters and derive the value of the other model's parameters. For example, the values of the onset component (t_R, t_{ST}, A_r/A_{st}) may not be very convincing based on a fit of the PST histograms for low spontaneous rate fibers, and we may prefer some model parameters to be fixed across all model fibers (e.g. x or u for Meddis model). We can fix these values (x, u, A_{sp}, A_{ss}, etc.) and derive the values of the others in the model. The new properties of the parameter values derived in this way may provide more insight into the differences underlying the different AN fiber types.

The study shows that a simple shift can improve the model's offset adaptation without compromising the model's onset adaptation characteristics. Inclusion of offset adaptation in the model improves the model's prediction of the temporal aspects of the AN fiber responses and thus benefits other modeling studies based on the temporal responses of AN fibers (e.g. modulation studies). Such a shift in the model structure can be interpreted as a constant leak in the synapse cleft before the transmitter reaches the post-synaptic site. This interpretation is different from that in the Meddis model, where the loss of transmitter in the cleft is proportional to the transmitter concentration. Since the offset adaptation of the AN fiber is related to the spontaneous rate (Harris and Dallos, 1979; Westerman, 1985), it is reasonable for the value of the shift to vary as a function of the spontaneous rate of the fiber as well. A further study of how the value of the shift is related to the other characteristics of the model fiber responses will be helpful. Also, the results show that the offset adaptation time constant is closely related to the short-term constant of the onset adaptation for the model explored. In fact, if the model onset adaptation is measured from level L_0 to level L_1 (L_0 is the level before the onset and L_1 is the level after onset) and offset adaptation is measured from level L_2 to L_1 (L_2 is the level before the offset and L_1 is the level after the offset), the time constants for the onset short-term component and for the offset adaptation are the same, based on Eq. A16. Whether this is a limitation of the model or reflects some underlying mechanism of the synapse still requires examination.

How the synapse dynamics are affected by the IHC voltage (or by the stimulus) is very important to the synapse response properties, and yet this relationship is highly simplified in most modeling studies. Even here, where we assume that only one parameter (the immediate permeability, P_I) in the synapse is affected by V_{ihc}, the model is complicated since the V_{ihc}-P_I relationship varies across different fibers: the voltage responses of the IHC at different BM locations to CF tone have different AC/DC components (Cheatham and Dallos, 1993); and responses from different AN fiber types have different threshold and rate-level curves that depend on the V_{ihc}-P_I relationship. The analysis of the adaptation of the synapse model provides some general constrains on the V_{ihc}-P_I relationship. For example, the value of P_I at rest (V_{ihc}=0) and at a certain high stimulus level L (the value of V_{ihc} at level L can be estimated from the composite model), with desired adaptation properties, can be derived from the synapse model directly. The value of P_I at threshold level can also be determined from the synapse model responses. Another useful conclusion from this study is that the saturation of the synapse response is largely contributed by the synapse model itself, rather than by the value of P_I (Eq. A21). That fact that P_I does not saturate as V_{ihc} increases at high levels (Carney, 1993, Zhang et al., 2001) is important for the model to predict AN fiber temporal response properties (e.g. synchronization index) across different levels.

Modeling the IHC-AN synapse is still a very challenging task due to both the synapse structure (how the reservoirs are connected to each other) and the processing dynamics (how the V_{ihc} affects model parameters such as permeabilities). While it is easy to complicate the IHC-AN synapse model with more physiologically realistic structures, whether such complication will benefit the model and provide more accurate responses should be carefully studied. Also modeling studies of other adaptation measures (Hewitt and Meddis, 1991; Sumner et al., 2003) which focused on models with preset parameters should be extended. An analysis of the relationship of these adaptation measures and how they are affected by the model parameters will greatly improve our knowledge of both IHC-AN synapse models and of the underlying physical system.

Appendix B. Relationship between adaptation characteristics and the Meddis-model parameters

The Meddis model with three neurotransmitter reservoirs (q, w, c) can be described by the simplified Eqs. A9 and A10. The model parameters M, y, x, u are all constant, and k is the level-dependent permeability that is denoted as k1 before the onset and as k2 after the onset. If we assume kq is the model synapse output which matches the characteristic function of the onset adaptation

$$R_{on}(t) = A_{ss} + A_r e^{-t/t_R} + A_{st} e^{-t/t_{ST}} \tag{B1}$$

with a spontaneous rate of A_{sp}, the relationship between the model parameters and adaptation characteristics can be described by the following equations ($-1/t_r$ and $-1/t_{st}$ are the poles of Q(s) in Eq. A12):

$$\frac{k_1}{k_2} = \frac{A_{sp}}{A_r + A_{st} + A_{ss}} \tag{B2}$$

$$\frac{yMk_1}{y + k_1(1-u)} = A_{sp} \tag{B3}$$

$$\frac{yMk_2}{y + k_2(1-u)} = A_{ss} \tag{B4}$$

$$x + k_2 + y = 1/t_R + 1/t_{ST} \tag{B5}$$

$$x(y + k_2(1-u)) = 1/(t_R t_{ST}) \tag{B6}$$

$$\frac{k_2}{k_1}(k_2 - k_1)A_{sp} = A_r/t_R + A_{ss}/t_{ST} \tag{B7}$$

Several intermediate parameters are useful for the derivation of the model parameters:

$$S_r = 1/t_R + 1/t_{ST} \tag{B8}$$

$$S_{r2} = A_r / t_R + A_{st} / t_{ST} \tag{B9}$$

$$\mathrm{P_r} = 1/(t_R t_{ST}) \tag{B10}$$

From Eqs. B2 and B7, k1 and k2 can be specified as:

$$k_2 = S_{r2} /(A_{on} - A_{sp}), k_1 = A_{sp} / A_{on}\, k_2 . \tag{B11}$$

Several other intermediate parameters are defined as:

$$\beta = (A_{ss} - A_{sp}) k_1 k_2 /(A_{sp} k_2 - A_{ss} k_1) , \tag{B12}$$

$$a = (\beta + k_2)\beta,\ b = -(S_r - k_2)(\beta + k_2),\ c = \mathrm{P_r} . \tag{B13}$$

$$z = -\frac{-b + \sqrt{b^2 - 4ac}}{2a} \tag{B14}$$

The other model parameters then are set as follows:

$$u = 1 - z , \tag{B15}$$

$$y = \beta z , \tag{B16}$$

$$x = S_r - k_2 - y , \tag{B17}$$

$$M = \frac{A_{sp}(y + k_1 z)}{y k_1} . \tag{B18}$$

References

Abeles,M. (1982). Role of the cortical neuron: integrator or coincidence detector? Israel journal of medical sciences 18, 83-92.

Berg,B.G., Nguyen,Q.T., and Green,D.M. (1992). Discrimination of narrow-band spectra. I: Spectral weights and pitch cues. Journal of the Acoustical Society of America 92, 1911-1918.

Bernstein,L.R. and Green,D.M. (1987). Detection of simple and complex changes of spectral shape. Journal of the Acoustical Society of America 82, 1587-1592.

Blackburn,C.C. and Sachs,M.B. (1989). Classification of unit types in the anteroventral cochlear nucleus: PST histograms and regularity analysis. Journal of neurophysiology 62, 1303-1329.

Blackburn,C.C. and Sachs,M.B. (1992). Effects of OFF-BF tones on responses of chopper units in ventral cochlear nucleus. I. Regularity and temporal adaptation patterns. Journal of neurophysiology 68, 124-143.

Bourk,T. (1976) Electrical responses of neural units in the anteroventral cochlear nucleus of the cat. PhD Dissertation. Massachusetts Institute of Technology.

Braida,L.D. and Durlach,N. (1988). Peripheral and Central Factors in Intensity Perception. In Auditory function : neurobiological bases of hearing, G. M. Edelman, W. E. Gall, W. M. Cowan, and Neurosciences Institute (eds.) (New York: Wiley), pp. 515-558.

Burkitt,A.N. and Clark,G.M. (2001). Synchronization of the neural response to noisy periodic synaptic input. Neural Computation 13, 2639-2672.

Buus,S. (1985). Release from masking caused by envelope fluctuations. Journal of the Acoustical Society of America78, 1958-65.

Buus,S., Zhang,L., and Florentine,M. (1996). Stimulus-driven, time-varying weights for comodulation masking release. Journal of the Acoustical Society of America 99, 2288-2297.

Cant,N.B. (1992). The Mammalian auditory pathway : neuroanatomy. In The Mammalian auditory pathway : neuroanatomy, D. B. Webster, A. N. Popper, and R. R. Fay, eds. (New York: Springer-Verlag), pp. 66-116.

Carney,L.H. (1990). Sensitivities of cells in anteroventral cochlear nucleus of cat to spatiotemporal discharge patterns across primary afferents. Journal of neurophysiology 64, 437-456.

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

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

Carney,L.H., Heinz,M.G., Evilsizer,M.E., Gilkey,R.H., and Colburn,H.S. (2002). Auditory phase opponency: A temporal model for masked detection at low frequencies. Acta Acustica United with Acustica 88, 334-347.

Carney,L.H. and Yin,T.C. (1988). Temporal coding of resonances by low-frequency auditory nerve fibers: single-fiber responses and a population model. Journal of neurophysiology 60, 1653-1677.

Caspary,D.M., Backoff,P.M., Finlayson,P.G., and Palombi,P.S. (1994). Inhibitory inputs modulate discharge rate within frequency receptive fields of anteroventral cochlear nucleus neurons. Journal of neurophysiology 72, 2124-2133.

Cheatham,M.A. and Dallos,P. (1993). Longitudinal Comparisons of Ihc Ac and Dc Receptor Potentials Recorded from the Guinea-Pig Cochlea. Hearing Research 68, 107-114.

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

Colburn,H.S. (1977). Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise. Journal of the Acoustical Society of America 61, 525-533.

Colburn,H.S., Carney,L.H., and Heinz,M.G. (2003). Quantifying the information in auditory-nerve responses for level discrimination. J Assoc Res Otolaryngol 4, 294-311.

Colburn,H.S., Han,Y.A., and Culotta,C.P. (1990). Coincidence model of MSO responses. Hearing Research 49, 335-346.

Colburn,H.S. (1996). Computational Models of Binaural Processing. In Auditory computation, H. L. Hawkins, ed. (New York: Springer Verlag), pp. xii, 517.

Colburn,H.S. and Durlach,N. (1978). Models of binaural interaction. In Hearing, E. C. Carterette and M. P. Friedman, eds. (New York: Academic Press), pp. 467-518.

Colburn,H.S. and Moss,P.J. (1981) Binaural interaction models and mechanisms IN Neuronal mechanisms of hearing. Syka, Josef and Aitkin, Lindsay (eds.). pp.283-288. New York, Plenum..

Cooper,N.P. and Rhode,W.S. (1997). Mechanical responses to two-tone distortion products in the apical and basal turns of the mammalian cochlea. Journal of neurophysiology 78, 261-270.

Cox,D.R. (1962). Renewal theory. (London, New York, Methuen; Wiley), pp. 142.

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

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

Delgutte,B. (1996). Physiological models for basic auditory percepts. In Auditory computation, H. L. Hawkins and T.A.McMullen, eds. (New York: Springer Verlag), pp. 157-220.

Deng,L. and Geisler,C.D. (1987). A composite auditory model for processing speech sounds. Journal of the Acoustical Society of America 82, 2001-2012.

Durlach,N.I., Braida,L.D., and Ito,Y. (1986). Towards a model for discrimination of broadband signals. Journal of the Acoustical Society of America 80, 63-72.

Durlach,N.I., Mason,C.R., Kidd,G., Jr., Arbogast,T.L., Colburn,H.S., and Shinn-Cunningham,B.G. (2003). Note on informational masking. Journal of the Acoustical Society of America 113, 2984-2987.

Durlach,N. and Colburn,H.S. (1978). Binaural phenomena. In Hearing, E. C. Carterette and M. P. Friedman (eds.), (New York: Academic Press), pp. 369-462.

Eddins,D.A. and Green,D.M. (1995). Temporal Integration and Temporal Resolution. In Hearing, B. C. J. Moore, ed. (San Diego: Academic Press), pp. 207-242.

Edwards,B. and Wakefield,G. (1990). On the statistics of binned neural point processes: the Bernoulli aproximation and AR representation of the PST histogram. Biological cybernetics 64, 145-153.

Evans,E.F., Pratt,S.R., and Cooper,N.P. (1989). Correspondence between behavioral and physiological frequency selectivity in the guinea pig. British journal of audiology 23, 151-152.

Fletcher,H. (1940). Auditory patterns. Review of Modern Physics, 12, 47-65.

Florentine,M. (1986). Level discrimination of tones as a function of duration. Journal of the Acoustical Society of America 79, 792-798.

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

Furukawa,T. and Matsuura,S. (1978). Adaptive rundown of excitatory post-synaptic potentials at synapses between hair cells and eight nerve fibres in the goldfish. The Journal of physiology 276, 193-209.

Gaumond,R.P., Kim,D.O., and Molnar,C.E. (1983). Response of cochlear nerve fibers to brief acoustic stimuli: role of discharge-history effects. Journal of the Acoustical Society of America 74, 1392-1398.

Gaumond,R.P., Molnar,C.E., and Kim,D.O. (1982). Stimulus and recovery dependence of cat cochlear nerve fiber spike discharge probability. Journal of neurophysiology 48, 856-873.

Gerstner,W. and Kistler,W.M. (2002). Spiking neuron models: single neurons, populations, plasticity. (Cambridge, U.K.; New York: Cambridge University Press).

Gilkey,R.H. (1987). Spectral and temporal comparisons in auditory masking. In Auditory Processing of Complex Sounds, W. Yost and C. Watson (eds.), pp. 26-36.

Gilkey,R.H. and Robinson,D.E. (1986). Models of auditory masking: a molecular psychophysical approach. Journal of the Acoustical Society of America 79, 1499-1510.

Goldberg,J.M. and Brown,P.B. (1969). Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: Some physiological mechanisms of sound localization. Journal of neurophysiology 22, 613-636.

Gray,C.M. (1999). The temporal correlation hypothesis of visual feature integration: still alive and well. Neuron 24, 31-25.

Green,D.M., Kidd,G., Jr., and Picardi,M.C. (1983). Successive versus simultaneous comparison in auditory intensity discrimination. Journal of the Acoustical Society of America 73, 639-643.

Green,D.M. (1988). Profile analysis : auditory intensity discrimination. (New York: Oxford University Press).

Green,D.M. and Swets,J.A. (1966). Signal detection theory and psychophysics. (New York, Wiley).

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

Hall,J.W. (1986). The effect of across-frequency differences in masking level on spectro-temporal pattern analysis. Journal of the Acoustical Society of America 79, 781-787.

Hall,J.W., III and Grose,J.H. (1988). Comodulation masking release: evidence for multiple cues. Journal of the Acoustical Society of America 84, 1669-1675.

Hall,J.W., Haggard,M.P., and Fernandes,M.A. (1984). Detection in noise by spectro-temporal pattern analysis. Journal of the Acoustical Society of America 76, 50-56.

Harris,D.M. and Dallos,P. (1979). Forward masking of auditory nerve fiber responses. Journal of neurophysiology 42, 1083-1107.

Heinz,M.G., Colburn,H.S., and Carney,L.H. (2001a). Evaluating auditory performance limits: I. one-parameter discrimination using a computational model for the auditory nerve. Neural Computation 13, 2273-316.

Heinz,M.G., Colburn,H.S., and Carney,L.H. (2001b). Evaluating auditory performance limits: II. One-parameter discrimination with random-level variation. Neural Computation 13, 2317-38.

Heinz,M.G., Colburn,H.S., and Carney,L.H. (2001c). Rate and timing cues associated with the cochlear amplifier: level discrimination based on monaural cross-frequency coincidence detection. Journal of the Acoustical Society of America110, 2065-84.

Heinz,M.G., Colburn,H.S., and Carney,L.H. (2002). Quantifying the implications of nonlinear cochlear tuning for auditory-filter estimates. Journal of the Acoustical Society of America111, 996-1011.

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

Hewitt,M.J. and Meddis,R. (1991). An Evaluation of 8 Computer-Models of Mammalian Inner Hair-Cell Function. Journal of the Acoustical Society of America 90, 904-917.

Houtsma,A.J.M. (1995). Pitch Perception. In Hearing, B. C. J. Moore, ed. (San Diego: Academic Press), pp. 267-291.

Irvine,D.R.F. (1992). Physiology of the Auditory Brainstem. In The Mammalian auditory pathway : neurophysiology, A. N. Popper and R. R. Fay, eds. (New York: Springer-Verlag), pp. 153-231.

Jeffress,L.A. (1948). A place theory of sound localization. Journal of comparative and physiological psychology, 41, 35-39.

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

Johnson,D.H. and Swami,A. (1983). The transmission of signals by auditory-nerve fiber discharge patterns. Journal of the Acoustical Society of America 74, 493-501.

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

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

Joris,P.X., Smith,P.H., and Yin,T.C. (1998). Coincidence detection in the auditory system: 50 years after Jeffress. Neuron 21, 1235-1238.

Joris,P.X. and Yin,T.C. (1992). Responses to amplitude-modulated tones in the auditory nerve of the cat. Journal of the Acoustical Society of America 91, 215-232.

Kalluri,S. and Delgutte,B. (2003a). Mathematical models of cochlear nucleus onset neurons: I. Point neuron with many weak synaptic inputs. Journal of computational neuroscience. 14, 71-90.

Kalluri,S. and Delgutte,B. (2003b). Mathematical models of cochlear nucleus onset neurons: II. model with dynamic spike-blocking state. Journal of computational neuroscience. 14, 91-110.

Kay,R.H. (1982). Hearing of modulation in sounds. Physiological reviews. 62, 894-975.

Kempter,R., Gerstner,W., van Hemmen,J.L., and Wagner,H. (1998). Extracting oscillations: Neuronal coincidence detection with noisy periodic spike input. Neural Computation 10, 1987-2017.

Kiang,N.Y.S., Watanabe,T., Thomas,E.C., and Clark,L.F. (1965). Discharge Patterns of SIngle Fibers in the Cat's Auditory Nerve. MIT Press).

Kidd,G., Jr., Mason,C.R., Brantley,M.A., and Owen,G.A. (1989). Roving-level tone-in-noise detection. Journal of the Acoustical Society of America86, 1310-7.

Kidd,G., Jr., Mason,C.R., Uchanski,R.M., Brantley,M.A., and Shah,P. (1991). Evaluation of simple models of auditory profile analysis using random reference spectra. Journal of the Acoustical Society of America 90, 1340-1354.

Kidd,G., Jr., Uchanski,R.M., Mason,C.R., and Deliwala,P.S. (1993). Discriminability of narrow-band sounds in the absence of level cues. Journal of the Acoustical Society of America 93, 1028-1037.

Kipke,D. and Levy,K. (1997). Sensitivity of the cochlear nucleus octopus cell to synaptic and membrane properties: A modeling study. Journal of the Acoustical Society of America 102, 403-412.

Kistler,W.M., Gerstner,W., and van Hemmen,J.L. (1997). Reduction of Hodgkin-Huxley equations to a single-variable threshold model. Neural Computation 9, 1015-1045.

Konig,P., Engel,A.K., and Singer,W. (1996). Integrator or coincidence detector? The role of the cortical neuron revisited. Trends in Neurosciences 19, 130-137.

Kopp-Scheinpflug,C., Dehmel,S., Dorrscheidt,G.J., and Rubsamen,R. (2002). Interaction of excitation and inhibition in anteroventral cochlear nucleus neurons that receive large endbulb synaptic endings. Journal of neuroscience 22, 11004-11018.

Lentz,J.J. and Richards,V.M. (1997). Sensitivity to changes in overall level and spectral shape: an evaluation of a channel model. Journal of the Acoustical Society of America 101, 3625-3635.

Liberman,M.C. (1991). Central projections of auditory-nerve fibers of differing spontaneous rate. I. Anteroventral cochlear nucleus. J.Comp Neurol. 313, 240-258.

Liberman,M.C. (1993). Central projections of auditory nerve fibers of differing spontaneous rate, II: Posteroventral and dorsal cochlear nuclei. J.Comp Neurol. 327, 17-36.

Licklider JCR (1951). The duplex theory of pitch perception. Experientia 7, 128-137.

Loeb,G.E., White,M.W., and Merzenich,M.M. (1983). Spatial cross-correlation. A proposed mechanism for acoustic pitch perception. Biological cybernetics 47, 149-163.

Manis,P.B. and Marx,S.O. (1991). Outward currents in isolated ventral cochlear nucleus neurons. Journal of neuroscience 11, 2865-2880.

May,B.J. and Sachs,M.B. (1992). Dynamic range of neural rate responses in the ventral cochlear nucleus of awake cats. Journal of neurophysiology 68, 1589-1602.

Meddis,R. (1986). Simulation of mechanical to neural transduction in the auditory receptor. Journal of the Acoustical Society of America 79, 702-711.

Meddis,R. (1988). Simulation of Auditory Neural Transduction - Further-Studies. Journal of the Acoustical Society of America 83, 1056-1063.

Molnar, C.E. and Pfeiffer, R.R. (1968). Interpretation of spontaneous spike discharge patterns of neurons in the cochlear nucleus. Proceedings of the IEEE 56, 993-1004.

Moore,B.C. (1978). Psychophysical tuning curves measured in simultaneous and forward masking. Journal of the Acoustical Society of America 63, 524-532.

Moore,B.C. (1986). Parallels between frequency selectivity measured psychophysically and in cochlear mechanics. Scandinavian audiology. Supplementum. 25, 139-152.

Moore,B.C.J. (1995). Hearing. (San Diego: Academic Press).

Neff,D.L. and Green,D.M. (1987). Masking produced by spectral uncertainty with multicomponent maskers. Perception & Psychophysics. 41, 409-415.

Nelson, P.C. and Carney, L.H (submitted). A phenomenological model of peripheral and central neural responses to amplitude-modulated tones.

Oertel,D. (1983). Synaptic responses and electrical properties of cells in brain slices of the mouse anteroventral cochlear nucleus. Journal of neuroscience 3, 2043-2053.

Oertel,D. (1985). Use of brain slices in the study of the auditory system: spatial and temporal summation of synaptic inputs in cells in the anteroventral cochlear nucleus of the mouse. Journal of the Acoustical Society of America 78, 328-333.

Oertel,D., Bal,R., Gardner,S.M., Smith,P.H., and Joris,P.X. (2000). Detection of synchrony in the activity of auditory nerve fibers by octopus cells of the mammalian cochlear nucleus. Proceedings of the National Academy of Sciences U.S.A 97, 11773-11779.

Oxenham,A.J. and Bacon,S.P. (2003). Cochlear compression: perceptual measures and implications for normal and impaired hearing. Ear and Hearing, 24, 352-366.

Oxenham,A.J. and Shera,C.A. (2003). Estimates of human cochlear tuning at low levels using forward and simultaneous masking. Journal of the Association for Research in Otolaryngology 4, 541-554.

Patterson,J.H. and Green,D.M. (1970). Discrimination of transient signals having identical energy spectra. Journal of the Acoustical Society of America 48, 894-905.

Patterson,R.D. (1976). Auditory filter shapes derived with noise stimuli. Journal of the Acoustical Society of America 59, 640-654.

Plack,C.J. (1998). Beneficial effects of notched noise on intensity discrimination in the region of the "severe departure". Journal of the Acoustical Society of America 103, 2530-2538.

Plack,C.J. and Viemeister,N.F. (1992). The effects of notched noise on intensity discrimination under forward masking. Journal of the Acoustical Society of America 92, 1902-1910.

Rhode,W.S. and Greenberg,S. (1992). Physiology of the Cochlear Nuclei. In The Mammalian auditory pathway : neurophysiology, A. N. Popper and R. R. Fay, eds. (New York: Springer-Verlag), pp. 94-152.

Rhode,W.S. and Greenberg,S. (1994). Lateral suppression and inhibition in the cochlear nucleus of the cat. Journal of neurophysiology 71, 493-514.

Rhode,W.S. and Smith,P.H. (1985). Characteristics of tone-pip response patterns in relationship to spontaneous rate in cat auditory nerve fibers. Hearing Research 18, 159-168.

Richard Lyon and Shihab Shamma (1996). Auditory Representations of Timbre and Pitch. In Auditory computation, H. L. Hawkins, ed. (New York: Springer Verlag), pp. 221.

Richards,V.M. (1992). The detectability of a tone added to narrow bands of equal-energy noise. Journal of the Acoustical Society of America 91, 3424-3435.

Richards,V.M. and Buss,E. (1996). Frequency correlation functions for the detection of a tone added to modulated noise maskers. Journal of the Acoustical Society of America 99, 1645-1652.

Richards,V.M., Buss,E., and Tian,L. (1997). Effects of modulator phase for comodulation masking release and modulation detection interference. Journal of the Acoustical Society of America 102, 468-476.

Richards,V.M. and Nekrich,R.D. (1993). The incorporation of level and level-invariant cues for the detection of a tone added to noise. Journal of the Acoustical Society of America 94, 2560-2574.

Rothman,J.S. and Manis,P.B. (2003a). Differential expression of three distinct potassium currents in the ventral cochlear nucleus. Journal of neurophysiology 89, 3070-3082.

Rothman,J.S. and Manis,P.B. (2003b). Kinetic analyses of three distinct potassium conductances in ventral cochlear nucleus neurons. Journal of neurophysiology 89, 3083-3096.

Rothman,J.S. and Manis,P.B. (2003c). The roles potassium currents play in regulating the electrical activity of ventral cochlear nucleus neurons. Journal of neurophysiology 89, 3097-3113.

Rothman,J.S. and Young,E.D. (1996). Enhancement of Neural Synchronization in Computational Models of Ventral Cochlear Nucleus Bushy Cells. Auditory Neuroscience 2, 47-62.

Rothman,J.S., Young,E.D., and Manis,P.B. (1993). Convergence of auditory nerve fibers onto bushy cells in the ventral cochlear nucleus: implications of a computational model. Journal of neurophysiology 70, 2562-2583.

Roy,S.A. and Alloway,K.D. (2001). Coincidence detection or temporal integration? What the neurons in somatosensory cortex are doing. Journal of neuroscience 21, 2462-2473.

Russell,I.J. and Sellick,P.M. (1978). Intracellular studies of hair cells in the mammalian cochlea. The Journal of physiology 284, 261-290.

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

Schreiner, C. E. and Langner, G. (1988). Coding of Temporal Patterns in the Central Auditory Nervous System. In Auditory function: neurobiological bases of hearing, G. M. Edelman, W. E. Gall, W. M. Cowan, and Neurosciences Institute (eds.), (New York: Wiley), pp. 337-362.

Schwid,H.A. and Geisler,C.D. (1982). Multiple reservoir model of neurotransmitter release by a cochlear inner hair cell. Journal of the Acoustical Society of America 72, 1435-1440.

Shamma,S. and Klein,D. (2000). The case of the missing pitch templates: how harmonic templates emerge in the early auditory system. Journal of the Acoustical Society of America 107, 2631-2644.

Shamma,S.A. (1985). Speech processing in the auditory system. II: Lateral inhibition and the central processing of speech evoked activity in the auditory nerve. Journal of the Acoustical Society of America 78, 1622-1632.

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

Singer,W. and Gray,C.M. (1995). Visual feature integration and the temporal correlation hypothesis. Annual Review Neuroscience. 18, 555-586.

Slaney,M. and Lyon,R. (1993). On the importance of time - a temporal representation of sound. In Visual representations of speech signals, M. Cooke, S. Beet, and M. Crawford, eds. (Chichester: J. Wiley & Sons).

Smith,P.H. (1995). Structural and Functional Differences Distinguish Principal from Nonprincipal Cells in the Guinea-Pig Mso Slice. Journal of Neurophysiology 73, 1653-1667.

Smith,R.L. (1977). Short-term adaptation in single auditory nerve fibers: some poststimulatory effects. Journal of neurophysiology 40, 1098-1111.

Smith,R.L., Brachman,M.L., and Frisina,R.D. (1985). Sensitivity of auditory-nerve fibers to changes in intensity: a dichotomy between decrements and increments. Journal of the Acoustical Society of America 78, 1310-1316.

Smith,R.L. and Zwislocki,J.J. (1975). Short-term adaptation and incremental responses of single auditory-nerve fibers. Biological cybernetics 17, 169-182.

Southworth,C. and Berg,B.G. (1995). Multiple Cues for the Discrimination of Narrow-Band Sounds. Journal of the Acoustical Society of America 98, 2486-2492.

Stein,R.B. (1965). A Theoretical Analysis of Neuronal Variability. Biophys.J. 91, 173-194.

Stover,L.J. and Feth,L.L. (1983). Pitch of narrow-band signals. Journal of the Acoustical Society of America 73, 1701-1707.

Sumner,C.J., Lopez-Poveda,E.A., O'Mard,L.P., and Meddis,R. (2003). Adaptation in a revised inner-hair cell model. Journal of the Acoustical Society of America 113, 893-901.

Tuckwell,H.C. and Richter,W. (1978). Neuronal interspike time distributions and the estimation of neurophysiological and neuroanatomical parameters. J.Theor.Biol. 71, 167-183.

Tuckwell,H.C. (1988). Introduction to theoretical neurobiology. (Cambridge [Cambridgeshire] ; New York: Cambridge University Press).

Tuckwell,H.C. (1989). Stochastic processes in the neurosciences. (Philadelphia, Pa.: Society for Industrial and Applied Mathematics).

Viemeister,N.F. (1979). Temporal modulation transfer functions based upon modulation thresholds. Journal of the Acoustical Society of America 66, 1364-1380.

Viemeister,N.F. (1988). Intensity coding and the dynamic range problem. Hearing Research 34, 267-274.

Westerman,L.A. (1985) Adaptation and Recovery of Auditory Nerve Responses. PhD Dissertation. Syracuse University.

Westerman,L.A. and Smith,R.L. (1984). Rapid and short-term adaptation in auditory nerve responses. Hearing Research 15, 249-260.

Westerman,L.A. and Smith,R.L. (1988). A Diffusion-Model of the Transient-Response of the Cochlear Inner Hair Cell Synapse. Journal of the Acoustical Society of America 83, 2266-2276.

Winslow,R.L. and Sachs,M.B. (1988). Single-tone intensity discrimination based on auditory-nerve rate responses in backgrounds of quiet, noise, and with stimulation of the crossed olivocochlear bundle. Hearing Research 35, 165-189.

Winter,I.M. and Palmer,A.R. (1991). Intensity coding in low-frequency auditory-nerve fibers of the guinea pig. Journal of the Acoustical Society of America 90, 1958-1967.

Wu,S.H. and Oertel,D. (1987). Maturation of synapses and electrical properties of cells in the cochlear nuclei. Hearing Research 30, 99-110.

Yin,T.C. and Chan,J.C. (1990). Interaural time sensitivity in medial superior olive of cat. Journal of neurophysiology 64, 465-488.

Yin,T.C.T. and Joseph C.K.Chan (1988). Neural Mechanisms Underlying Interaural Time Sensitivity to Tones and Noise. In Auditory function : neurobiological bases of hearing, G. M. Edelman, W. E. Gall, W. M. Cowan, and Neurosciences Institute (, eds. (New York: Wiley), pp. 385-430.

Young,E.D., Shofner,W.P., White,J.A., Heanne-Marie,R., and Voigt,H.F. (1988). Respons Properties of Cochlear Nucleus Neurons in Relationship to Physiological Mechanisms. In Auditory function : neurobiological bases of hearing, G. M. Edelman, W. E. Gall, W. M. Cowan, and Neurosciences Institute (New York N.Y.), eds. (New York: Wiley).

Zhang,X., Heinz,M.G., Bruce,I.C., and Carney,L.H. (2001). A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression. Journal of the Acoustical Society of America109, 648-70.

Biographical Note