BOSTON UNIVERSITY

COLLEGE OF ENGINEERING

Dissertation

COMPUTATIONAL AND STATISTICAL ANALYSIS OF AUDITORY PERIPHERAL PROCESSING FOR VOWEL-LIKE SIGNALS

by

QING TAN

B.S., Tsinghua University, 1997

M.S., Boston University, 2000

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2003

Approved by

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

Second Reader_____
                           Laurel H. Carney, Ph.D.
                           Professor of Bioengineering and Neuroscience, Syracuse University

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

Fourth Reader_____
                           Allyn E. Hubbard, Ph.D.
                           Associate Professor of Electrical and Computer Engineering

Fifth Reader   _____
                           Barbara G. Shinn-Cunningham, Ph.D.
                           Assistant Professor of Cognitive and Neural Systems

Acknowledgments

I would like to thank Steve Colburn, David Mountain, Allyn Hubbard, and Barbara Shinn-Cunningham for their participation on my dissertation committee. The completion of this dissertation work would not have been possible without the many meaningful discussions with them.

I consider myself extremely fortunate to have had the opportunity to work with Laurel Carney for the past five years. As a stimulating advisor and an understanding friend, she contributed so much to this work that it is very difficult for me to describe her invaluable help using my limited English vocabulary. She gave me the freedom of choosing the research topic I like and exploring my own "wild ideas". She also gave me the just-in-time guidance and advice whenever I needed them. I am sure my future work and life will greatly benefit from the experience of working with her.

I enjoyed studying and working in the Department of Biomedical Engineering and the Hearing Research Center of Boston University. Xuedong Zhang, Neil Letendre, Paul Nelson, and Michael Heinz have provided much help for this project. Additionally, I thank Sean Davidson, Michael Anzalone and Satish Iyengar for making our weekly net-meetings more joyful and productive. I thank the National Science Foundation and the National Institute of Health for their financial support in the past five years.

Finally, I am deeply grateful to my father, Guanrong Tan, and my mother, Luyue Sun, for their patience, support and encouragement ever since I was born. I would like to give special thanks to my wife, Haining Wang, for her never-ending love and smiles.

COMPUTATIONAL AND STATISTICAL ANALYSIS OF AUDITORY PERIPHERAL

PROCESSING FOR VOWEL-LIKE SIGNALS

(Order No.   )

QING TAN

Boston University, College of Engineering, 2003

Major Professor: Laurel H. Carney, Professor of Bioengineering and Neuroscience,

Syracuse University

ABSTRACT

The auditory nerve (AN) transfers information about sound stimuli from the inner

ear to the brainstem, which further projects to higher levels of the auditory nervous

system. A detailed knowledge of the discharge patterns of AN fibers is important for

understanding how sounds are encoded at the input stage of the auditory system. The

study presented here investigates neural mechanisms for encoding and processing sound

signals, including harmonic complexes and synthesized speech signals. A better

understanding of the sound encoding mechanisms in the auditory system can benefit

many other related studies, such as the building of speech-signal processors or

recognizers as well as the design of hearing aids.

The initial focus of this study is the coding of complex sounds in a quiet

background. The analysis is then extended to the coding of complex sounds in a noisy

background. Two different approaches, one directly based on the discharge pattern of AN

fibers and the other one based on the activity of a coincidence-detection mechanism

operating on the AN fiber patterns, are used to evaluate the role of signal processing in

the auditory periphery, which includes the outer, middle, and inner ears. Predictions

based on average discharge count information (the number of neural action potentials in

one trial) are compared with results based on the mechanisms using both count and

temporal information (the times of neural action potentials). The results show that the

predictions using both count and temporal information are more consistent with the

psychophysical data than the predictions based on average-count information. The results

of this study also suggest that the prediction based on only a small population of the AN

model fibers, which have characteristic frequencies near the signal frequency, can

account for the results of the hearing experiments discussed in this project. The trends of

the predicted thresholds based on the coincidence-detection mechanism are similar to

those based directly on the discharge patterns of AN fibers.

Table of Contents

List of Figures

CHAPTER 1 INTRODUCTION

1.1 Background

The cues used by listeners to detect spectral changes in vowels have been studied for many years. However, both the cues embedded in vowel signals and the processing mechanisms used by the human auditory system are still not completely clear. Formant frequencies are important characteristics of vowel signals and estimating the resolving ability of the auditory system for formant-frequency changes is the first step in understanding speech-signal processing in the auditory system. Many psychophysical studies of human listeners have addressed this task. For example, human performance in frequency discrimination of band-limited sounds including band-limited harmonic complexes (Lyzenga and Horst, 1995) and synthetic vowels (Flanagan, 1995, Mermelstein, 1978 and Kewley-Port and Watson, 1994) have been studied in a number of investigations.

Lyzenga and Horst (1995) measured human subjects' just noticeable differences (JND) in the center frequency of band-limited harmonic complexes, which are a convenient simplification of synthetic vowel signals. The threshold of center-frequency discrimination of a harmonic complex with a triangular spectral envelope has a concave shape in the tested frequency range, i.e. the lowest threshold is observed when the center-frequency is half-way between two harmonic components. For the center-frequency discrimination task with a trapezoidal spectrum, the thresholds at multiples of the fundamental frequency (F0, 100 Hz) are lower than the thresholds at frequencies in

between multiples of F0. These experiment results cannot be explained by the changes of the total energy of the signal from trial to trial (Lyzenga and Horst, 1995).

In the same study, JNDs for the same task but with a randomly roving signal level were measured (Lyzenga and Horst, 1995). This roving-level condition made it impossible for the subject to use signal energy as a cue to detect frequency change. The performances for the conditions with and without the roving signal level showed similar trends, although the JNDs for the non-roving condition are lower than those for the roving condition. The ratio of roving vs. non-roving JNDs (keeping all the other parameters the same) is about 1.5 in most cases (Lyzenga and Horst, 1995). This suggests that the auditory system takes advantage of some information coded in the signal other than the total energy, or the level cue, of the stimuli.

A more complicated signal, the synthesized vowel, was also included in the simulations and analyses of this study. An important goal of this project was to test the conclusions from the predictions related to simple signals (harmonic complexes) with the predictions related to more complicated signals (synthesized vowels). In addition, this study further extended the quantitative analysis to the simulation and understanding of the decoding scheme in the auditory periphery when a background noise is present.

Formant-frequency discrimination thresholds of cat have been measured by Hienz *et al.* (1998). They generated synthesized vowels with a cascade speech synthesizer (Klatt, 1980) and reported that the thresholds at low and medium noise levels (signal-to-noise ratios at 23 and 13 dB, respectively) are similar to the threshold in quiet, while the threshold at a high noise level (signal-to-noise ratio at 3 dB) is significantly higher than

the threshold in quiet. In this study, model performance based on the information in the discharge patterns of AN model fibers or monaural cross-frequency coincidence detectors was calculated and the thresholds were compared to cat performance.

1.2 Coding mechanisms discussed in this study

A study of what coding mechanisms are used by the peripheral auditory system is important for understanding how speech signals are encoded in auditory system. The purpose of this project is to explore the cues used by the auditory system in formant-frequency discrimination tasks. The limits on formant-frequency discrimination performance imposed by the random nature of AN discharges are estimated by using a computational AN model. The study of Siebert (1965) provided a general approach of using a combination of analytical models of the peripheral auditory system with an ideal central processor to predict human performance limits in psychophysical tasks. This approach assumes that the discharge pattern of the AN population is used by an ideal central processor to discriminate changes in stimulus parameters. By employing methods from the theory of statistical hypothesis testing, the discrimination ability of this central processor can be estimated. Colburn (1969, 1973, 1977) extended this approach for binaural psychophysical studies with a simple interaural coincidence-counter mechanism. Heinz *et al.* (2001b) adopted both the ideal processor mechanism and the cross-fiber coincidence counter mechanism and employed a computational AN model in his study of monaural psychophysics. Both Colburn (1969, 1973, 1977) and Heinz *et al.* (2001b) only considered the counts of the coincidence detectors.

3

The project proposed here used both the optimal processor mechanism and the coincidence detection mechanism to predict human performance limits in frequency-discrimination tasks with vowel-related signals. A new analysis for the coincidence-detection mechanism was also developed in this project. This new analysis uses not only the count information but also the temporal information encoded in the responses of the coincidence detectors. In our study, a computational AN model (Tan, 2000, Appendix A) is used to process the band-limited harmonic-complex signal and synthetic-speech stimuli. The results for the harmonic complexes will be compared with human performance (Lyzenga and Horst, 1995). The frequency-discrimination limits with synthetic vowels based on the same computational AN model will be compared with the experimental results of Hienz *et al.* (1998).

It is important to keep in mind that this study was not trying to build the mechanism that can achieve best performance (i.e., lowest threshold); instead, the goal of this study was to understand what coding mechanism is used by the auditory periphery in the psychophysical experiments described above. Thus, more attention was paid to the trends than to the absolute values of the thresholds. In general, the model thresholds were in general lower than human performance, but the model thresholds could easily be elevated by assuming that fewer AN model fibers were engaged in the task.

1.3 Features of the AN model of special interest in this study

The simulations of AN responses in this project were based on a nonlinear computational AN model (Tan, 2000; Appendix A). The output of this AN model is

4

designed to simulate the discharge rate of an AN fiber responding to an arbitrary sound stimulus.

One important feature of this AN model was the nonlinear compression in its response property. For many hearing-impaired listeners, hearing loss is related to the damage of the functioning of the outer hair cells (OHCs) (Patuzzi et al., 1989, Ruggero and Rich, 1991), which play an important role in the active nonlinear mechanism of the cochlea. One goal of this project is to explore the effect of the active nonlinearity of the cochlea and the associated nonlinear phase property on encoding and processing of speech signals and the potential benefit of this nonlinear mechanism.

Another focus of this AN model was to included the instantaneous frequency (IF) glide in its reverse-correlation (revcor) function. This IF glide has been observed in reverse-correlation functions of the mammalian auditory periphery (Recio *et al.*, 1997; de Boer and Nuttall, 1997, Carney et al., 1999). Carney *et al*. (1999) showed that revcor functions have different frequency glide patterns for AN fibers with different characteristic frequencies. For CFs higher than 1500 Hz, the glide has an upward trend, i.e. from low frequency to high frequency. For CFs between 750 Hz and 1500 Hz, the instantaneous frequency is nearly constant. For CFs lower than 750 Hz, a downward trend is observed. This IF glide affects the temporal pattern of AN discharges, which may be important in encoding speech signals. In the AN model, this IF glide is achieved by the manipulation of the pole positions in control space (see methods section in Appendix A for more detail).

Both the nonlinear compression property and the IF glide in the revcor function can be "turned off" by manipulating the parameters of the model. The threshold predictions based on a linear model (i.e., without the nonlinear compression) and a model without the IF glide in its revcor function were computed and compared with the performances of the original model with the same set of psychophysical tasks and no significant differences were found (see Chapter 5 for more detail).

1.4 Overview of the dissertation document

Chapters 2-4 of this document were designed in the format of a manuscript. Thus, each chapter has an individual introduction, method, result and discussion section.

The second chapter of this document describes the limit of an ideal central processor's performance assuming that the central processor could optimally use the information in the response patterns of an AN model population. The simulation work was focused on the AN model population's responses to harmonic complexes. Statistical methods (Siebert, 1965; Colburn, 1969, 1973; Heinz *et. al.*, 2001a) were employed to estimate the central processor's ability to discriminate the center-frequency change in the harmonic complex.

The third chapter also concerns the simulation of the same set of psychophysical tasks as in the second chapter, except that the prediction was computed by the coincidence-detection mechanism, either based on the count information or both count and temporal information (fine structure of the response) of the coincidence-detector output.

The fourth chapter extends the study to the simulation of psychophysical results based on synthesized-speech signals. The trends of the degradation in performance as background-noise level increases were compared for the two decoding mechanisms described in the previous two chapters: one is based on the AN model fiber response patterns and the other is based on the coincidence detector response patterns. For both decoding mechanisms, the predictions are performed first with the rate/count information approach and then with the approach based on both rate/count and temporal information.

The fifth chapter serves as the general discussion for this study. Limitations and future work are discussed.

Chapter 2 Harmonic-Complex Center-Frequency Discrimination Ability of AN Model Population

2.1 Introduction

Although many psychophysical studies related to speech recognition have been done, it is still not clear how speech signals are processed and encoded in the auditory system. The purpose of this chapter is to explore the cues used by the auditory system in formant-frequency discrimination tasks using simple harmonic-complex stimuli. A better understanding of speech-signal processing in the auditory periphery will benefit the development of artificial processors for speech signals and the design of hearing aids.

Formant frequencies are important features of the spectrum of speech signals, especially vowels. The values of formant frequencies characterize the basic shape of the speech spectrum and are important for phonetic identification (Rabiner and Schafer, 1978). Psychophysical experiments have been done to estimate the formant-frequency discrimination ability of human subjects (Flanagan, 1955; Mermelstein, 1978; Sinnott and Kreiter, 1991; Kewley-Port and Watson, 1994). However, different thresholds of the formant-frequency discrimination tasks were recorded in different studies due to the complexity of the stimuli and the differences in the experimental settings. For example, Mermelstein (1978) found the threshold for discriminating changes in the first formant at 350 Hz was 50 Hz, much higher than the result of Flanagan (1955), who reported that the thresholds for the first formant (at 300 Hz) was 12 to 17 Hz.

Harmonic complexes, having either a triangular shaped or a trapezoidal shaped spectrum (Fig. 2-1) were proposed as a convenient simplification of vowel signals by Lyzenga and Horst (1995). For the triangular spectral envelope, the threshold of the center-frequency discrimination as a function of the center frequency has a concave shape in the frequency range from 2000 Hz to 2100 Hz [Fig. 2-2(a)], i.e. the lowest threshold occurs for a center frequency at 2050 Hz. The lowest and highest thresholds for the triangular spectrum with a slope of 400 dB/oct are about 1 Hz (0.05% of 2050 Hz) and 6 Hz (0.3% of 2100 Hz), respectively. For the trapezoidal spectral envelope, the threshold of center frequency has an M-shaped curve from 2000 Hz to 2200 Hz [Fig. 2-2(b)], i.e. the thresholds at multiples of the fundamental frequency (F0, 100 Hz) are lower than the thresholds at frequencies in between the multiples of F0. The lowest and highest thresholds for the trapezoidal spectrum with a slope of 400 dB/oct are about 1.74 Hz (0.087% of 2000 Hz) and 24.3 Hz (1.2% of 2025 Hz), respectively.

To explain their results, Lyzenga and Horst (1995) examined various cues and decoding mechanisms possibly used by the subjects, including a profile comparison model and an explanation based on amplitude-modulation detection thresholds. They proposed an excitation-profile comparison model, which can roughly predict the lowest threshold for the triangular spectrum and the lowest threshold for the trapezoidal spectrum. However their explanation requires a physiologically unreasonable assumption that the excitation difference includes only negative values for results of the trapezoidal envelope while both positive and negative values are included for the triangular envelope

(Fig. 10 of Lyzenga and Horst, 1995). They also suggested that the results for the
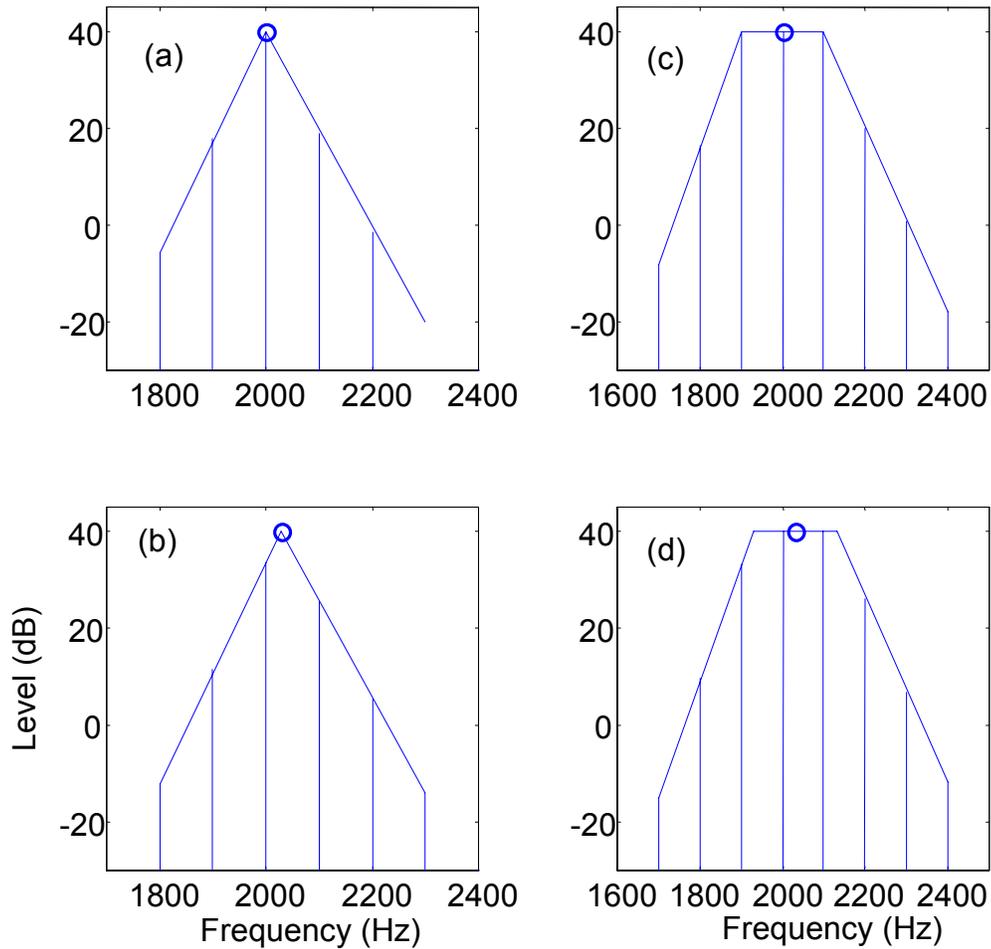
Figure 2-1 Examples of the harmonic-complex spectra, with a triangular envelope (a and b) or a trapezoidal envelope (c and d). All the spectra have a fundamental frequency of 100 Hz. The bold circles on top of each spectrum indicate the center frequency. The spectra on the top row are centered at 2000 Hz while the spectra on the bottom row are centered at 2020 Hz (shifted 20 Hz away from 2000 Hz).

(a) Triangular spectrum   (b) Trapezoidal spectrum
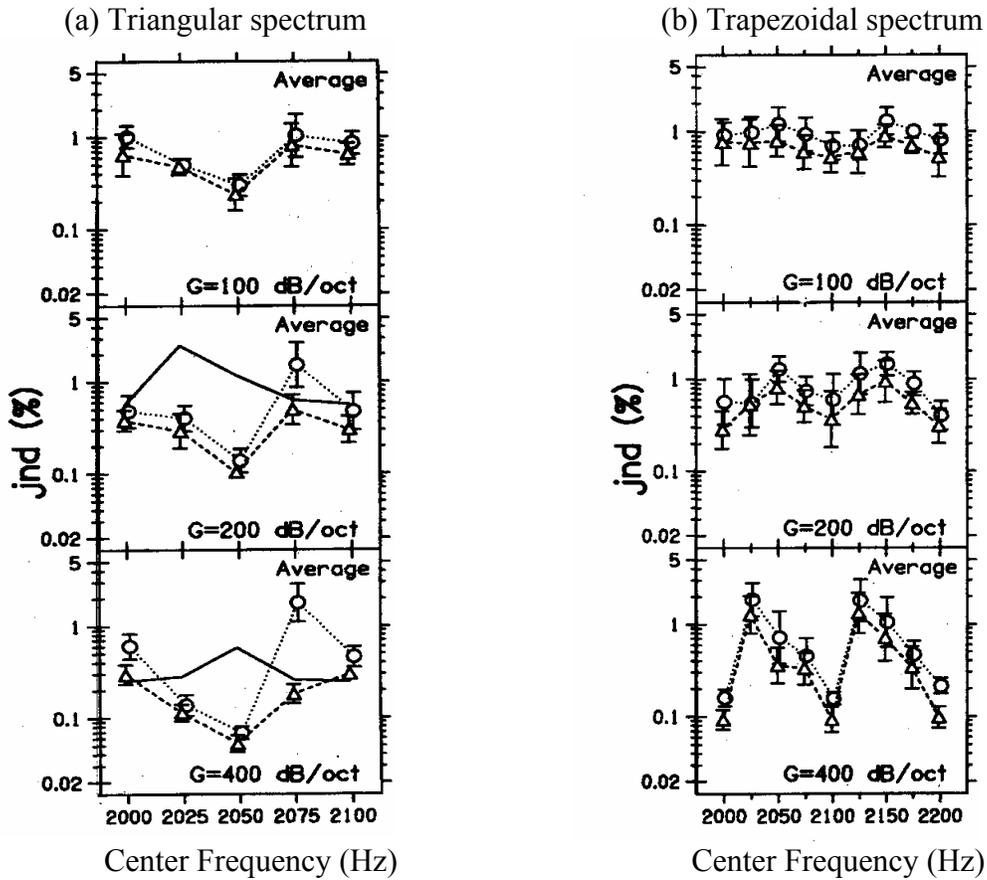
Figure 2-2. Human threshold for center-frequency discrimination of the harmonic-complexes (from Lyzenga and Horst, 1995) with: (a) triangular spectrum envelope and (b) trapezoidal spectrum envelope. Each row corresponds to the thresholds for a different slope of the harmonic envelope: 100, 200, and 400 dB/oct from top to bottom. The solid lines are predictions based on the change in the overall energy of the stimuli. The dashed lines with triangle signs are human performance for the experiments described in text. The dotted lines with circles are human performance with random roving levels of the stimuli.

triangular spectrum can be partially explained by sensitivity to amplitude-modulation depth (Fig. 9 of Lyzenga and Horst, 1995). However Lyzenga and Horst (1995) did not explain the trends of the thresholds as a function of the center frequency for the triangular spectrum and the trapezoidal spectrum.

Lyzenga and Horst (1997) extended their earlier study and concluded that in the frequency region near 2000 Hz, phase cues influence the threshold. They suggested that when the fundamental frequency is 100 Hz, there are three harmonic components falling into one critical band (roughly 250 Hz wide) and therefore that the excitation-profile model cannot explain the data due to its insensitivity to the relative phase relations of the harmonic components. They also calculated the envelope-weighted or intensity-weighted averaged instantaneous frequency (EWAIF or IWAIF; Feth, 1974) and concluded that EWAIF and IWAIF showed little correspondence with the psychophysical data. Additionally Lyzenga and Horst (1997) pointed out that the peaks in the second-order derivative of the signal's temporal envelope with the triangular spectrum was clearly influenced by the center frequency of the harmonic complex, as well as by the phase relation of the harmonic components. This indicates the potential importance of temporal cues in predicting the performance.

In the study presented here, the limit on center-frequency discrimination performance was estimated based on the response patterns of a computational AN model. A general approach was proposed by Siebert (1965) using a combination of an analytical model of the peripheral auditory system and an ideal central processor to predict human performance limits in psychophysical tasks. The discrimination ability of this ideal

central processor can be estimated using methods from the theory of statistical hypothesis testing. More recently, Heinz (2000) adopted the ideal-processor mechanism and employed a computational AN model in his study of monaural level and frequency discrimination.

The study presented here used the optimal processor mechanism and compared the predictions based on only the average rate of the AN responses with the predictions based on both the average rate and the fine structure of the AN response patterns (i.e., the timing information). In our study, a computational AN model (Tan, 2000; Appendix A) was used to process the band-limited harmonic-complex signals. The frequency-discrimination ability of an ideal central processor based on the population model response was computed. The results were compared with human performance (Lyzenga and Horst, 1995).

2.2 Methods

2.2.1 Stimuli

Two center-frequency discrimination experiments by Lyzenga and Horst (1995) were simulated using band-limited harmonic complexes with a fundamental frequency of 100 Hz as stimuli (Fig. 2-1). Stimulus parameters were the shape (triangle or trapezoid), the slope (100, 200 or 400 dB/oct), and the center frequency (varied from 2000 Hz to 2100 Hz for triangular envelope and from 2000 Hz to 2200 Hz for trapezoidal envelope) of the spectral envelope. In the first experiment, the spectral envelope was triangular on a

log-log scale [Fig. 2-1(a) and (b)]. In the second experiment, the spectral envelope was trapezoidal with a 200 Hz-wide constant-level plateau [Fig. 2-1(c) and (d)]. The fundamental frequency was 100 Hz and all frequency components of the complexes had a starting phase angle of zero degree. In each trial the signal had a duration of 250 ms, including a 25 ms onset and offset time shaped by a raised cosine.

The frequencies of the harmonic components (the vertical bars in Fig. 2-1) were kept the same from trial to trial. The task was to discriminate changes in the center-frequency (the circles in Fig. 2-1) of the envelope (the dashed lines in Fig. 2-1). The magnitudes of the harmonic components change as the center frequency of the spectrum envelope shifts to lower or higher frequencies. For example, the center frequency of the harmonic complex in Fig. 2-1 (a) is at 2000 Hz. If this center frequency is shifted to a frequency slightly higher than 2000 Hz, e.g. 2005 Hz, the magnitude of all the components with frequency higher than 2000 Hz will increase and the magnitude of all the components with frequency lower than 2000 Hz will decrease. If the center frequency decreases, the magnitudes of the components with frequencies lower than 2000 Hz will increase and the magnitudes of the components with frequencies higher than 2000 Hz will decrease.

The spectrum of a harmonic-complex signal usually has five to eight harmonic components. In order to understand better the features in the harmonic complexes, as well as the predicted performance based on the AN population model, it is useful to consider simple signals with a reduced number of components. This simplification can also make the mathematical analysis easier. Figure 2-3 demonstrates such a simplification for the

14
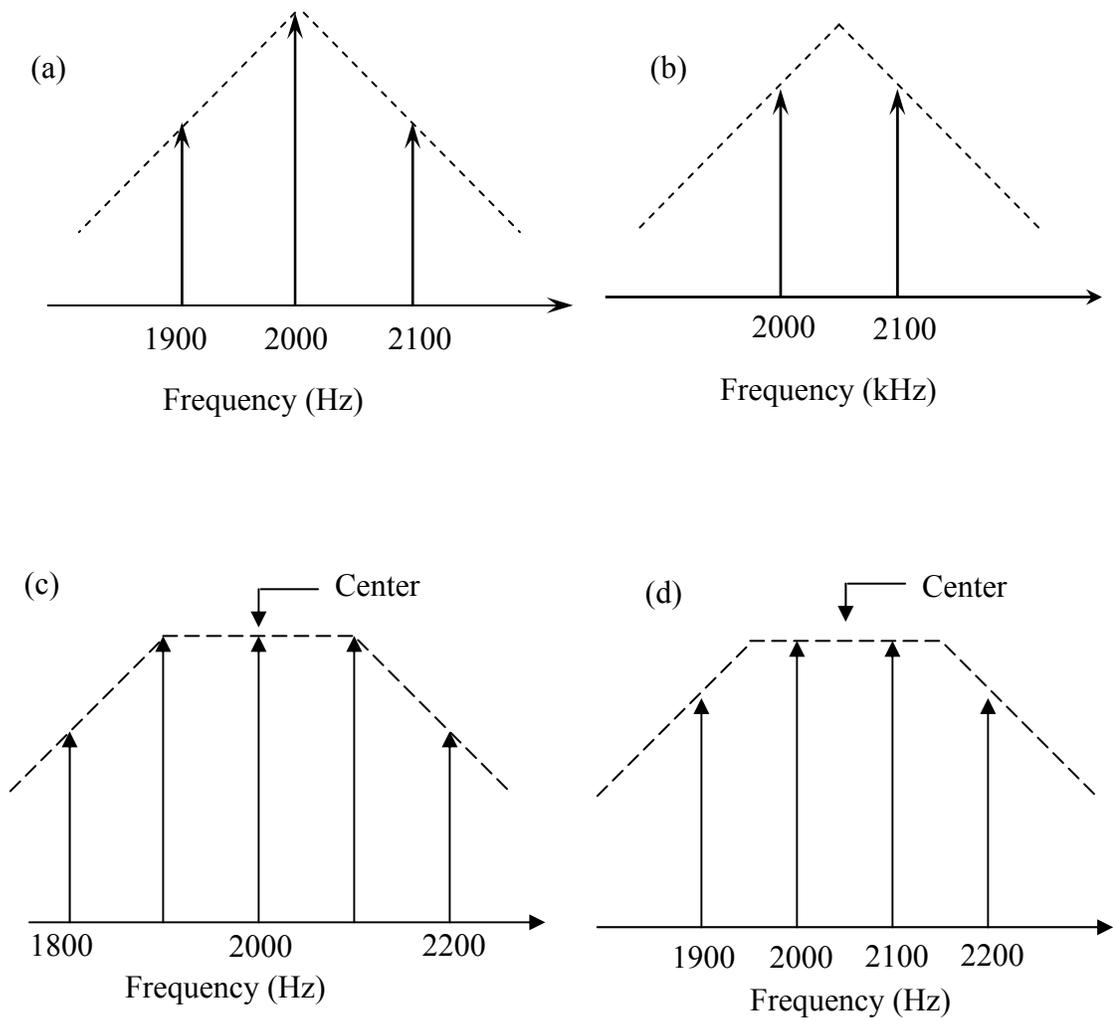
Figure 2-3. Simplified harmonic-complex signals: (a) triangular envelope with center frequency at 2000 Hz (b) triangular envelope with center frequency at 2050 Hz (c) trapezoidal envelope with center frequency at 2000 Hz (d) trapezoidal envelope with center frequency at 2050 Hz.

signal with triangular spectrum (the upper row) and for the signal with trapezoidal

spectrum (the bottom row). For the triangular spectrum with center frequency at 2050 Hz

[Fig. 2-3 (b)], two harmonic components, which are closest to the center of the envelope,

are kept and all the other components, which are relatively far away from the center

frequency, are ignored. In this case, the signal is the combination of two sinusoids with

the same amplitude in the time domain. This combination of signals can be transformed

to a sinusoidal signal modulated by a cosine (Equation 2.1).

$$\sin(2\pi * f_1 * t) + \sin(2\pi * f_2 * t) = 2 * \underbrace{\sin(\frac{2\pi * (f_1 + f_2) * t}{2})}_{\text{Carrier}} * \underbrace{\cos(\frac{2\pi * (f_1 - f_2) * t}{2})}_{\text{Modulator}}$$

$$(2.1)$$

The cosine modulator serves as the envelope of the signal. An interesting feature

of this simplified signal is that at the zero-crossing point of this cosine signal, i.e. when

the cosine signal changes from positive to negative or from negative to positive, there is a

180-degree phase change in the harmonic complex.

Figure 2-3 (a) shows the simplified signal when the center frequency is 2000 Hz:

three harmonic components are kept in this case. Because of the existence of the center

component, the combination of these three harmonic components does not show the 180-

degree phase change in the time domain. As described in the following sections, the

presence or absence of this 180-degree phase shift can explain the threshold difference

for different center frequencies.
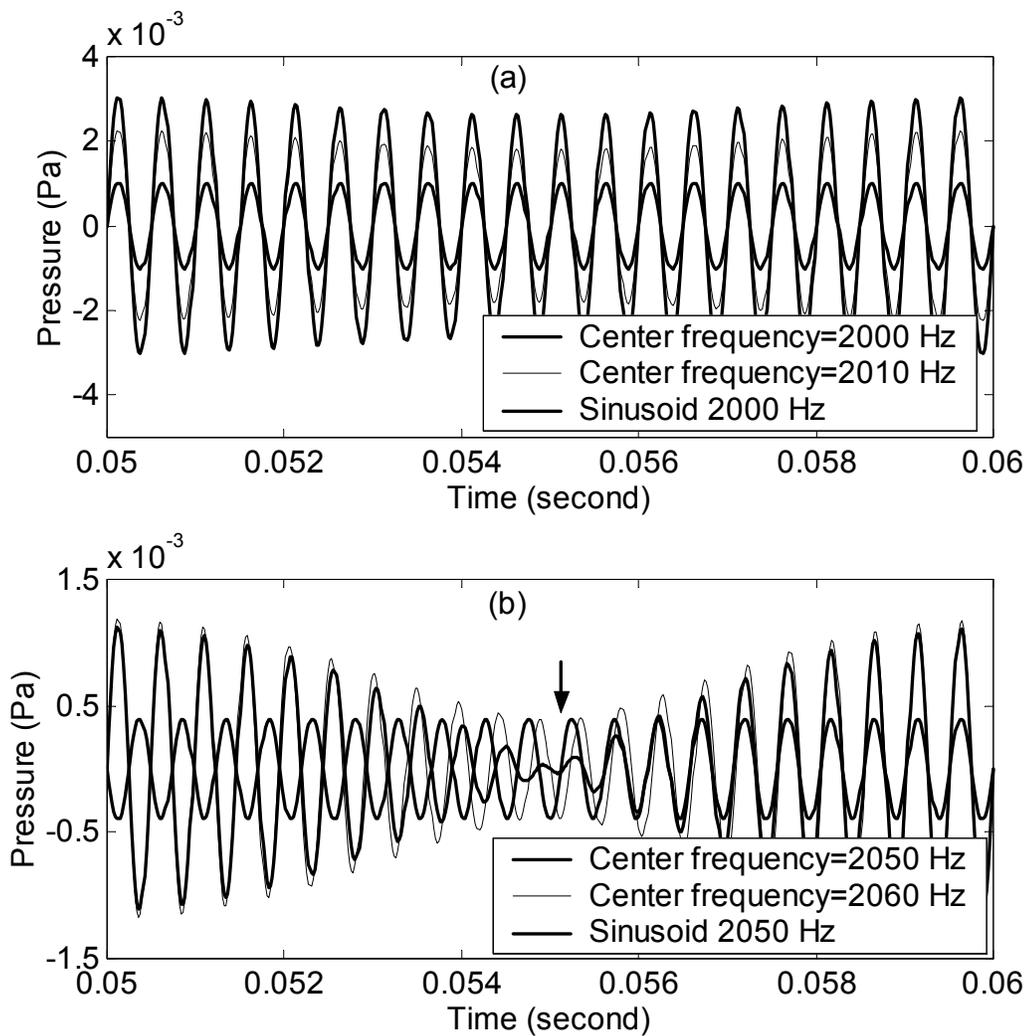
Figure 2-4 Simplified harmonic-complex signals for triangular spectrum in time domain. The upper panel corresponds to the center frequency of 2000 Hz and the lower panel corresponds to the center frequency of 2050 Hz. In each panel, the thick solid line illustrates a signal without the center frequency and the thin solid line illustrates a signal with a 10 Hz center-frequency shift. The dotted lines are reference sinusoidal signals of 2000 Hz (upper panel) and 2050 Hz (lower panel).

Figure 2-4 shows the simplified signals in the time domain. In Fig. 2-4(a), the thick solid line is the simplified signal with center frequency at 2000 Hz [corresponding to the spectrum in Fig. 2-3 (a)] and the thin solid line is the simplified signal with center frequency at 2010 Hz (shifted 10 Hz from 2000 Hz). The dotted line is a reference sinusoidal signal (2000 Hz). It is clear that the result of the 10 Hz shift is primarily a magnitude change in the time domain.

In Fig. 2-4(b), the thick solid line is the simplified signal with center frequency at 2050 Hz [corresponding to the spectrum in Fig. 2-3 (a)] and the thin solid line is the simplified signal with center frequency at 2060 Hz (shifted 10 Hz from 2050 Hz). The dotted line in Fig. 2-4 (b) is a sinusoidal signal at 2050 Hz. By comparing the thick and the thin solid line with the dotted reference line in Fig. 2-4 (b), the 180-degree phase transition can be observed. On the right side of the marker (the downward arrow), the thick and the thin solid lines have the same phase as the dotted sinusoid signal. On the left side of the marker, the thick and the thin solid lines have a 180-degree phase difference from the dotted reference line. The phase transition in the thick solid line is slightly different from that in the thin solid line. The thin solid line has a relatively slower phase shift: the phase shift in the thin solid line starts earlier and ends later than in the thick solid line. This difference in the phase transient provides information for center-frequency discrimination if we assume the AN response phase-locks to the fine structure of the sound stimulus.

The same simplification strategy was applied to the stimuli with the trapezoidal spectrum [Fig. 2-3 (c, d)], except that a larger number of components were kept in the

18

simplified signal for the trapezoidal spectrum than for the triangular spectrum. The central five and four components were kept while the other components were removed for center frequency at 2000 Hz [Fig. 2-3 (c)] and 2050 Hz [Fig. 2-3 (d)], respectively. Figure 2-5 (a) shows the simplified signals with center frequency at 2000 Hz (thick solid line) and 2010 Hz (thin solid line, with a 10 Hz deviation from 2000 Hz) in the time domain. The arrows in Fig. 2-5 (a) indicate two sudden 180-degree phase reversals in the thick solid line while the phase reversals in the thin solid line are relatively smooth. Figure 2-5 (b) shows the simplified signal with a center frequency of 2050 Hz (half-way between two harmonic components). A 180-degree phase reversal can also be seen in this case. However there are more phase-reversal cues in Fig. 2-5(a) than in Fig. 2-5(b): For the same period of time, there are more reversals in Fig. 2-5(a) (two times between 0.05 and 0.06 second) than in Fig. 2-5(b) (only once between 0.05 and 0.06 second). This is a potential reason for the relatively lower discrimination threshold at 2000 Hz as compared to 2050 Hz.

This simplification method was used only for the purpose of observing the cues in the signal. The threshold predictions were all based on the original signals, unless specifically mentioned.
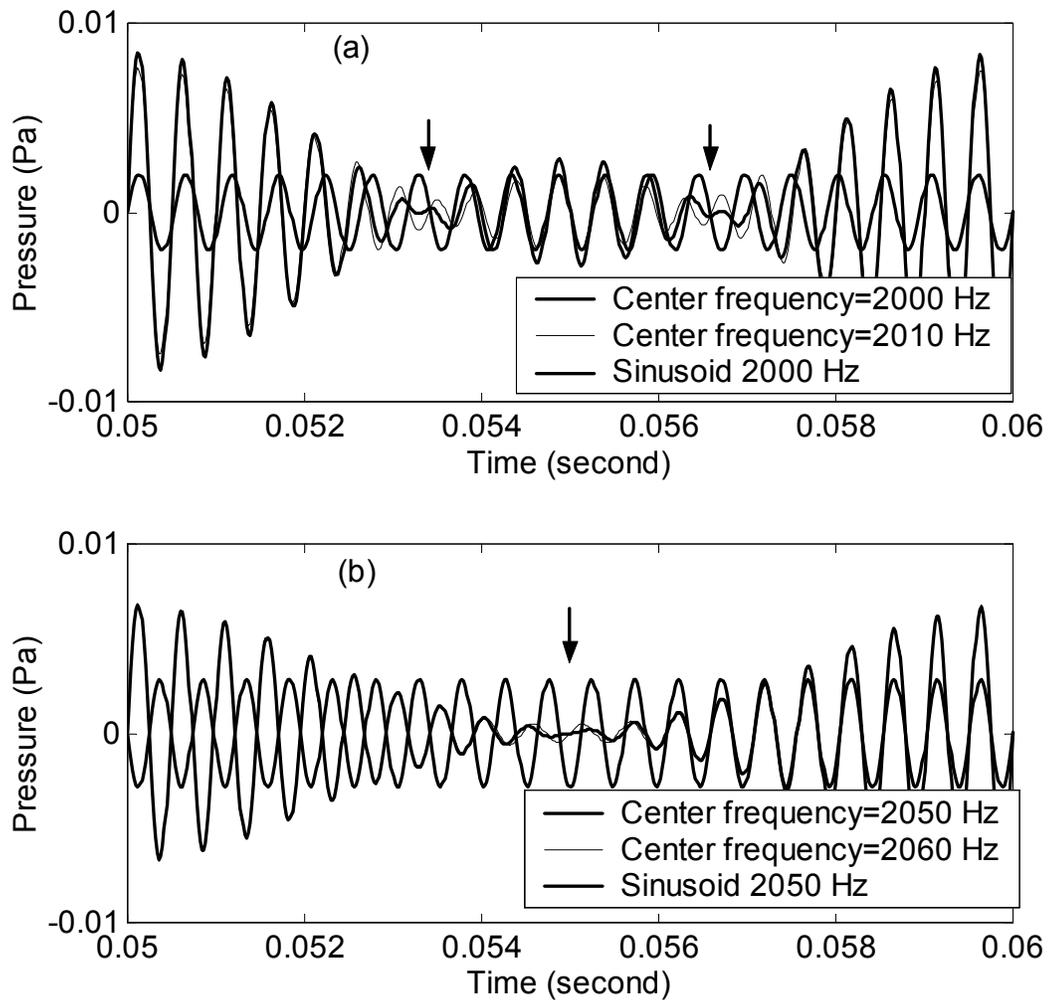
Figure 2-5 Simplified harmonic-complex signals for trapezoidal spectrum in time domain. The upper panel corresponds to the center frequency of 2000 Hz and the lower panel corresponds to the center frequency of 2050 Hz. In each panel, the thick solid line illustrates a signal without the center frequency and the thin solid line illustrates a signal with a 10 Hz center-frequency shift. The dotted lines are reference sinusoidal signals of 2000 Hz (upper panel) and 2050 Hz (lower panel).

2.2.2 The Nonlinear AN Model

A nonlinear computational auditory-nerve model was developed as part of the preliminary work for this project (Tan, 2000; Appendix A). The block diagram in Fig. 2-6 shows the basic components of the model. The model consists of a time-varying band-pass filter as the signal path and a nonlinear feed-forward control path. The control path changes the bandwidth and gain of the signal path instantaneously and therefore the model response has a nonlinear compression property. Due to the configuration of the locations of the poles and zeros in the band-pass filter of the signal path, the reverse-correlation function of this model's response to broadband noise contains instantaneous frequency glides similar to those reported for AN fibers (Carney *et al.*, 1999). The output of this model is the time-varying discharge rate that simulates the auditory-nerve fiber response to arbitrary sounds. This model produces realistic response features to various stimuli, including pure tones, two-tone combinations, wide-band noise and clicks.

A total of 30,000 AN fibers in human (Rasmussen, 1940) are assumed to have CFs evenly distributed on a log scale from 20 Hz to 20,000 Hz which is a simplified version of the human cochlear map of Greenwood (1990). All the calculations of the computational AN model presented here are based on 50 model AN fibers whose CFs are evenly distributed on a log scale, from 1500 Hz to 3000 Hz (CFs beyond this range are not considered for efficiency in computation). Thus, these 50 models roughly represent 10% of the 30,000 total AN fibers ($\frac{\log(20000/20)}{\log(3000/1500)} \times 100\% = 10\%$). Each of the 50 AN fiber models represents about 60 independent AN fibers, for a total of 3000 fibers in the
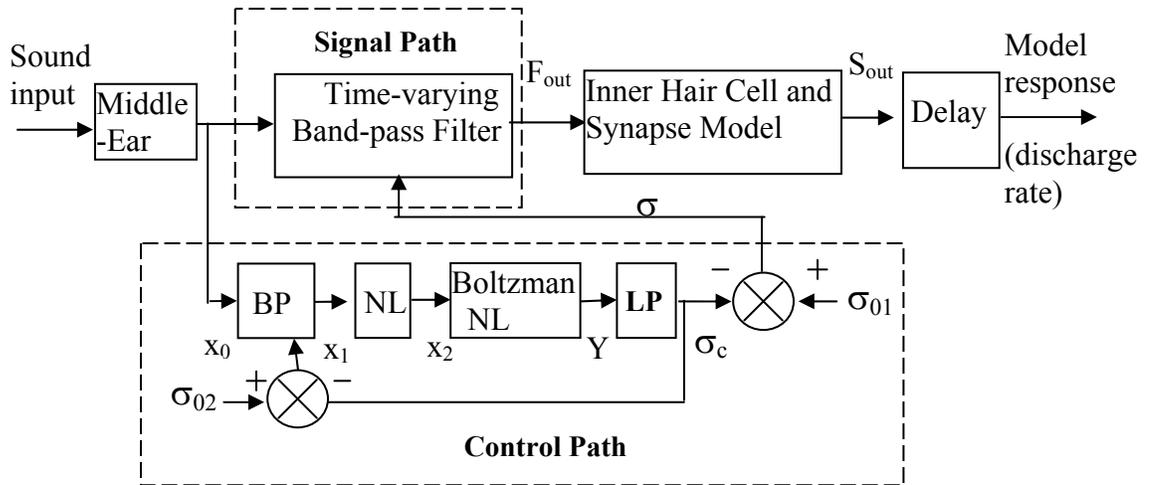
Figure 2-6 Schematic diagram of the auditory-nerve model (Tan 2000; Appendix A). The model includes a signal path, a control path, the inner hair cell and synapse model and a time delay.

1500-3000 Hz range. The random nature of the AN discharges is assumed to be Poisson. The predictions [just-noticeable-difference (JND) of center frequency] are made based on the assumption that the observations in the AN-model response is a set of independent non-stationary Poisson processes. These assumptions for AN models are used for all of the predictions presented here.

2.2.3 Statistical methods

Predictions of model performance in this study are based on statistical methods. There are two general approaches used in this study. The first one uses the discharge patterns of the AN model population to calculate the model performance. In the second approach, a coincidence detection mechanism is used and the discharge patterns of a population of model coincidence cells are used. This chapter is focused on the first approach, where an ideal central processor is assumed to optimally use the information encoded in the response pattern of each AN model fiber and the performance of this central processor is estimated. In a later chapter, the performance of the coincidence-detection mechanism will be discussed.

The AN discharge activity can be described as a Poisson random process. The bound on the variance of the estimate of a variable can be described by the Cramér-Rao bound (Cramér, 1951; van Trees, 1968). The variance σ_i of the estimate of any signal parameter (e.g. Fc, the center frequency of the harmonic complex) based on the observation from the i-th AN fiber can be estimated by (Siebert, 1965):

$$\frac{1}{\sigma_i^2} \leq \int_0^T \frac{1}{r_i(t)} \left[\frac{\partial r_i(t)}{\partial F_c} \right]^2 dt \tag{2.2}$$

where $r_i(t)$ is the ith AN fiber's instantaneous discharge rate.

Equation 2.2 represents the relative sensitivity of the i-th AN fiber to the center frequency change of the signal. By assuming that the discharge patterns of all AN fibers are statistically independent (Johnson and Kiang, 1976), the bound of the variance of the observation based on the AN population's response pattern is the summation of the variance's bound for each single AN fiber, i.e. $\frac{1}{\sigma_{all}^2} = \sum_i \frac{1}{\sigma_i^2}$. The just-noticeable

difference (JND) of the ideal central processor corresponding to $d' = \frac{F_{cJND}}{\sqrt{\sigma_{all}^2}} = 1$ can be

computed by (Siebert, 1965):

$$F_{cJND} = \left[\frac{1}{\sum_i \frac{1}{\sigma_i^2}} \right]^{\frac{1}{2}} = \left[\frac{1}{\sum_i \int_0^T \frac{1}{r_i(t)} \left[\frac{\partial r_i(t)}{\partial F_c} \right]^2 dt} \right]^{\frac{1}{2}} \tag{2.3}$$

If only the average-rate information of the AN model responses is used, equation 2.3 can be simplified to be:

$$F_{cJND} = \left[\frac{1}{\sum_i \frac{1}{\sigma_i^2}} \right]^{\frac{1}{2}} = \left[\frac{1}{\sum_i \frac{1}{Y_i} \left[\frac{\partial Y_i}{\partial F_c} \right]^2} \right]^{\frac{1}{2}} \tag{2.4}$$

24

where $Y_i = \int_0^T r_i(t)dt$ is the expected number of spikes (representing the average-rate information) from the i-th model AN fiber in one trial and T is the duration of one trial in the psychophysical experiment.

The calculation of the partial derivatives was approximated by calculating the ratio of the changes in the response due to a small change in the center frequency of the signal and the small change in the center frequency, e.g.,

$$\frac{\partial r_i(t)}{\partial f} \cong \frac{r_i(t \mid f + \Delta f) - r_i(t \mid f)}{\Delta f} \tag{2.5}$$

In this project, the approximation was computed using $\Delta f = 1$ Hz.

2.3 Results

2.3.1 Predictions for Signals with a Triangular Spectrum

Figure 2-7 shows the predictions of threshold for center-frequency discrimination for the harmonic complexes with triangular spectra. Center-frequency discrimination thresholds (JND of center frequency change) are plotted as functions of the center frequency of the spectrum envelope. Each panel corresponds to predictions for one value of the spectral slope. The lines with asterisks are predictions based on the combined rate and timing information of the AN model population response and the lines with circles
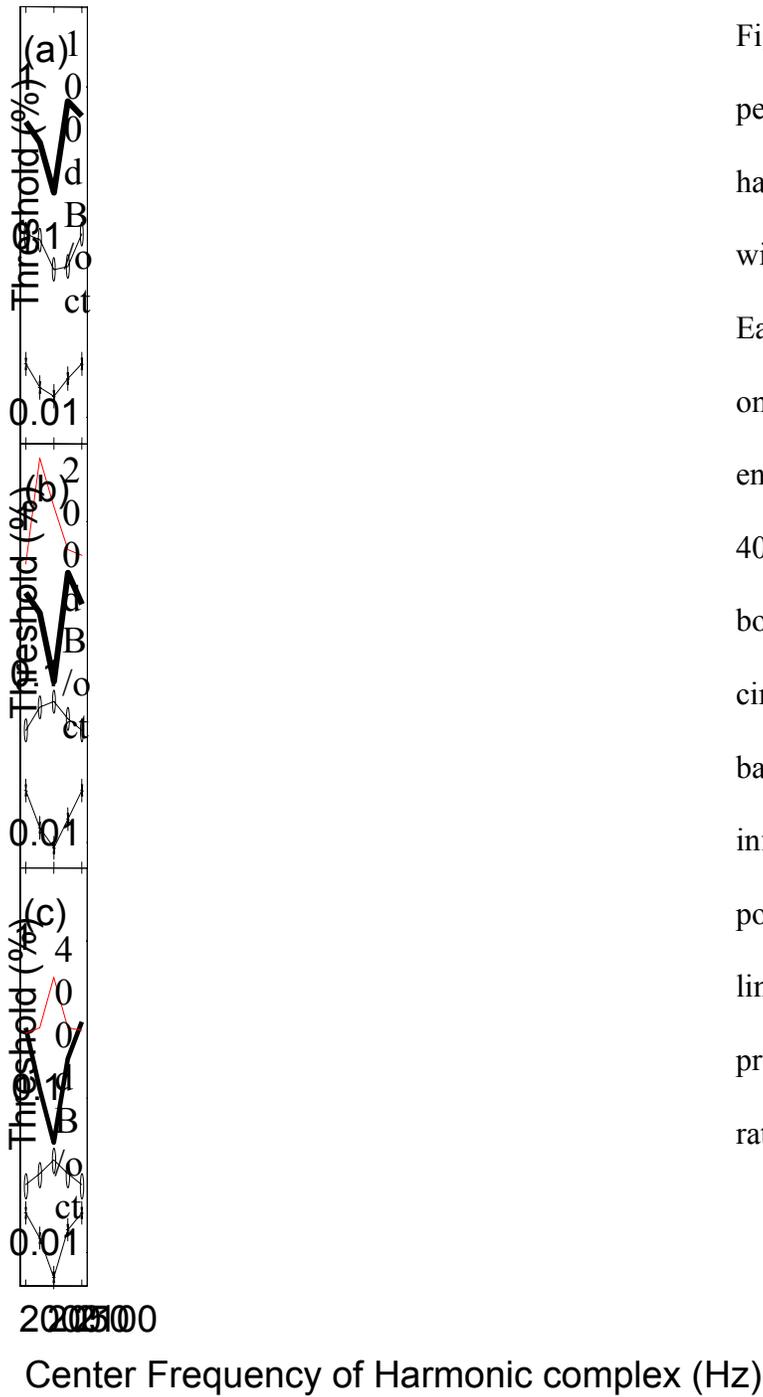
Figure 2-7 Model performances for harmonic-complex signals with triangular spectrum. Each panel corresponds to one slope of the spectrum envelope (100, 200, and 400 dB/oct from top to bottom). The lines with circles are predictions based on only average-rate information of the AN population responses. The lines with asterisks are predictions based on both rate and temporal

are based on only the rate information. The predictions based on the combination of rate

and timing information (asterisks) showed the desired concave shape for all three spectral

slopes. The prediction based on only the rate information showed a convex shape, which

is similar to Lyzenga and Horst's (1995) prediction based on the total energy of the

signals [solid lines in Fig. 2-2(a)].

How the model performance predicts the concave trend in the thresholds can be

revealed by observing the response patterns of the AN model population to the simplified

stimuli. Figure 2-8 shows the normalized sensitivity [in units of $1/(Hz)^2$] as a function of

model CF for the triangular spectrum with a slope of 400dB/oct. Each row corresponds to

one envelope center frequency (as indicated to the left of the figure). The normalized

sensitivity based on both rate and timing information (left column) is defined as

$\int_0^T \frac{1}{r_i(t)}[\frac{\partial r_i(t)}{\partial F_c}]dt$. The normalized sensitivity based on average-rate information (right

column) is defined as $\frac{1}{Y_i}\left[\frac{\partial Y_i}{\partial F_c}\right]^2$. The average-rate-only approach ignores the timing

information and therefore the normalized sensitivity based on average-rate is always

lower than the normalized sensitivity based on both rate and timing information. The

solid lines in all ten panels are computed with the original signals (before simplification).

The lines with asterisks are the results based on the simplified harmonic complex signals.

When the simplified signals are used (center frequency at 2000 Hz, 2050 Hz or 2100 Hz),

the results for the simplified signals overlapped with those for the original signals and
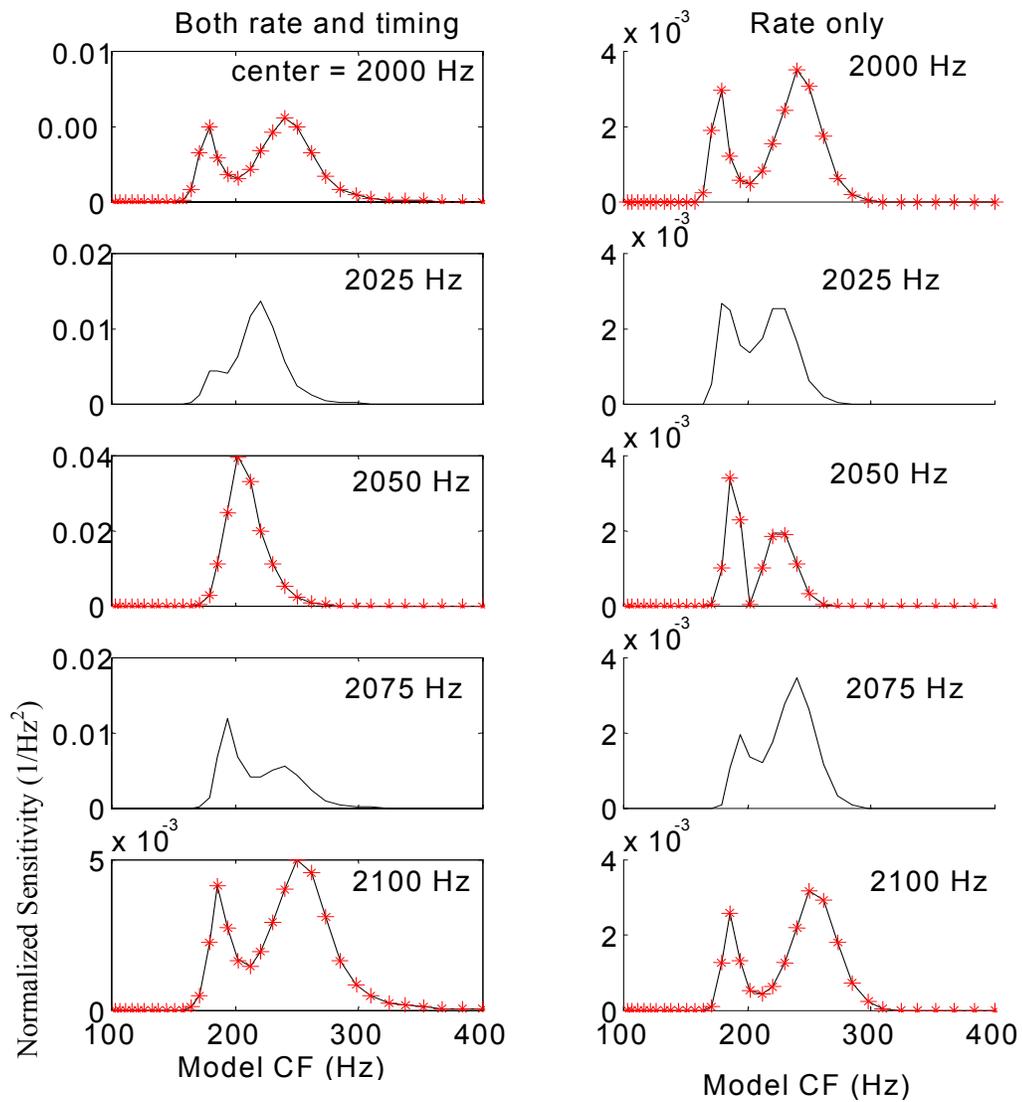
Figure 2-8 Sensitivity of model fibers to the changes of the harmonic-complex center frequency as a function of model CF. Each row corresponds to one center frequency, (2000 Hz to 2100 Hz). The solid lines are based on the original stimuli (for all center frequencies) and the asterisks are based on the simplified signals (for 2000 Hz, 2050 Hz, and 2100 Hz only).
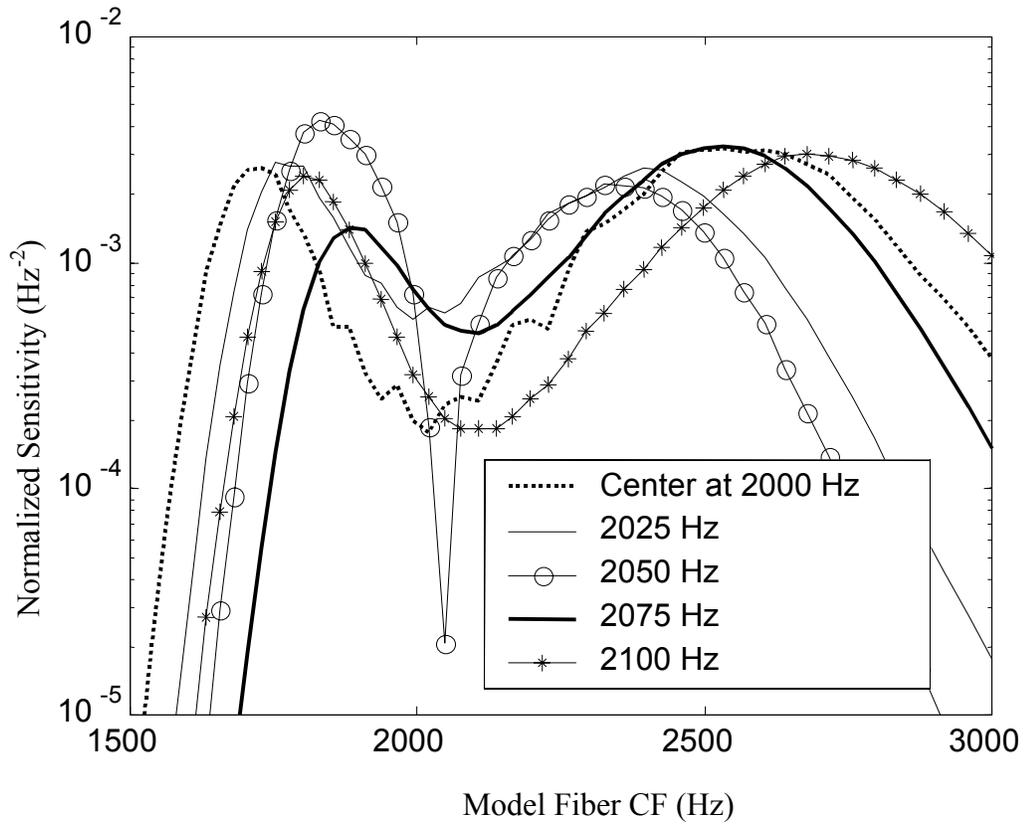
Figure 2-9 Normalized sensitivity patterns for triangular spectrum (not simplified) at various center frequencies (as marked in the figure) based on average-rate information of AN model responses.
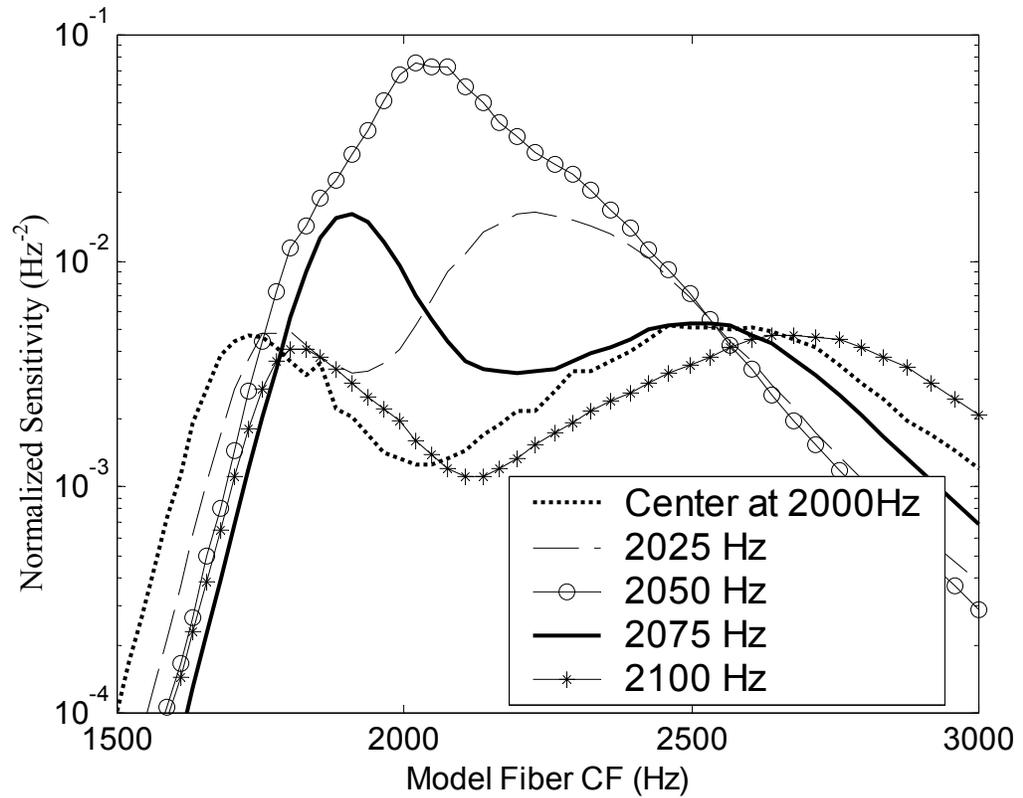
Figure 2-10 Normalized sensitivity patterns for triangular spectrum (not simplified) at various center frequencies (as marked in the figure) based on rate-and-timing information of AN model responses.

thus only one line can be seen in each of those panels. This overlap indicates that the predictions based on the simplified signal set would be the same as the predictions based on the original signals and therefore the simplified signals have the same cues (phase transitions) for the center-frequency discrimination as the cues in the original signals.

To compare the results across different center frequencies, all the sensitivity results are plotted together in Fig. 2-9 for the rate-only calculation and Fig. 2-10 for the rate-and-timing calculation. Different center frequencies are marked differently in Fig. 2-9 and Fig. 2-10. One interesting feature of this figure is that there is always a drop at about 2050 Hz for the results based on average-rate information (Fig. 2-9). However for results based on both rate and temporal information, the shape of the curve (Fig. 2-10) changes for different envelope center frequencies: some curves show a drop in threshold while others show a peak at the center frequency.

The sensitivity patterns in Fig. 2-9, based on average-rate information only, always show a threshold drop between 1900 Hz and 2200 Hz. The overall sensitivities (integral over CF) for the cases of center frequency at 2000 Hz (the dotted line) and 2100 Hz (the line with asterisks) are higher than that for center frequency 2050 Hz (the line with circles). Thus the threshold is higher for center frequency at 2050 Hz than at 2000 Hz and 2100 Hz, if only rate information is used.

In Fig. 2-10, due to the notch (reduction in threshold) of the dotted line (center frequency at 2000 Hz) and the line with asterisks (center frequency at 2100 Hz), the overall sensitivity for the stimulus with a center frequency right at a harmonic frequency (2000 Hz or 2100 Hz) is lower than the overall sensitivity forcenter frequency at 2050

31

Hz, where a peak is observed in the sensitivity pattern (the solid line with circles). Thus the threshold is lower at 2000 Hz and 2100 Hz than at 2050 Hz for the results based on both rate and timing information.

After explaining the general trends in the model performance using the results in Fig. 2-9 and Fig. 2-10, the next question is how to understand the difference in threshold predictions between center frequency at 2050 Hz as compared to a lower frequency (2000 Hz) and a higher frequency (2100 Hz). According to the above analysis, the difference in threshold based on both rate and timing information (solid lines in Fig. 2-7) comes from the difference between the lines in Fig. 2-10, especially the difference between the peak of the solid line with circles and the notch in the solid line with asterisks. However, the difference is reduced by the side bands in Fig 2-10. Thus it is reasonable to expect that predictions based on a smaller population of AN model fibers, centered at about 2050 Hz, would have a larger difference in threshold across center frequencies. This assumption is also feasible physiologically, i.e., it is reasonable to assume that the brain uses the information from a sub-group of AN fibers instead of all 30,000 fibers.

Figure 2-11 compares the results for a smaller population [using AN model fibers with CFs from 1900 Hz (the lowest possible frequency of the harmonic component in the simplified signal set) to 2200 Hz (the highest possible frequency of the harmonic component in the simplified signal set)] with the previous results (using the original model population with CFs from 1500 Hz to 3000 Hz). Figure 2-11 (a) is the model performance with the wider CF range (1500 Hz to 3000 Hz, same as the right column in Fig. 2-9) and Fig. 2-11 (b) is the model performance with smaller CF range. The

32
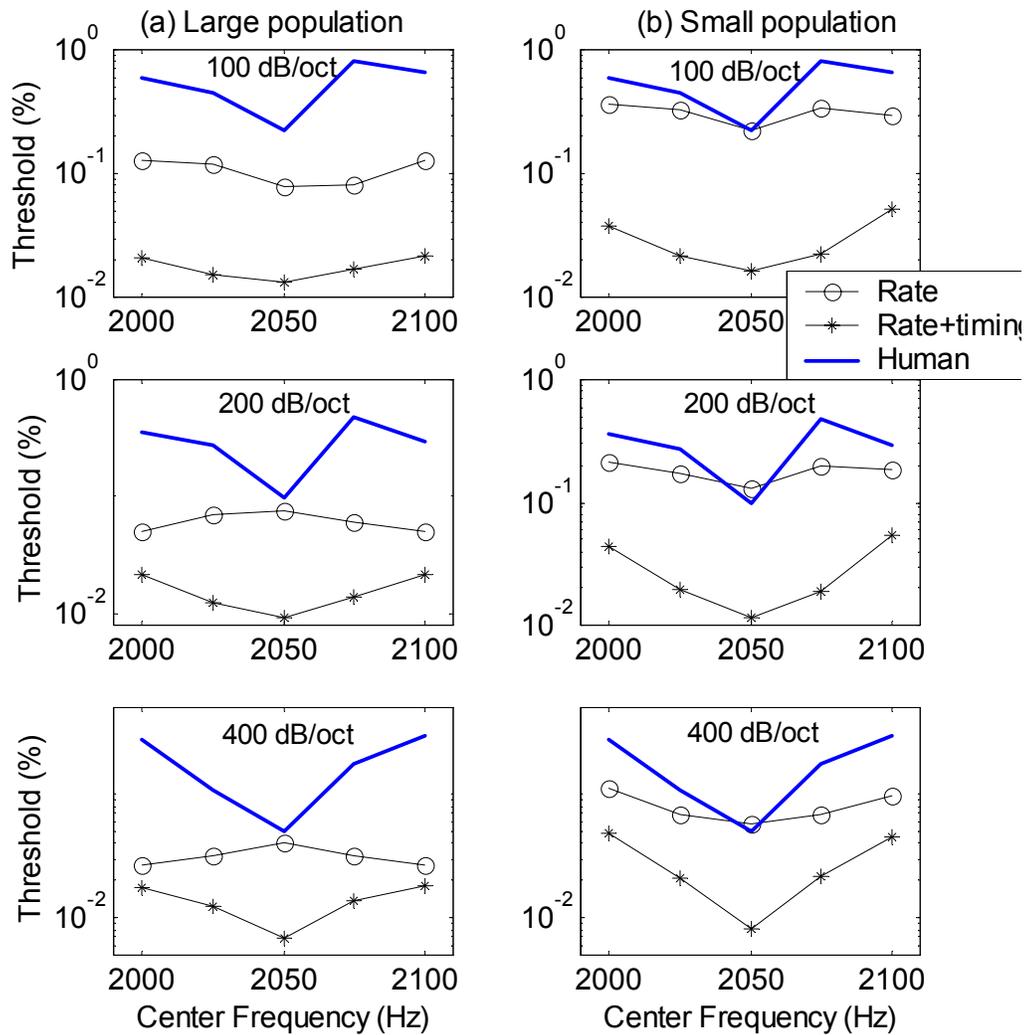
Figure 2-11 Comparing predictions based on (a) a larger population with CFs between 1500 Hz and 3000 Hz, and (b) a smaller population of AN model fibers with CFs between 1900 Hz and 2200 Hz. The rate prediction and the rate-and-timing prediction are marked by circles and stars, respectively.
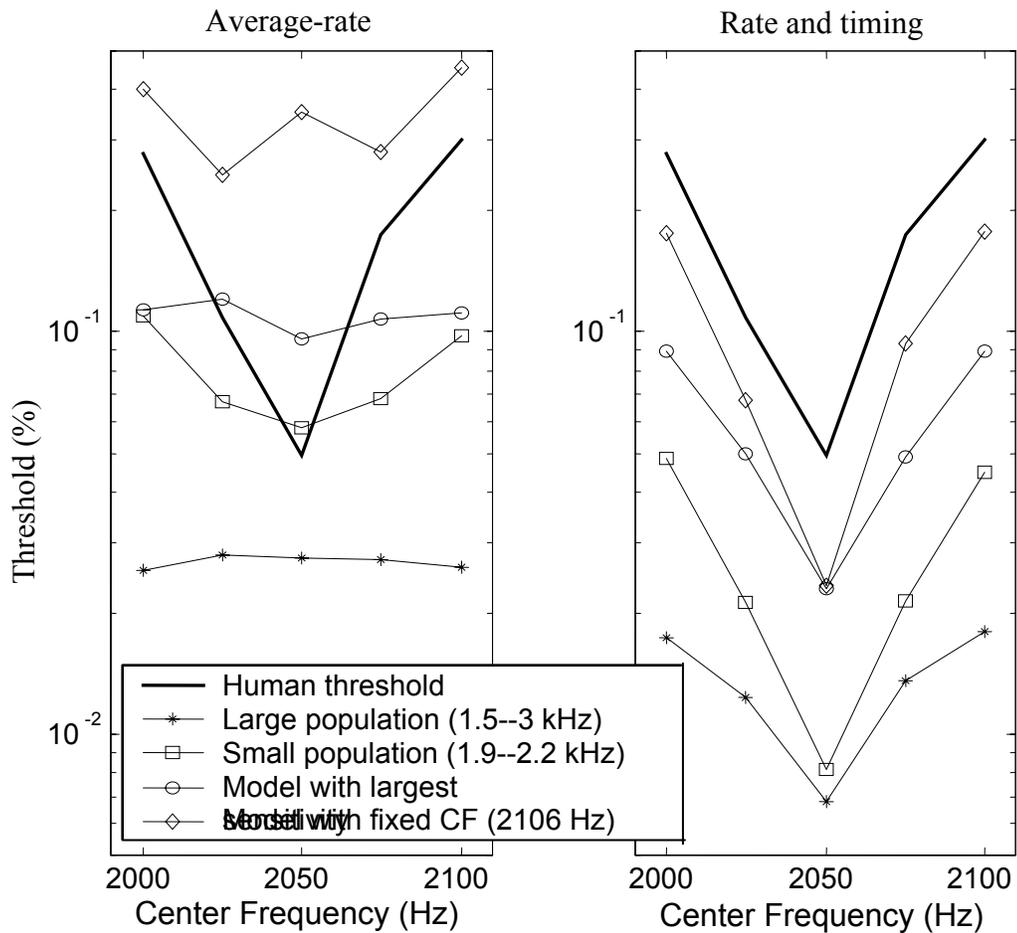
Figure 2-12 Thresholds for triangular spectrum (400 dB/oct slope) with various selections of AN model fibers based on (a) average-rate information (b) both rate and timing information. Human performance (Lyzenga and Horst, 1995) is shown as the solid line with no markers. The selections of the CFs of AN models for the small-population predictions are distinguished by different markers. The line with circles is based on the model fiber with highest sensitivity at each center frequency (The selection of this best model fiber could be different for different center frequencies). The line with diamonds is based on a single AN model fiber with CF at 2106 Hz.

difference in threshold across center frequencies in Fig. 2-11 (b) is larger than the difference in Fig. 2-11 (a). This suggests that a relatively smaller population of AN model fibers could predict a trend more similar to human performance than the previously used larger population of AN model fibers.

Because no matter which prediction method is used, the threshold differences are smaller as the slope of the signal's spectrum envelope gets smaller, starting from Fig. 2-12 the remaining figures focus on only the results for signals with a spectral slope of 400 dB/oct. The discussion of the trend of the thresholds will also focus on spectra with slopes of 400 dB/oct.

Figure 2-12 (a and b) compares the predictions based on various subsets of the model-fiber populations. The asterisks are predictions based on model fibers with CF between 1500 Hz and 3000 Hz. The squares are predictions based on the model fibers with CF between 1900 Hz and 2200 Hz. The circles are predictions in which the model fiber used for each center frequency had the highest sensitivity to that center frequency. The diamonds indicate that a single model fiber with CF at 2100 Hz was used.

In Fig. 2-12 (a) the results are based on only the rate information. The prediction based on the small population with CFs between 1900 Hz and 2200 Hz (marked by squares) shows the correct trend (i.e., lowest threshold at 2050 Hz and highest thresholds at 2000 Hz and 2100 Hz), however the threshold difference in this prediction (about a factor of 2) is much smaller than the difference in human performance (about a factor of 10). The other prediction methods based on only the rate information but with other subsets of the model fibers all fail in predicting the trend of the psychophysical results.

In Fig. 2-12 (b) the predictions are based on both rate and timing information. The prediction based on only the model fibers with CF at 2106 Hz (diamonds) shows a shape that is most similar to the trend seen in human performance, the lowest threshold appears at 2050 Hz, the difference between the lowest threshold and the highest threshold is about a factor of 10, and the trend of the threshold is actually asymmetrical (the threshold at 2075 Hz is higher than the threshold at 2025 Hz). All the other predictions also show the correct trend of the psychophysical data; however, they either do not have the asymmetry (squares and circles) or do not have as great a difference in their lowest and highest thresholds as that in the psychophysical results (asterisks).

2.3.2 Predictions for Signals with a Trapezoidal Spectrum

Figure 2-13 compares the model prediction with human performance for the trapezoidal spectrum's center-frequency discrimination. As stated previously, the analysis is focused on the spectrum slope at 400 dB/oct, because the difference in thresholds across center frequencies is smaller as the spectrum slope decreases. The experiments cover a frequency range equal to two times the fundamental frequency and thus the thresholds show the same patterns in the frequency range of 2100 Hz to 2200 Hz as in the range of 2000 Hz and 2100 Hz. This study thus focused only on the range of 2000 Hz to 2100 Hz for the simulation and investigation of the trapezoidal spectrum, which also made it easier to compare the analysis for the trapezoidal spectrum and the analysis for the triangular spectrum.
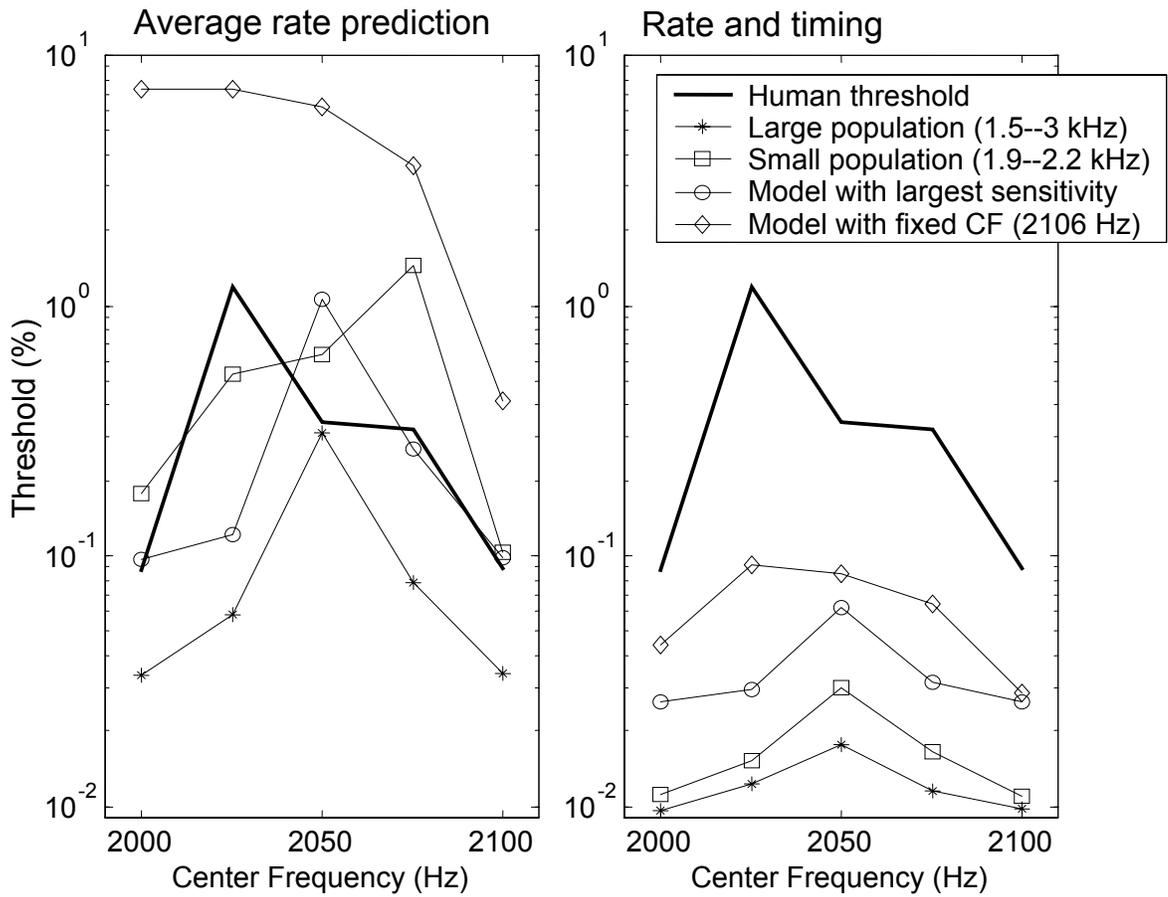
Figure 2-13, Thresholds for the trapezoidal spectrum with a slope of 400 dB/oct (a) with rate information only and (b) with timing information. The thick solid lines illustrate human performance. Different markers distinguish the model predictions with different selections of model CF range. The correspondences between the model CF range and the markers are the same as in Fig. 2-12.
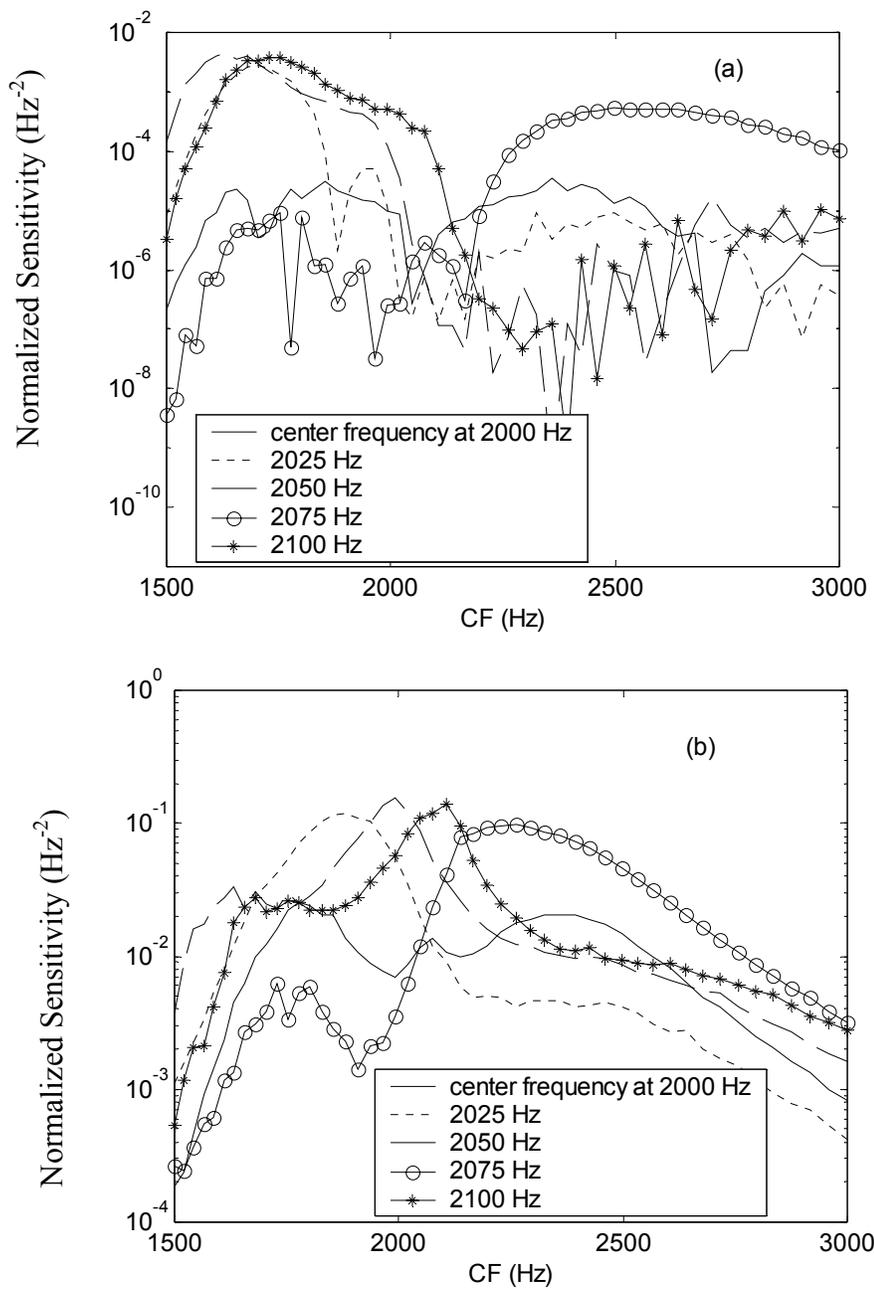
Figure 2-14 Sensitivity patterns as a function of model-fiber CF for the

trapezoidal spectrum (a) with only average rate information (b) with both rate and timing

information. Different harmonic-complex center frequencies are distinguished by

different signs as marked in the figure.

To understand the trends of the results in Fig. 2-13, Fig. 2-14 shows the sensitivity as a function of model CF for the signals with trapezoidal spectrum with an envelope slope of 400 dB/oct. Figure 2-14 (a) is based on the average-rate information only in AN model responses. The integral of the sensitivity over model CF for the center frequency at 2050 Hz [solid line in Fig. 2-14 (a)] is the lowest among all the center frequencies between 2000 Hz and 2100 Hz. Therefore the threshold at 2050 Hz is higher than the thresholds at 2000 Hz and 2100 Hz if the prediction is made based on the rate information of a large population of AN model fibers [CF between 1500 Hz and 3000 Hz; the solid line with asterisks in Fig. 2-13 (a)]. Fig. 2-14 (b) shows the sensitivity patterns based on both rate and timing information: the overall sensitivity for center frequency at 2050 Hz is lower than that for the other center frequencies, and thus the highest threshold appears at 2050 Hz for the prediction with all the model fibers between 1500 Hz and 3000 Hz in Fig. 2-13 (the line with asterisks in panel b).

The exploration of predictions based on smaller AN populations, as was done for the triangular spectrum, was also done for the trapezoidal spectrum. For both the rate-only prediction and the rate-and-timing prediction, the trends of the prediction based on the model fibers with CF between 1900 Hz and 2200 Hz (the lines with squares) and the prediction based on the model fibers with highest sensitivity (the lines with circles) are similar to the trend of the prediction based on all the model fibers with CFs within 1500 Hz and 3000 Hz. The performance based on a single model fiber with a CF equal to 2106 Hz was also calculated (the line with diamonds). For the rate-only prediction, the trend of this prediction is wrong (i.e., the highest threshold appears at 2000 Hz). The trend of the

single-model prediction is the best match to the trend of the human performance among all the predictions (based on both rate and timing information) in Fig. 2-13(b) because it shows the asymmetry of the threshold trend (i.e., the highest threshold is at 2025 Hz instead of 2050 Hz). And it is interesting that the same CF channel (i.e., the model fibers with CF equal to 2100 Hz) have the results that best match human performance for both the triangular spectrum and the trapezoidal spectrum for the rate-and-timing predictions.

2.4 Discussion

This chapter here is an extension of previous psychophysical modeling studies (Siebert, 1965; Heinz 2000) where psychophysical discrimination ability was quantitatively estimated. In this chapter, a harmonic-complex frequency discrimination experiment was simulated with a computational AN model, and the model's thresholds for the frequency discrimination tasks were evaluated. The model performance was predicted either with the rate information only or with both rate and timing information.

The rate-only prediction can be explained by the level cue in the stimuli. Lyzenga and Horst (1995) showed a prediction based on the overall-energy change in the stimuli for the harmonic-complex frequency discrimination. Their overall-energy prediction is in accordance with the rate-only prediction described here. The rate-level function (Tan, 2000; Appendix A) of the AN model used here is a monotonic function and thus the changes in the average response rate of the AN model preserve the changes in the overall energy of the stimuli for stimuli with the levels studied here.

A method of simplifying the harmonic-complex spectrum was useful for identifying potential timing cues encoded in the response to harmonic complexes. The simplified signals at some harmonic-complex center frequencies clearly showed the phase transition cues and could qualitatively explain the general trend of the performance. For the triangular spectrum, when the center frequency (2050 Hz) is between two harmonic components, the speed of the 180 degree phase transition provides important timing information and this information is not available when the center frequency is right at one harmonic component (2000 Hz or 2100 Hz). For the trapezoidal spectrum, the phase transients appear more often in the stimuli where the center frequency is at 2000 Hz or 2100 Hz than in the stimuli where the center frequency is at 2050 Hz, and thus the center frequency of 2050 Hz corresponds to a relatively higher threshold. The rate-and-timing predictions apparently take advantage of this phase transition cue and show the same trends as in the human performance.

Figure 2-15 illustrates this phase transition cue in the response of an AN model fiber with a CF of 2106 Hz, which is the model CF for the single-model predictions in the results section (line with diamonds in Fig. 2-12 and Fig 2-13). In Fig. 2-15 (a), the responses of the AN model fiber to harmonic complexes (triangular spectrum) with center frequencies of 2050 Hz (thick line) and 2060 Hz (thin line) are compared to a sinusoid signal (dashed line). The three signals in the left part of Fig. 2-15 (a) show a 180-degree phase difference between the responses and the reference signal and there is no such phase difference in the right part of Fig. 2-15 (a). Thus, the 180-degree phase reversal is observed in the responses of the AN model fiber. Figure 2-15 (b) shows the
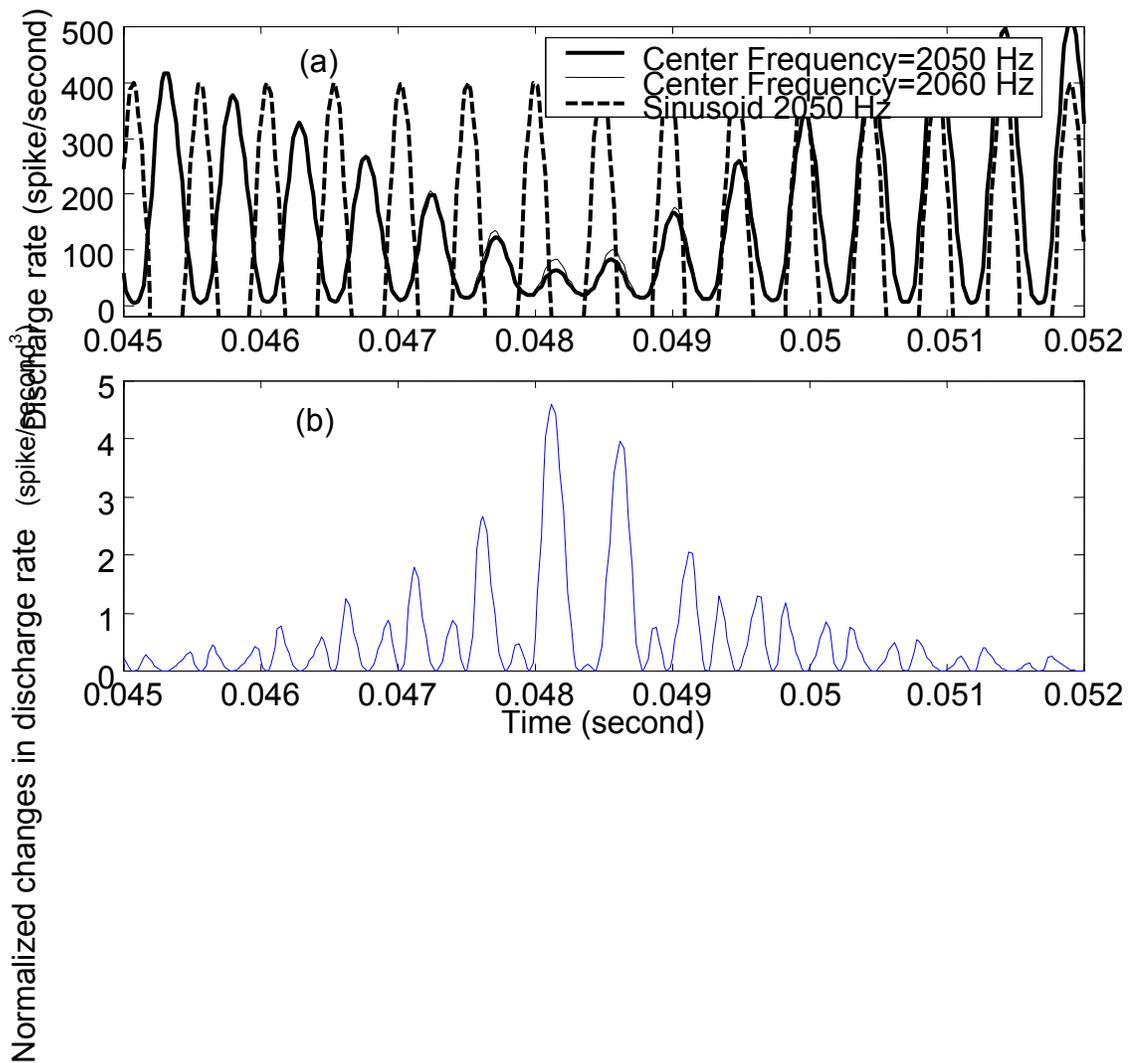
Figure 2-15 The phase transition in the response of an AN model fiber. In panel (a), the

dashed line is a reference signal (sinusoid, 2050 Hz). The thick and the thin solid lines are

responses of an AN model fiber (CF=2106 Hz) to harmonic complexes (triangular

spectrum) with center frequencies at 2050 Hz and 2060 Hz, respectively. Panel (b) shows

the difference between the responses to the harmonic complexes with and without the 10

Hz center-frequency change normalized by the response to the harmonic complex

without the 10 Hz frequency shift. See text for more detail.

normalized changes in the response of the AN model fiber due to a 10 Hz center-frequency change in the harmonic complex, i.e., the difference in the thin and the thick solid lines in Fig. 2-15 (a) normalized by the thick solid line:

$$R_{diff}(t) = \frac{1}{r_{CF=2106}(t \mid f = 2050)} \left[\frac{r_{CF=2106}(t \mid f = 2060) - r_{CF=2106}(t \mid f = 2050)}{2060 - 2050} \right]$$

(2.6)

$R_{diff}(t)$ illustrates the contribution of the response as a function of time and the integral of $R_{diff}(t)$ over time is the normalized sensitivity of this AN model fiber to the center frequency change in the signal. Figure 2-15 (b) shows that $R_{diff}(t)$ has relatively high values during the 180-degree phase reversal. This observation supports the suggestion that the phase transitions could be the cues used in the harmonic-complex center-frequency discrimination.

This study showed that fibers from a single frequency channel predicted better the trend of the threshold across center frequencies for both the triangular spectrum and the trapezoidal spectrum than the predictions based on a large population of AN models. This single frequency channel is on the high-frequency side of the harmonic-complex envelope. This is in agreement with a suggestion of Van Zanten (1980) that the temporal modulation transfer function (TMTF) is governed by the signal contents within the highest frequency region of the stimuli. Van Zanten (1980) measured TMTF for noise stimuli with various bandwidth and center frequencies and found that the TMTF for a wideband noise was the same as that for the highest frequency band of the stimulus. The suggestion of using a small number of AN fibers with CFs near the signal frequency is

also physiologically realistic, because it requires much less neural processing in the

human brain as compared with the processing of information from a huge population of

AN responses. If the stimuli are narrow-band signals, most of the AN fibers outside the

small population generally would not have much sensitivity to the changes of the stimuli,

especially when the signal is at low sound-pressure levels. If the stimuli are broadband

signals (such as speech or in the presence of a noise background) the response of the AN

fibers outside the small population are likely to be dominated by stimulus components

other than the target component. On the other hand, when the task in a psychophysical

experiment is to discriminate the changes of more than one frequency component (e.g.

different vowels usually have different frequencies for their first, second and third

formants), it is more realistic for the auditory system to discriminate a formant-frequency

change by focusing on the information from AN fibers tuned to frequencies near that

formant frequency.

The method of using fewer model fibers results in model threshold trends that are

more similar to the trends of the human performance than the method of using a larger

population for the triangular spectrum [Fig. 2-12 (b)] but not so significant for the

trapezoidal spectrum [Fig. 2-13 (b)]. The reason is likely to be that the trapezoidal

spectrum has a 200 Hz plateau and thus it has a larger bandwidth than the triangular

spectrum. More AN fibers could be involved in this discrimination task.

The assumption of an ideal central processor is not physiologically realistic. It

requires that the central neural system has a perfect memory for the response patterns to

each stimulus for the rate-and-timing prediction, which in general predicts the lowest

bound of the threshold and therefore always has a lower threshold than the human threshold. Thus we need some kind of realistic decoding mechanism that is also sensitive to timing information, such as the coincidence detection mechanism, which will be explored in the next chapter.

One goal of this study is to improve our understanding of speech processing in the auditory periphery. The harmonic-complex is an over-simplified version of a vowel signal. It is important to explore psychophysical experiments with stimuli closer to natural speech. In a later chapter, the model performance for discrimination of formant frequency in synthesized speech with various levels of background noise will be explored.

Chapter 3 Harmonic-Complex Center-Frequency Discrimination Based on a Cross-Frequency Coincidence-Detection Mechanism

3.1 Introduction

The performance limits of discriminating the center-frequency change of the harmonic complex envelope based on the response patterns of a population of AN model fibers were quantitatively evaluated in the previous chapter. The evaluation was based on the performance of an optimal processor, which could optimally use the response patterns of a population of auditory-nerve (AN) models. It was demonstrated in the previous chapter that the prediction based on both average rate and timing information of the AN model response can predict the trends of human thresholds better than predictions based only on the average-rate information. This result suggested that timing information of AN responses was important in the auditory processing of complex signals in the psychophysical experiments described in the previous chapter.

However, the assumption that a processor could optimally use both the rate and temporal information of the AN response pattern was not physiologically realistic. The coincidence-detection mechanism is a physiologically realistic sub-optimal processor of monaural temporal information, which is of special interest for the study presented here. A binaural coincidence-detection model has been proposed to decode temporal cues in complex sounds for binaural psychophysical studies (Colburn 1969, 1973, 1977). Heinz *et al.* (2001b) adopted the cross-fiber coincidence-counter mechanism and employed a

computational AN model in their study of monaural discrimination of level (Heinz *et al.*, 2001 a).

Each coincidence detector (the basic unit of the coincidence detection) receives inputs from a pair of AN models [i.e. the responses of the AN model fibers are used as the input of the coincidence detector, (Fig. 3-1)].
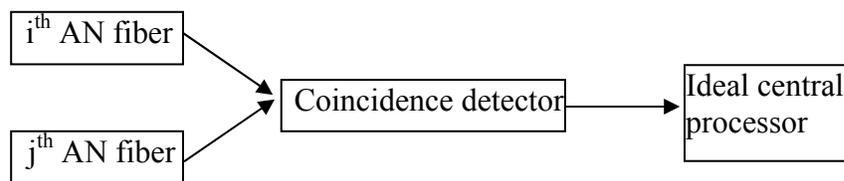
Figure 3-1 Structure of a coincidence detector

Because the bandwidth and the gain of the band-pass filter in the AN model's signal path are changed by the control signal (the output of the feed-forward control path of the AN model) continuously, the phase spectrum of this band-pass filter depends on the input sound intensity. This coincidence-detection mechanism takes advantage of this nonlinear phase cues in the responses of the two converging AN models. The nonlinear phase change in these two AN model responses can be different with a change in the input sound, and thus the response of the coincidence detector can be sensitive to a spectral change in the sound stimulus.

Both Colburn (1969, 1973, 1977) and Heinz *et al.* (2001b) only considered the count information of the coincidence detector output. The study presented here extended the coincidence-detection mechanism by discussing the predictions based on the fine structure (temporal information) of the coincidence-detector responses.

Due to the nonlinear nature of timing of AN responses, the coincidence detector is sensitive to changes in both the intensity and the frequency of the stimulus. When the sound stimulus changes, the timing of the responses of the AN model inputs tuned to different frequencies change differently. This difference results in changes of the response of the coincidence detectors. The predicted performance based on a population of coincidence detectors will be compared with human performance in psychophysical tasks (Lyzenga and Horst, 1995).

3.2 Methods

In the study presented in this chapter, center-frequency discrimination experiments (Lyzenga and Horst, 1995) were simulated. The experiments with both the triangular and the trapezoidal spectrum envelope were simulated. A detailed description of these psychophysical experiments was included in the previous chapter. The simulations and predictions for this chapter were focused on the harmonic complexes with an envelope slope of 400 dB/oct because the difference in threshold is always smaller with a smaller slope (Lyzenga and Horst, 1995).

Model performance was quantitatively evaluated using the response patterns of model coincidence detectors. Each of the coincidence detectors receives a pair of AN model outputs. Whenever the two AN model fibers both discharge within a short coincidence window, the coincidence detector discharges. Thus the discharges of the coincidence detector can be described by:

$$Cij(t) = \sum_{m=1}^{N_i} \sum_{n=1}^{N_j} g(t - t_m^i) \times g(t - t_n^j) \qquad\qquad (3.1)$$

where: (1) i and j are the indices of the AN fibers that converge on a coincidence cell, (2)

g(t) is a short rectangular coincidence window unit height, (3) and $\{t_1^i, t_2^i, ..., t_{Ni}^i\}$ and

$\{t_1^j, t_2^j, ..., t_{Ni}^j\}$ are discharge times of the i-th and the j-th fiber, respectively. The

probability of the coincidence detector's discharge within the coincidence time window

at time t is equal to the probability that the two input AN model fibers discharge within

that time window at time t: $P_{CD}(t) = P_{AN1}(t) \times P_{AN2}(t)$ and thus

$r_{ij}(t) \times \Delta t = [r_{ani}(t) \times \Delta t] \times [r_{anj}(t) \times \Delta t]$, where r_{ij} is the instantaneous discharge rate of the

coincidence detector, r_{ani} and r_{anj} are the instantaneous discharge rates of the two input

AN models, and Δt is the size of the coincidence time window. Δt is set to be 20µs in this

project, which is about 1/25 of one cycle of a 2000 Hz sinusoid. This Δt value is small

enough for the purpose of the coincidence-detection calculation in this work as the

stimulus is dominated by the frequency components near 2000 Hz. Further decreasing or

slightly increasing the value of Δt will not change the trend of the performance because

the effect of the changes in Δt is simply the scaling of $r_{ij}(t)$. The sensitivity of the

coincidence counter to the variations in the center-frequency of the stimulus was

calculated based on the expected response pattern of each coincidence detector, either the

instantaneous discharge rate $r_{ij}(t)$ or the discharge count $Y_{ij} = \int_0^T r_{ij}(t) \times dt$ where T is the

duration of one trial.

Two different analyses of the coincidence-detection mechanism are considered here. In both approaches the responses of a single coincidence detector are assumed to be a Poisson process. The first one uses only the discharge count of the coincidence detectors. Thus the prediction method can be easily adapted from the prediction based on the average rate of the AN model responses (Eq. 2-4), except that the response rate of the AN model is substituted by the response rate of the coincidence detector: The sensitivity of a single coincidence detector is:

$$Q_{ij} = \frac{1}{Y_{ij}} \left[\frac{\partial Y_{ij}}{\partial F_c} \right]^2 \tag{3.2}$$

The threshold (just noticeable difference) of the simulation based on the coincidence-detector population performance is predicted by the square root of the sum of the sensitivities of the population (the inputs and activity of the coincidence detectors were assumed to be independent):

$$F_{cJND} = \left[\frac{1}{\sum_i \sum_j Q_{ij}} \right]^{\frac{1}{2}} = \left[\frac{1}{\sum_i \sum_j \frac{1}{Y_{ij}} \left[\frac{\partial Y_{ij}}{\partial Fc} \right]^2} \right]^{\frac{1}{2}} \tag{3.3}$$

The second analysis uses timing information (fine structure in the time domain) of the coincidence-detector response patterns and thus the sensitivity of a single coincidence detector is

$$Q_{ij} = \int_0^T \frac{1}{C_{ij}(t)} \left[\frac{\partial C_{ij}(t)}{\partial F_c} \right]^2 dt \tag{3.4}$$

50

and the threshold based on the coincidence detector population is

$$
F_{cJND} = \left[\frac{1}{\sum_i \sum_j Q_{ij}} \right]^{\frac{1}{2}} = \left[\frac{1}{\sum_i \sum_j \int_0^T \frac{1}{C_{ij}(t)} \left[\frac{\partial C_{ij}(t)}{\partial F_c} \right]^2 dt} \right]^{\frac{1}{2}}
\qquad (3.5)
$$

For simplicity of calculation, the inputs of the coincidence detectors were restricted to the responses of AN models with characteristic frequencies (CFs) between 1500 Hz and 3000 Hz. The configuration of the distribution of the characteristic frequencies of AN models are the same as in the previous chapter. It was assumed that the total population of the AN fibers in the CF range of interest was represented by 50 model AN fibers, each of which represented about 60 AN fibers. Each AN fiber is assumed to project to only one coincidence detector. Thus, the 50 model fibers in this project require 1275 model coincidence detectors to represent all possible pair-wise combinations of the CFs of the input AN fibers. The performance based on a smaller population of coincidence detectors is also discussed in this chapter.

The AN model (Tan, 2000; Appendix A) used in this study was the same as described in the previous chapter and thus its description is not included here.

3.3 Results

3.3.1 Predictions for Signals with a Triangular Spectrum

Figure 3-2 compares the performance of the coincidence detection mechanism with human performance (thick solid line in both panels) for the triangular spectrum with an envelope slope of 400 dB/oct. The prediction based on the count of the coincidence detector [line with asterisks in Fig. 3-2 (a)] showed the wrong trend across center frequencies: the threshold at 2050 Hz was the highest instead of the lowest. The prediction based on both the average rate and timing information [line with asterisks in Fig. 3-2 (b)] had the correct trend, yet the threshold difference across different center frequencies was not as large as in human performance (thick solid line). All the predicted performances have higher thresholds than the human performance.

The sensitivity (Q_{ij} in Eq. 3-2 and Eq. 3-4) patterns for each center frequency of the harmonic complex are plotted in Fig. 3-3, where the left column is based on the count of the coincidence detector responses and the right column is based on both the average-rate and timing information of the coincidence detector responses. In each of the ten panels, either based on count information or based on temporal information, there are one or two groups of coincidence detectors having relatively higher sensitivity on the diagonal (i.e., the coincidence detector receives inputs from two AN model fibers with same or similar CF). However the distribution of these high sensitivity coincidence detectors on the diagonal are different for the count prediction and the rate-and-timing prediction. Because these coincidence detectors receive inputs from AN models with the same CF, their relative sensitivities are basically proportional to the sensitivities of the corresponding AN model fibers and thus the sensitivity pattern on the diagonal is similar to the sensitivity patterns of the AN model fibers at each center frequency. For the count

prediction, there are always two groups of units with high sensitivity, similar to the sensitivity pattern of AN rate prediction (Fig. 2-11) where there are always two peaks with one notch in between. For the rate-and-timing prediction, there is a single group of high-sensitivity coincidence detectors on the diagonal for center frequency equal to 2050 Hz which corresponds to the single peak in the sensitivity pattern using temporal information of AN model for the same center frequency (line with circles in Fig. 2-12) and there are two groups of high-sensitivity coincidence detectors on the diagonal for the other center frequencies, corresponding to the other lines in Fig. 2-12. For both the count prediction and the rate-and-timing prediction, the sensitivity pattern always has a group of coincidence detectors with relatively high sensitivity off the diagonal (i.e., they receive inputs from two AN model fibers with different CFs). These off-diagonal coincidence detectors, for both the count prediction and the rate-and-timing prediction, are more sensitive to the relative timing information in the response patterns of the two input AN fibers instead of the average-rate information of the input AN fibers.
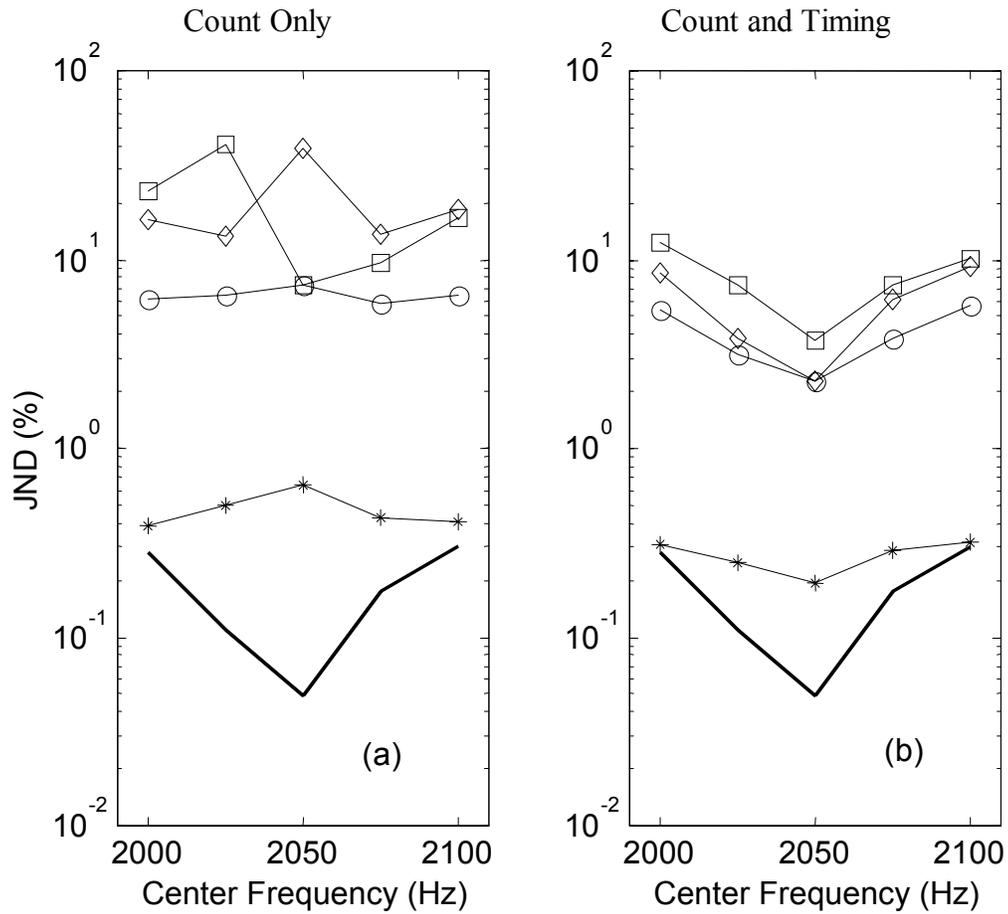
Figure 3-2 Thresholds for triangular spectrum with a slope of 400 dB/oct based on (a)
the count information of the coincidence detectors (b) both count and timing information
of the coincidence detectors. The thick solid line with no markers is human performance
(Lyzenga and Horst, 1995). The line with asterisks is based on the whole population as
described in the introduction section. The line with circles is based on the model
coincidence detectors that have the highest sensitivity at each center frequency. The line
with diamonds and the line with squares are based on two different model coincidence
detectors, the selection of which is described in text.
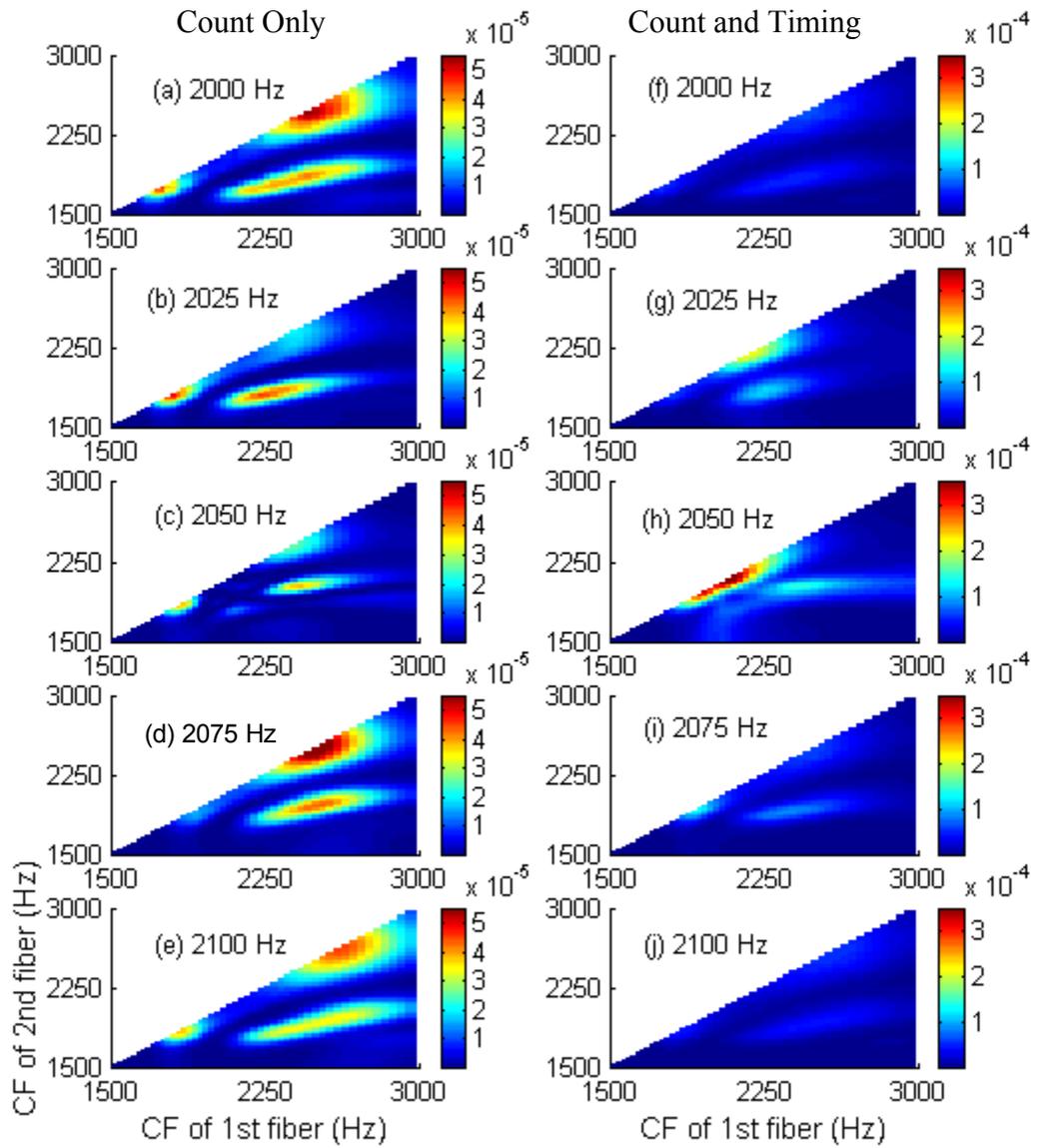
Figure 3-3 Sensitivity (Qij) patterns of the coincidence detectors. Blue shade indicates

smaller sensitivity and red shade indicate a larger sensitivity, as described in the color

bars. The two axes are the CFs of the two input AN model fibers. The left column is

based on the count information. The right column is based on rate-and-timing

information. Each row corresponds to one center frequency of the harmonic complex

(from top to bottom: 2000 Hz, 2025 Hz, 2050 Hz 2075 Hz and 2100 Hz).

As in the previous chapter where the method of using a smaller group of AN model fibers was proposed, the predictions based on only a small group of coincidence detectors was calculated here. The coincidence detectors with the highest sensitivity for each center frequency (the selection of the coincidence detector might be different for different harmonic-complex center frequencies) were first selected to calculate the threshold at that center frequency of the harmonic complex (line with circles). The absolute thresholds are all far higher than human performance for both count-only and rate-and-timing predictions, though the threshold difference across different center frequencies are larger for the rate-and-timing prediction. The count prediction shows the incorrect trend and the rate-and-timing prediction shows the correct trend in threshold as a function of center frequency.

Figure 3-3 (h) corresponds to the highest overall sensitivity (rate-and-timing prediction for center frequency at 2050 Hz). In Fig. 3-3 (h), there is one group of coincidence detectors with high sensitivity on the diagonal and another group away from the diagonal. The model coincidence-detectors that have the highest sensitivities in the first group (on the diagonal) were selected to represent the "best coincidence detectors" for the "easiest task (the center frequency with lowest threshold)" and their performance is shown in Fig. 3-2 (diamonds), based on their count information (panel a) and based on rate-and-timing information (panel b). The count prediction has an incorrect trend (highest threshold at 2050 Hz) while the rate-and-timing prediction has the correct trend (highest threshold at 2000 Hz and 2100 Hz and lowest threshold at 2050 Hz). The

coincidence detectors selected each have two inputs from model AN fibers with CF equal to 2100 Hz, which is the same CF as the best AN model fiber selected in Chapter 2.

Another group of model coincidence detectors was selected: the coincidence detectors with the highest sensitivity in the off-diagonal group in the right panel on the third row (Fig. 3-3). The count prediction of these selected coincidence detectors [line with squares in Fig 3-4 (a)] has the lowest threshold at 2050 Hz however the highest threshold is at 2025 Hz, different from the trend of the human thresholds. The rate-and-timing prediction of these selected coincidence detectors [line with squares in Fig 3-4(b)] has a trend similar to the human performance.

3.3.2 Predictions for Signals with a Trapezoidal Spectrum

As in the previous chapter, the simulation was focused on the harmonic complexes with an envelope slope of 400 dB/oct and center-frequencies between 2000 Hz and 2100 Hz. Figure 3-4 compares the predictions (JNDs) of the coincidence detection mechanism with human performance using different coincidence-detector populations. The threshold predictions based on all the coincidence-detector populations discussed in this chapter, either using count information [Fig. 3-4 (a)] or using rate-and-timing information [Fig. 3-4 (b)] have the correct trends, though the threshold difference across center frequencies are smaller than that of human performance. The line with circles represents the performance based on the "best coincidence detectors" (the coincidence detectors with highest sensitivity) at each center frequency. In the previous
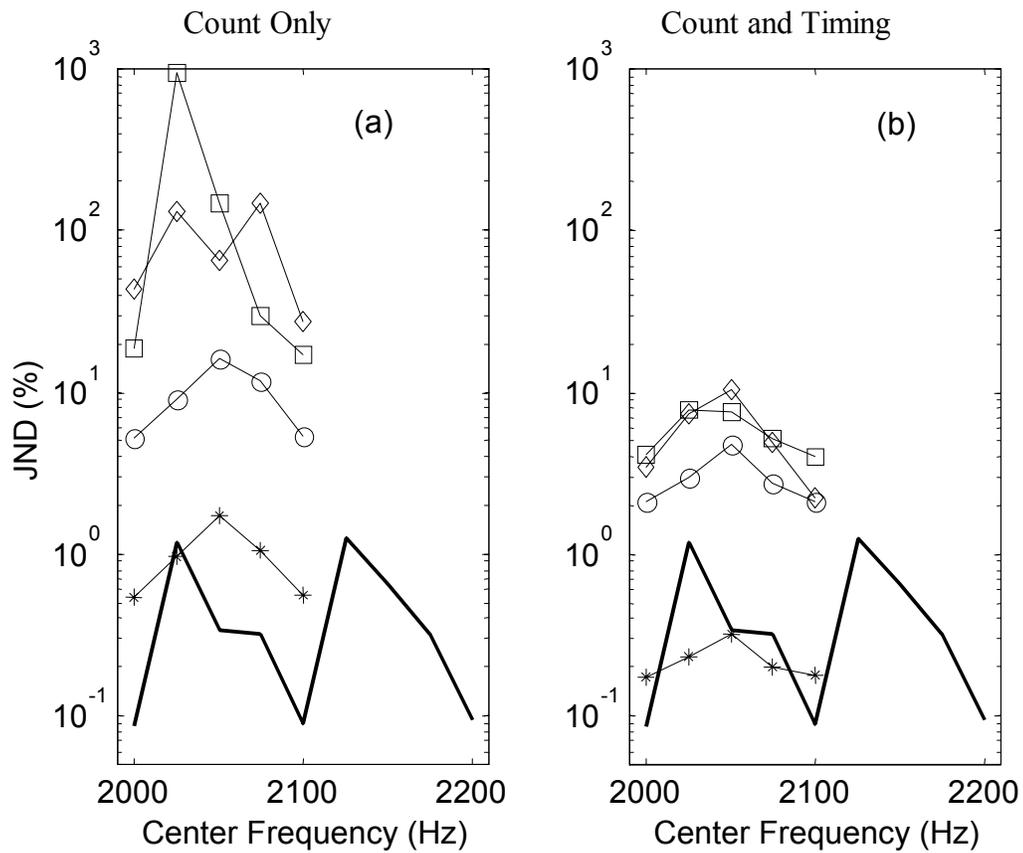
Figure 3-4 Thresholds for trapezoidal spectrum with a slope of 400 dB/oct based on (a) the count information of the coincidence detectors (b) both count and timing information of the coincidence detectors. The thick solid line is human performance (Lyzenga and Horst, 1995). The line with asterisks is based on the whole population as described in the introduction section. The line with circles is based on the model coincidence detectors that have the highest sensitivity at each center frequency. The line with diamonds and the line with squares are based on two different model coincidence detectors, the selection of which is same as in Fig. 3-2.

section for the prediction of triangular-spectrum results, two model coincidence detectors were selected for prediction based on a small group of coincidence detectors: one of them receives inputs from two AN models with same CF and the other one receives inputs from two AN models with different CFs both close to the center frequency of the stimuli. The predictions based on exactly the same two small selections of the coincidence detectors were computed for the trapezoidal spectrum results (line with diamonds in Fig. 3-4 for the coincidence detectors receiving inputs from AN models with same CF (2077 Hz); line with squares in Fig. 3-4 for the coincidence detectors receiving inputs from AN models with CFs of 2019 Hz and 2426 Hz). The rate-and-timing predictions based on these two small populations of coincidence detectors have the correct trend with the lowest threshold at 2000 Hz and 2100 Hz. The absolute threshold is generally higher than the human thresholds in most cases in Fig 3-4, except that for some center frequencies the threshold predictions based on the large population (line with stars) are lower than human thresholds.

3.4 Discussion

The primary goal of this chapter was to describe and test predictions of center-frequency discrimination for harmonic complexes based on a cross-frequency coincidence-detection mechanism. The coincidence-detection mechanism was proposed partly because coincidence detector's response is sensitive to the timing information in the stimuli. It was hypothesized that the spectral changes in the signal associated with the changes in harmonic-complex center frequency could result in different phase cues in

59

pairs of neighboring AN model fibers due to the nonlinear nature of the AN responses. This coincidence-detection mechanism has been successful in predicting the human performance in pure-tone level discrimination tasks (Heinz *et al.*, 2001b).

In general, the predictions based on the coincidence-detection mechanism had trends similar to those based on the optimal processor for a population of AN models (Fig. 2-14): For the simulations with triangular spectrum, the predictions based on count information of the coincidence detectors had the wrong trend (i.e. with a highest threshold at 2050 Hz) and the predictions based on both rate and timing information had the correct trend (i.e. with a lowest threshold at 2050 Hz), though the threshold difference across different center frequencies was smaller than that for human performance. For the simulations of the trapezoidal spectrum results, both prediction mechanisms predicted the correct trend but the threshold difference across the center frequencies was still smaller than that from human performance. This suggested that the coincidence-detection mechanism did not focus on the particular temporal cue (i.e. the speed of the180-degree phase transition) that was apparently important for the center-frequency discrimination of the harmonic complexes. This 180-degree phase transition in the stimuli might require an "intra-channel" temporal-information decoder instead of an "inter-channel" decoder to best match human performance.

The previous chapter suggested that the prediction based on the model fibers with CFs from a small frequency range could better predict the trend of the psychophysical thresholds. This suggestion was also tested with the coincidence-detection mechanism. The same selection of small populations of the coincidence detectors predicted the correct

threshold trend for both the triangular-spectrum results and the trapezoidal-spectrum results based on rate-and-timing information [line with diamonds or squares in Fig. 3-2 (b) and Fig. 3-4 (b)]. This suggested again that only a small portion of AN fibers were involved in the psychophysical tasks with narrow-band stimuli described in this study.

Coincidence-detection predictions are not better than AN predictions in terms of predicting the trends of the performance, however the coincidence-detection mechanism is suggested to be robust in the presence of background noise because temporal cues are more robust than level cues in noise (Carney *et al.*, 2002). The simulations in this chapter were all in quiet (i.e., no noise was included). It is possible that the coincidence-detection mechanism could do better in predicting human performance when a noisy background is added to the psychoacoustical experiments, which will be discussed in the next chapter.

As shown in Fig. 3-2 and Fig. 3-3, some of the predicted performances have threshold trends similar to human performance. However the absolute thresholds in the present study were higher than human performance, which means that the coincidence detection mechanism have degraded some cues important for harmonic-complex center-frequency discrimination. For example, the discharge pattern of the coincidence detectors with inputs from two AN model fibers with the same CF usually has a response rate $r_{ii}(t) = r_{ani}(t) \times r_{ani}(t) \times \Delta t$ (Δt is the size of the coincidence time window), which is a scaled version of the square of the corresponding AN model fiber's discharge rate, $r_{ani}(t)$. Due to the squaring, the information encoded at times when the AN model fiber has a relatively large discharge rate is emphasized more than the information encoded at times when the discharge rate is low. As suggested by the previous chapter, the 180-

degree phase transition could be an important cue for the detection tasks in this study and the phase transitions generally happen when the sound intensity is relatively low (Fig. 2-4, 2-5, 2-6 and 2-7). Thus the coincidence-detection mechanism might have reduced the contribution of this phase transition cue.

What made the coincidence-detection mechanism different from the prediction mechanism directly based on AN model response patterns are mostly the coincidence detectors that receive inputs from AN models with different CFs (those off-diagonal units in the sensitivity patters). The coincidence detectors that receive inputs from AN model fibers with same CF (the units on diagonal in the sensitivity patterns) generally have similar sensitivity pattern as a function of the corresponding AN model's CF to the sensitivity pattern of AN model fibers. This suggests that it might be interesting if the central processor could pay more attention to the off-diagonal units and/or limit the attention to the on-diagonal units.

Chapter Four: Prediction of Formant-Frequency Discrimination in Noise Based on Auditory-Nerve Model Responses

4.1 Introduction

The auditory system has a remarkable ability to extract speech information in a noisy environment. This suggests that incorporation of signal-processing mechanisms based on knowledge of the auditory system could enhance a signal-recognition system's performance, especially in a noisy environment. Thus it is important to know how speech signals are processed in the auditory system. The study presented here intends to improve our understanding of the signal-processing mechanisms in the peripheral auditory system by simulating a formant-frequency discrimination experiment with various decoding mechanisms and comparing the predicted performance with psychophysical data (Hienz *et al.,* 1998) at different background-noise levels.

Formant-frequency discrimination thresholds are quite different from frequency-discrimination thresholds based on pure tones. Human thresholds for tone frequency discrimination are about 1.1 and 3.2 Hz at 600 and 2000 Hz, respectively (Wier et al., 1977) while the human threshold for formant-frequency discrimination is larger: 1.5% of the center frequency (Kewley-Port and Watson, 1994), i.e. 15 Hz at 600 Hz. This difference may indicate that different mechanisms are used to process tonal signals in the tone-frequency discrimination task and speech signals in the formant-frequency discrimination task.

Hienz *et al.* (1998) measured cat formant-frequency discrimination behavioral thresholds in quiet and with a noise background at various signal-to-noise ratios. Their results show that the formant-discrimination thresholds of cat at low and medium noise levels (signal-to-noise ratio at 23 and 13 dB, respectively) are similar to the threshold measured in quiet, and the threshold at high noise level (signal-to-noise ratio at 23 dB) is much higher than the threshold in quiet.

This psychophysical experiment was simulated with the same AN model as described in previous chapters. However the statistical methods used in previous chapters need some modification before they can be applied to the simulations in this chapter. Chapter two showed that auditory-nerve (AN) model predictions matched the trends of human performance in center-frequency discrimination for harmonic complexes, which is a simplification of formant-frequency discrimination. The prediction methods in Chapter two assume that the ideal central processor has detailed knowledge of the AN response pattern for each stimulus. However, this assumption is not realistic when random noise is added in the stimuli.

The study presented here followed the approaches of Heinz (2000) and Heinz *et al.* (2002), who made an assumption that the central processor only has knowledge of the averaged response patterns to the stimulus with noise (i.e., the AN model response averaged across the noise ensemble).

This chapter used the conclusions from the previous chapters as a basis and extended those results to explore the coding mechanisms used in the auditory periphery

for speech-signal processing, or more specifically the coding mechanisms for formant-frequency discrimination.

The prediction of formant-frequency discrimination was calculated first based on AN-model-fiber response patterns. The performance of a smaller population of AN model fibers, which was suggested by the previous results for harmonic complexes, was also tested. Predictions were also made using the coincidence-detection mechanism, which was suggested to be successful in predicting psychophysical thresholds for tone detection in noise (Carney *et al.*, 2002).

4.2 Methods

The specifications of the stimuli followed Hienz *et. al.* (1998): Synthesized vowel signals were produced by a cascade Klatt Speech Synthesizer (Klatt, 1980) for the vowel /ɛ/ with a sampling rate of 50 kHz. The first, third, fourth and fifth formant frequencies of the synthesized speech signals were fixed at 500, 2500, 3300 and 3750 Hz, respectively. The second formant frequency varied over a frequency range between 1700 Hz and 1900 Hz. The threshold was computed for the second formant frequency change near 1700 Hz with different signal-to-noise ratios based on the AN model fibers' response pattern. The stimuli were gated with a 250-ms time window and with rise/fall times of 20 ms. The sound pressure levels of the synthesized speech signals were computed based on the total power in the first 30 harmonics (100 Hz – 3 kHz). The signal-to-noise ratios were computed relative to the total noise power at frequencies below 3 kHz. Fig. 4-1 shows the
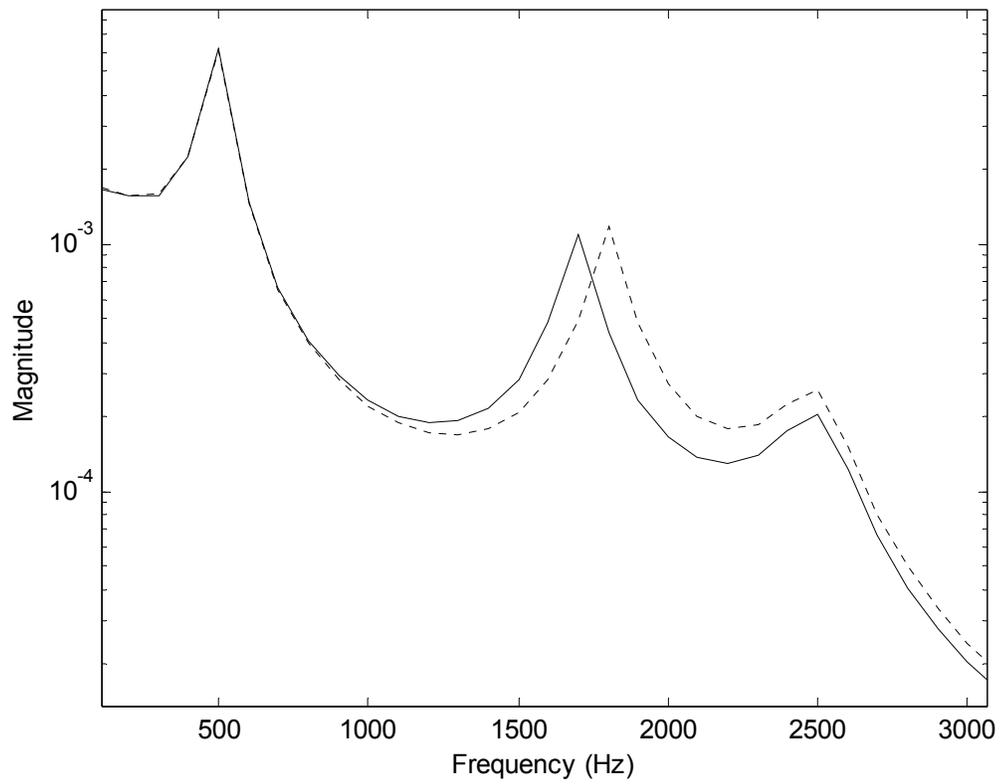
Figure 4-1. Examples of synthesized speech spectra, representing signal with standard F2 (1700 Hz, solid) and with a small frequency shift in F2 (1800 Hz, dotted) in quiet (i.e., no background noise was added).

spectra of two sample speech signals with different second-formant frequencies: 1700 Hz (solid line) and 1800 Hz (dashed line).

The computation of the AN responses in this chapter was based on the same AN model as in the previous chapters (Tan, 2000; Appendix A). 50 model fibers with CFs evenly distributed on a log frequency scale were used to represent all the AN fibers in the auditory system with CFs between 1200 Hz and 2500 Hz. Each of these 50 model fibers represented 63 independent AN fibers, for a total of 3180 fibers in the 1200-2500 Hz CF range. This density of the model CF distribution was similar to that in previous chapters, where 50 model fibers were used to represent the AN fibers with CFs between 1500 Hz and 3000 Hz.

It is not realistic to assume that the ideal central processor knows every detail about the response to each stimulus when random noise is added to the stimuli. Thus the method in Chapter 2 (Eq. 2.3 and 2.4) could not be directly applied to the study presented in this chapter. However, Heinz (2000, see also Heinz *et al.*, 2002) extended this approach to explore the psychophysical performance limits with the presence of random noise by assuming that the central processor knows the *expected* response waveform corresponding to each formant frequency: i.e., the central processor uses the same expected response pattern (averaged response waveform across a group of random noises) as *a priori* information for each trial.

In Heinz (2000), the sensitivity index Q for this central processor is defined by [Eq. 5.11 in Heinz (2000)]

$$Q(f, f + \Delta f) = \frac{\{E_{n,\tau}[Y(\tau) \mid f + \Delta f] - E_{n,\tau}[Y(\tau) \mid f]\}^2}{Var_{n,\tau}[Y(\tau) \mid f]} \qquad (4.1)$$

where $E_{n,\tau}[Y(\tau) \mid f + \Delta f]$ and $E_{n,\tau}[Y(\tau) \mid f]$ are the *expected* value of the decision

variable used by the central processor with and without the formant-frequency shift (Δf)

and $Var_{n,\tau}[Y(\tau) \mid f]$ is the variance of the decision variable due to the random noise and

due to the Poisson randomness of the AN discharges. The decision variable based on the

observation of the AN model population response is an equally weighted combination of

the decision variables based on single AN model responses, i.e.,

$$Y(\tau) = \sum_{i=1}^{M} Y_i(\tau^i) \qquad (4.2)$$

where $Y_i(\tau^i)$ is the decision variable based on the observation of the i-th fiber's

response pattern. Thus the AN model fibers with relatively large changes in their

responses [large $Y_i(\tau^i)$] contribute more to the total value of the decision variable. The

normalized sensitivity squared is given by

$$[\delta'(f)]^2 = \frac{Q(f, f + \Delta f)}{(\Delta f)^2} \qquad (4.3)$$

By combining Eqs. 4.1-3, the normalized sensitivity squared for the whole AN

model population is [Eq. 5.34 in Heinz (2000)]

$$[\delta'(f)]^2 = \frac{\left\{ \sum_{i=1}^{M} \int_0^T \frac{1}{\overline{r_i}(t \mid f)} \left[\dot{\overline{r_i}}(t \mid f) \right]^2 dt \right\}^2}{\left(\sum_{i=1}^{M} \int_0^T \frac{1}{\overline{r_i}(t \mid f)} \left[\dot{\overline{r_i}}(t \mid f) \right]^2 dt + Var_n \left\{ \sum_{i=1}^{M} \int_0^T \left[\frac{\dot{\overline{r_i}}(t \mid f)}{\overline{r_i}(t \mid f)} \right] r_i(t \mid f, n) dt \right\} \right)} \qquad (4.4)$$

where $r_i(t \mid f, n)$ is the response of the i-th AN model fiber to the synthesized vowel with the second formant frequency f and with the presence of the n-th noise; $\overline{r_i}(t \mid f)$ is the averaged response of the i-th model fiber to all the noise, i.e. $\overline{r_i}(t \mid f) = E_n[r_i(t \mid f, n)]$;

$\dot{\overline{r_i}}(t \mid f) = E_n[\dot{r_i}(t \mid f, n)] = E_n\left[\dfrac{\partial}{\partial f} r_i(t \mid f, n)\right]$ is the averaged derivative of the response of the i-the fiber. The first term in Eq. 4.4 corresponds to the variance due to the Poisson randomness of the AN discharges and the second term corresponds to the variance due to the random noise. The model threshold corresponding to $Q(f, f + \Delta f_{JND}) = 1$ is

$$\Delta f_{AN} = \frac{1}{\delta'(f)} \tag{4.5}$$

The prediction based on only the average-rate information of the AN responses was also calculated by assuming that the central processor knows only the number of discharges or the average discharge rate for each trial and thus the response of the i-th fiber can be assumed to be a stationary Poisson process with a discharge rate of $R_i(f, n)$. The performance is thus based only on the information in the average response rate of the AN model fibers [Eq. 5.36 in Heinz (2000)]:

$$[\delta'(f)]^2 = \frac{\left\{\displaystyle\sum_{i=1}^{M}\left(\dfrac{1}{\overline{R_i(f)}}\left[\dot{\overline{R_i}}(f)\right]^2\right) \times T\right\}^2}{\left(\displaystyle\sum_{i=1}^{M}\left(\dfrac{1}{\overline{R_i(f)}}\left[\dot{\overline{R_i}}(f)\right]^2\right) \times T + Var_n\left\{\displaystyle\sum_{i=1}^{M}\left[\dfrac{\left[\dot{\overline{R_i}}(f)\right]}{\overline{R_i(f)}}\right] \times R_i(f, n) \times T\right\}\right)} \tag{4.6}$$

Cross-frequency coincidence detection has been proposed as a mechanism to detect temporal cues encoded in sound stimuli (Heinz *et al.,* 2001b). As described in the Chapter 3 of this document, the coincidence-detection mechanism did not provide better predictions for the trend of the psychophysical thresholds than the prediction based on the AN model response patterns for discrimination of center frequency of harmonic complexes. However the coincidence-detection mechanism was reported to be successful in predicting psychophysical thresholds for detection of tones in the presence of background noise (Carney *et al.*, 2002), which was similar to the focus in this chapter. Thus the performance of the coincidence-detection mechanism was quantitatively evaluated here.

For simplicity, the outputs of the coincidence detectors were assumed to be independent non-stationary Poisson processes. The quantification methods were adapted from those for the AN model fiber performance (Eq. 4.2), except that the total sensitivity was accumulated over the whole population of coincidence detectors instead of AN model fibers:

$$[\delta'(f)]^2 = \frac{\left\{\sum_i \sum_j \int_0^T \frac{1}{\overline{C_{ij}(t|f)}} \left[\overline{\dot{C}_{ij}(t|f)}\right]^2 dt\right\}^2}{\left(\sum_i \sum_j \int_0^T \frac{1}{\overline{C_{ij}(t|f)}} \left[\overline{\dot{C}_{ij}(t|f)}\right]^2 dt + Var_n\left\{\sum_i \sum_j \int_0^T \left[\frac{\overline{\dot{C}_{ij}(t|f)}}{\overline{C_{ij}(t|f)}}\right] C_{ij}(t|f,n)dt\right\}\right)}$$

(4.7)

where Cij(t|f,n) is the response of the coincidence detector that receives the responses of the i-th and the j-th AN model fibers.

The calculation of the partial derivative with respect to the formant frequency in this chapter was approximated by taking the difference between the AN (or the coincidence detector) discharge rate at two different formant frequencies (with and without the frequency shift) and dividing this difference by the change in the formant frequency.

4.3 Results

Figure 4.2 illustrates the thresholds at various background-noise levels. Cat performance [thick line, from Hienz *et al*., (1998)] shows similar thresholds in quiet, low noise level and medium noise level (the curve is almost flat) and the threshold at high noise level is about two times that at the other noise levels. The prediction based on the average response rate of the AN model fibers (squares) showed an increasing trend from quiet to low noise level, to medium noise level, and to high noise level, indicating that the performance degraded as the background noise increased across the entire range of noise levels. The performance based on both average rate and timing information of the AN model fibers (circles, including all the model fibers with CFs between 1200 Hz and 2500 Hz) also showed an increasing trend. The prediction based on a single frequency channel (diamonds, all AN fibers with CF equal to 1720 Hz) showed similar thresholds in quiet, low noise level, and medium noise level and had a threshold increase at high noise level, which was in agreement with the cat performance data. This agreement between the performance a small group of AN model fibers and psychophysical data is consistent with the conclusions from previous chapters.

71

The relative contribution of each AN model fiber for predictions based on average-rate is shown in Fig. 4-3(a). In general, the sensitivity decreased quickly as the noise level increased, though there was relatively less difference between the sensitivities in quiet and for the low noise. The drop of the sensitivity for CFs between 1700 Hz and 1800 Hz was due to the saturation of the rate responses of the corresponding AN model fibers. More contribution was from model fibers with CFs about 100 Hz to 200 Hz below or above the second formant frequency (1700 Hz) than from the model fibers near the second formant frequency.

The relative contribution of each AN model fiber to the overall sensitivity at each signal-to-noise ratio was calculated by quantifying the integral in the numerator in Eq. 4.1 for the predictions based on both average rate and timing information, as shown in Fig. 4-3(b). The contribution of the AN model fibers near the second formant frequency (1700 Hz) was always higher than that of other model fibers at each noise level. The curve was clearly wider at quiet (solid line in bold) or low noise level (dashed line) than at high noise level (solid line with diamonds). The responses of the model AN fibers with CFs away from the second formant frequency might be masked by the noise at high noise levels. This suggests that if the whole population of AN model fibers were used in the predictions, a larger number of fibers contribute in quiet or at low noise levels than at high noise levels.
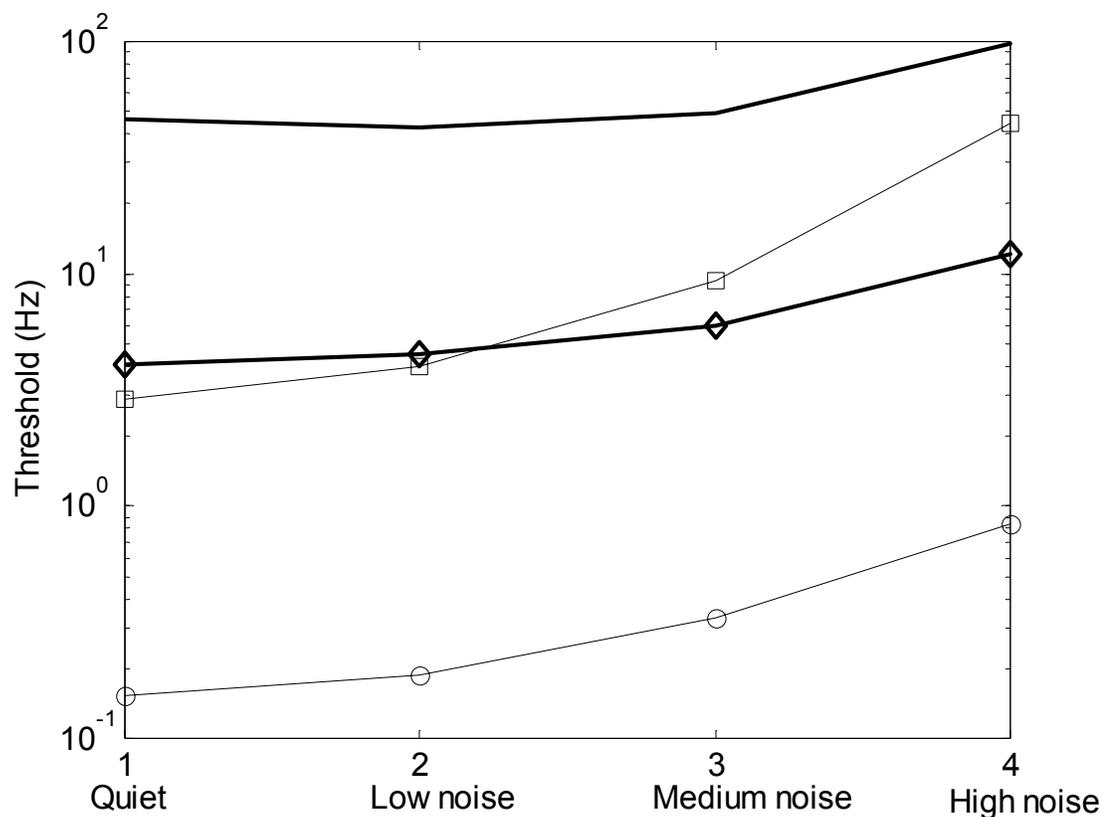
Figure 4-2 Formant-frequency discrimination thresholds at various background-noise

levels. The thick solid line is cat performance. The line with squares is the prediction

based on only rate information with the same selection of model fibers. The line with

circles is the prediction based on both rate and timing information based on model fibers

with CF between 1200 Hz and 2500 Hz. The dashed line with diamonds is the

performance using both rate and timing information based on a smaller group of model

fibers (63, all have same CF at 1720 Hz).
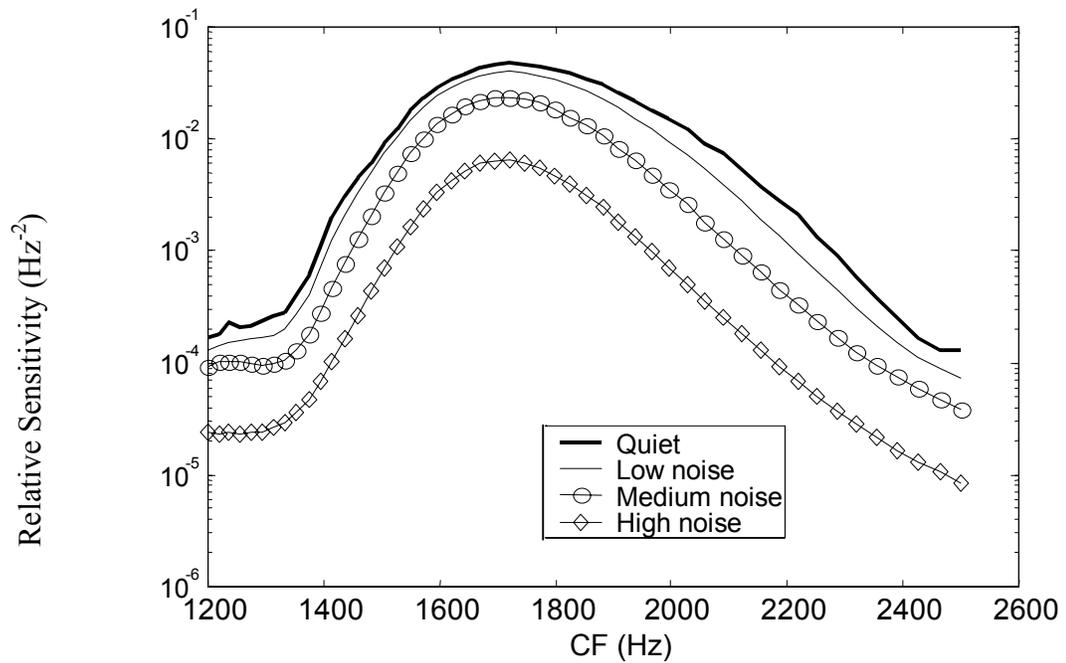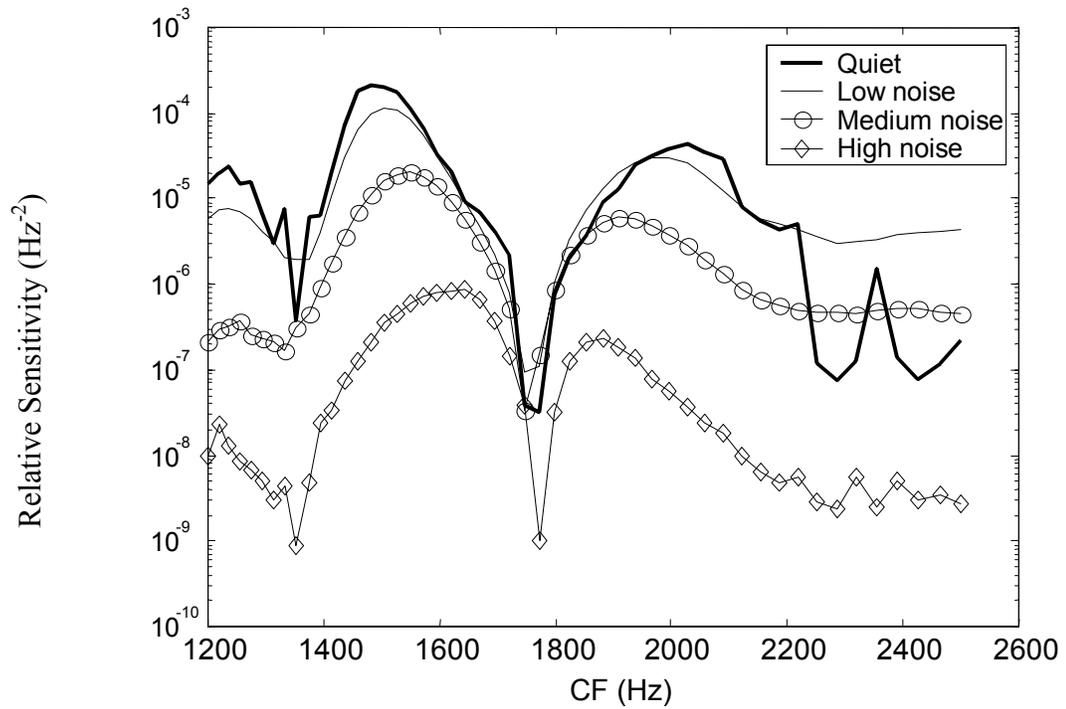
Figure 4-3 Relative contribution of the model fibers at different noise levels: (a) prediction based on average rate information (b) prediction based on both rate and timing information
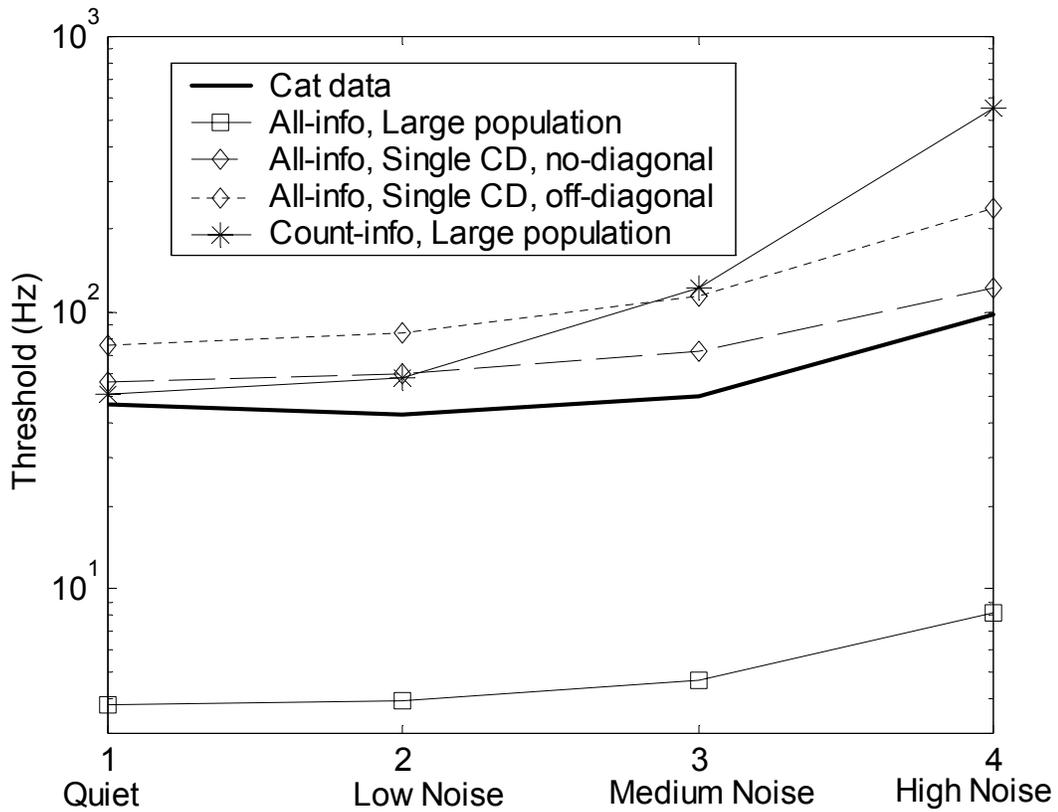
Figure 4-4 Thresholds for formant-frequency discrimination based on the coincidence-detection mechanism (Eq. 4.4). The thick solid line is cat performance. The lines with squares (both count and timing information) and asterisks (count information only) are based on all the coincidence detectors with inputs from model fibers with CFs between 1200 Hz and 2500 Hz. The dotted line with diamonds and the dashed line with diamonds are the performances based on a single coincidence detector on- and off-diagonal of Fig. 4-5 (a), respectively.

As described in Chapter 2, each coincidence detector receives inputs from two AN model fibers (Fig. 2-1). Figure 4-4 demonstrates predicted thresholds based on the coincidence detection mechanism (Eq. 4.4). The predicted threshold based on both count and timing information of a large coincidence-detector population (line with squares) shows an almost flat trend in quiet and with low and medium background noise levels and the threshold at high noise level is about two times of the threshold at medium noise level. This trend is in accordance with the psychophysical data. Thresholds were also predicted with smaller groups of coincidence detectors using both count and timing information. The selection of the smaller groups of the coincidence detectors are based on Fig. 4-5, which shows the relative sensitivity of each coincidence detector to the formant frequency change. In general, in each panel of Fig. 4-5 (corresponding to one signal-to-noise ratio) there are two groups of coincidence detectors with relatively high sensitivity: One group is the one on the diagonal (i.e., receiving inputs from AN fibers with same or similar CFs). The other group is away from the diagonal (i.e., receiving inputs from two AN fibers with different CFs) and has relatively smaller sensitivity than the first group. The coincidence detector with highest sensitivity in each of these two groups was used to compute the formant-frequency discrimination thresholds. In Fig. 4-4, the solid line with diamonds corresponds to the performance of the coincidence-detector from the first group (diagonal, with matched-AN-CFs) and the dashed line with diamonds corresponds to the performance of the coincidence-detector from the second group (off-diagonal, convergence of different AN CFs). The performance of each of these two selections of

76
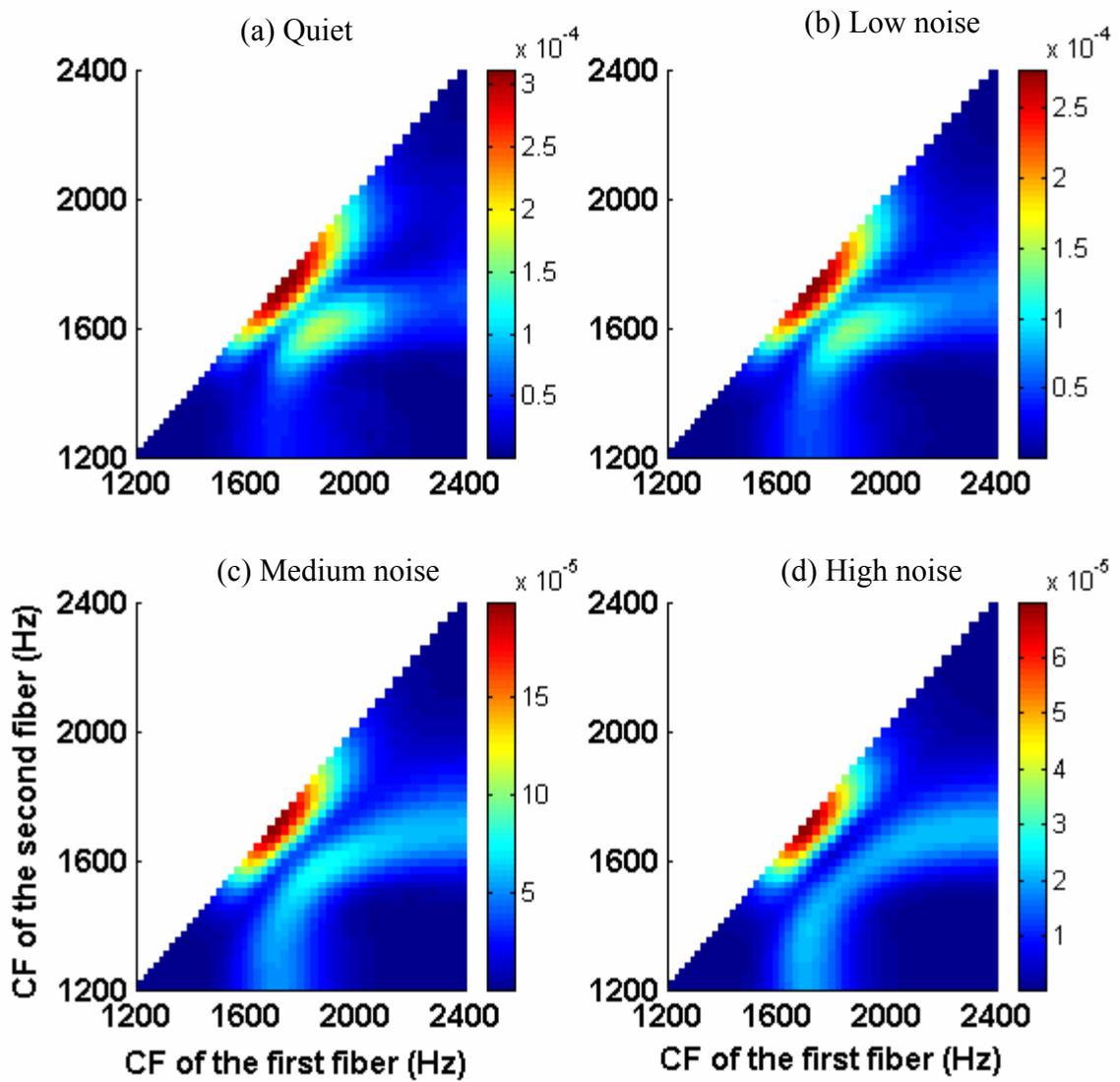
Figure 4-5 Relative sensitivities of coincidence detectors to the formant-frequency change. Each panel corresponds to one background-noise level: (a) Quiet (b) Low noise level (signal-to-noise ratio is 23 dB) (c) Medium noise level (signal-to-noise ratio is 13 dB) (d) High noise level (signal-to-noise ratio is 3 dB).

coincidence detectors using both count and timing information shows a similar trend to the cat performance and to the prediction based on the original population of coincidence detectors (line with squares in Fig. 4-4). The absolute thresholds of each of these two selections of coincidence detectors are slightly higher than psychophysical data. The threshold was also predicted using only count information of the coincidence detector (line with asterisks in Fig. 4-4). The threshold based on count information degraded quickly as the background-noise level increases.

4.4 Discussion

The study presented in this chapter predicted the model performance limits in formant-frequency discrimination based on two decoding mechanisms, one directly based on the response patterns of AN model fibers and the other based on a cross-frequency coincidence-detection mechanism.

The average response rate of the AN model fibers saturated at high noise levels and thus the model fiber sensitivity to the formant-frequency change was significantly reduced by the high-level background noise [Fig. 4-3(a)] if the prediction was based on average-rate information only. When the background-noise level increased, the performance based on average-rate information increased faster than cat performance as well as faster than the prediction based on both rate and timing information (Fig. 4-2), suggesting that timing information was needed in the prediction of this psychophysical performance.

For the predictions based on both rate and timing information, the relative contribution of AN model fibers as a function of CFs [Fig. 4-3(b)] indicated that the model fibers with CFs near but not too close to F2 may contribute to the overall sensitivity at low noise levels, yet they contribute less when the noise level is high because their response was overwhelmed by the noise at high level. Thus the prediction based on a larger population of model fibers degraded faster than the prediction based on only the fibers with CFs within a smaller range.

As suggested by our previous results (Chapter 2) in the simulations of psychophysical experiments, it is realistic and reasonable to predict psychophysical thresholds with a small population of AN model fibers with CFs close to the signal frequency when the task was to discriminate spectrum changes within a small bandwidth. The results presented in this chapter were in accordance with this suggestion of using a smaller population of model fibers.

The threshold prediction based on both rate and timing information of the coincidence-detection detectors shows a flat trend when the noise level is below medium (signal-to-noise ratio at 13 dB) and a small increase (about a factor of 2) at high noise level. This trend is more similar to psychophysical data than the trends of the predictions based on AN model fibers (with all the model fibers with CF between 1200 Hz and 2500 Hz). This suggested that the coincidence-detection mechanism (both rate and timing) might be more robust in a noise background than the mechanisms directly based on AN model response patterns. However, the trend of the prediction based on the AN model fibers with same CF (1720 Hz; dashed line with diamonds in Fig. 4-2) is actually as

similar to the trend of cat performance (thick solid line in Fig. 4-2 and Fig. 4-4) as the trend of the performance based on the coincidence detectors [either the original selection (the line with squares in Fig. 4-4) or the smaller groups (the line with diamonds in Fig. 4-4)]. This suggests that the coincidence-detection process does not specifically extract the cues related to the formant-frequency discrimination tasks from AN model responses.

The threshold trend of the coincidence-detector count prediction degraded faster than the trend predicted by count-and-timing information and is similar to the threshold trend predicted directly based on the average rate of AN model response.

At every background-noise level, the threshold prediction of the coincidence-detection mechanism always has a higher value than the prediction directly based on the response patterns of the AN model fibers, thus the coincidence detection process clearly decreased the total amount of information compared with the information carried by the AN model fiber response patterns, as expected. For either the AN model fiber prediction or the coincidence-detector prediction, the thresholds based on a smaller population always have larger values than the thresholds based on the original selection of model fibers or coincidence detectors, because when more model fibers or more coincidence detectors are used, the total amount of information is increased (Eq. 4-1, Eq. 4-3, and Eq. 4-4).

The work presented here was all based on the formant-frequency discrimination experiment in the second formant in synthesized speech stimuli. It would be interesting to see if the simulation and prediction results are similar if the experiments are done for a lower frequency area (first formant) or a higher frequency area (third formant).

Chapter 5 Summary and Discussion

Many psychophysical experiments have been done to estimate human thresholds for discrimination of frequency changes in speech or speech-related stimuli. However, the decoding mechanisms used by the auditory system are still not clear, partly because of the lack of corresponding modeling studies to understand these psychophysical results. Previous computational modeling studies were mostly focused on tones, modulated tones, or tone combinations. The study presented here simulated two sets of speech-related psychophysical experiments with a computational auditory-nerve (AN) model and compared the results with psychophysical data.

In this study, the limits of model performance were estimated by statistical methods applied from Heinz (2000), which is an extension of the study of Siebert (1965). The model thresholds were computed using the count or temporal information either directly based on the AN model response patterns (Chapter 2 and part of Chapter 4) or based on the coincidence-detector response patterns (Chapter 3 and part of Chapter 4).

The first experiment simulated in this study was the center-frequency discrimination of harmonic complexes, which are a convenient simplification of vowel signals. When average-rate or count information was used, the predicted trend in thresholds was correct for the trapezoidal spectrum yet incorrect for the triangular spectrum. When temporal information was used, the predicted thresholds had a trend similar to human thresholds for both triangular spectra and trapezoidal spectra. Further

analysis showed that the 180-degree phase transition in the stimuli could be associated with the performance.

The second set of experiments used synthesized speech signals as the stimuli. Unlike real speech signals, the synthesized speech signals produced by Klatt Speech Synthesizer (Klatt, 1980) can be reproduced easily and it is also easy to accurately manipulate the parameters of the signals. This is convenient for the current simulation work. The model performance was tested at various background-noise levels. When only the average-rate (or count) information was used, the predicted model performance was degraded more strongly than human thresholds as the background-noise level was increased. When temporal information was used, the degradation of the performance was similar to what was observed in psychophysical data (Hienz *et al.,* 1998).

The results for the harmonic-complex center-frequency discrimination task showed that the prediction based on a single AN model fiber had a threshold trend that matched the trend of human performance better than the performance based on a larger population of AN model fibers (with wider range of model CF).

5.1 Quantification and analysis of temporal and average-rate (count) information in the response of AN model or coincidence detector

This study started with simulations of the harmonic-complex center-frequency discrimination experiment (Lyzenga and Horst, 1995) using a computational AN model. Lyzenga and Horst (1995) showed that the predicted thresholds based on the overall-level change in the stimuli have the incorrect trend compared with human performance.

82

Lyzenga and Horst (1995; 1997) suggested that human performance in these psychophysical tasks were affected by the temporal cues associated with the stimuli. However they did not quantitatively analyze the trend of the thresholds.

Two decoding mechanisms were used in this study to quantitatively estimate the performance limits (JNDs). The first one was based on an ideal central processor that optimally uses the response patterns of the AN models (Chapter 2). The second mechanism involved a cross-frequency coincidence detector that receives inputs from two AN model fibers. The response patterns of the coincidence detectors were sent to an ideal central processor. In both decoding mechanisms, the trends in the predicted thresholds based on temporal information of the AN model fiber or the coincidence-detector response matched trends in human thresholds for both the triangular spectrum and the trapezoidal spectrum. The prediction based on the average response rate of AN model fiber or the counts of the coincidence detectors matched the performance trend for the trapezoidal spectrum but predicted an incorrect trend for the triangular spectrum. This suggested that temporal cues are important in this group of psychophysical experiments.

Lyzenga and Horst (1997) discussed the possibility of using the changes in the envelope-weighted averaged instantaneous frequency (EWAIF; Feth, 1974) or intensity-weighted averaged instantaneous frequency (IWAIF) to analyze the temporal information and account for the trends in performance. They claimed that neither EWAIF or IWAIF could provide a good explanation for the psychophysical data. However, as an extension of this study, the changes in the averaged instantaneous frequency (AIF, not weighted) was calculated for the harmonic complexes:
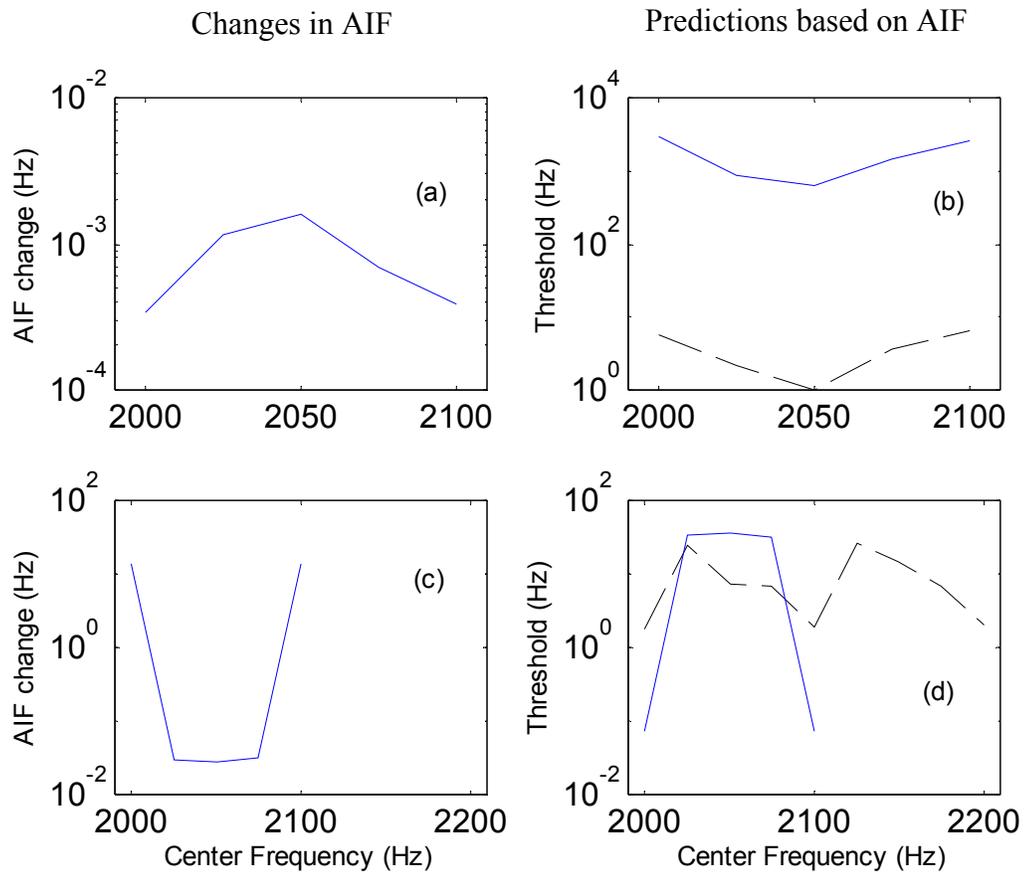
Figure 5-1 Using averaged instantaneous frequency (AIF) to predict center-frequency discrimination results for the triangular spectrum (top row) and the trapezoidal spectrum (bottom row). The left column shows ΔAIF as a function of the center frequency. The right column shows the reciprocal of ΔAIF as a function of the center frequency (solid line) and human performance (dashed line).

$$\Delta AIF = \left| \int_0^T f_1(t)dt - \int_0^T f_2(t)dt \right| \tag{5.1}$$

where T is the time duration of the stimuli; $f_1(t)$ is the instantaneous frequency of the stimulus at a particular center frequency (2000 Hz, 2025 Hz, 2050 Hz, 2075 Hz, or 2100 Hz) and $f_2(t)$ is the instantaneous frequency of the stimulus at a center frequency which has a 10 Hz shift from the center frequency for $f_1(t)$. If the auditory system indeed uses ΔAIF as described in Eq. 5-1 to decode the center-frequency change, then the ΔAIF should indicate the relative sensitivity of the auditory system to the center-frequency change and the reciprocal of ΔAIF should be proportional to the performance threshold. The ΔAIF and its reciprocal are shown in Fig. 5-1. The reciprocal of ΔAIF shows trends as a function of center frequency that are similar to human performance, and thus the changes in the mean value of the instantaneous frequency could roughly account for the trends of the performance. Because ΔAIF is defined as the instantaneous-frequency difference between two stimuli averaged over the duration of the stimulus, it is "averaged timing information" and thus it is a sub-optimal decoding mechanism for timing information.

As described in Chapter 2, the AN model discharge rate $r_i(t)$ preserved the 180-degree phase transition [Fig. 2-15(a)], which is closely related with the signal's instantaneous frequency. Thus the predicted threshold trend might be able to match that of human threshold better if a decoding mechanism that can extract this ΔAIF information from the AN model output is adopted.

The results in Chapter 2 also suggested that the timing information related to phase transitions is mostly near the "troughs" of the signal [i.e., when the signal envelope

is small, see Fig. 2-15 (b)]. However both EWAIF and IWAIF weight the IF at "ridges" of the signal more than the IF at "troughs" of the signal. This could be the reason that EWAIF and IWAIF could not account for the center-frequency discrimination results though the changes in un-weighted AIF might be able to account for the trend of the results.

5.2 Predictions using a smaller group of AN model fibers or model coincidence detectors

Enlightened by the sensitivity patterns of the AN model population to the center-frequency change in the harmonic complex, Chapter 2 evaluated the threshold performance based on a single AN model fiber. The results suggested that the trend of the single AN model fiber's performance could match the trend of human performance better than a larger population of AN model fibers could (CFs between 1500 Hz and 3000 Hz). The selection of the single AN model that best matched the human performance was on the high-frequency side of the harmonic-complex center frequency for both the triangular spectrum and the trapezoidal spectrum. This frequency deviation might be related to the level-dependent best-frequency (the frequency at which the AN model is most sensitive at a particular sound pressure level) shift of the AN model (Tan, 2000; Appendix A). A model fiber with CF (best frequency at threshold) slightly higher than the harmonic-complex center frequency will shift its best frequency to lower frequencies as level increases. Thus the BF could be closer to or right at the center frequency, and thus becomes the most sensitive model fiber to the center-frequency change.

The suggestion that using a small population of AN model fibers could provide a good match to the psychophysical performance was supported by the simulation results in Chapter 4, where formant-frequency discrimination performance (Hienz *et al.*, 1998) was predicted at various background-noise levels. The performance degradation of a single AN model fiber as a function of the noise level was similar to the trend in cat performance (Hienz *et al.*, 1998) yet the degradation increases faster if the prediction was based on a relatively larger AN model population (CFs between 1200 Hz and 2500 Hz). Thus the model predictions based on a smaller population of AN fibers could better account for the trends of psychophysical performance.

5.3 The effect of the nonlinear compression and the instantaneous frequency glide

The results in previous chapters were all based on a nonlinear AN model (Tan, 2000; Appendix A) with an instantaneous frequency glide in its impulse response. The nonlinear compression property of this model is important for the simulation results in this study because the level-dependent gain and bandwidth of the bandpass filter in the signal-path of this AN model resulted in level-dependent phase responses, and thus the temporal information encoded in the AN response patterns could be affected by the nonlinearity of the AN model.

The effect of this nonlinear compression was quantitatively analyzed first by comparing the performance of a linear version of the AN model with the performance of the original AN model (Fig. 5-2). The linear AN model was achieved by forcing the output of the control-path of the model to be zero and thus the signal-path became a

87
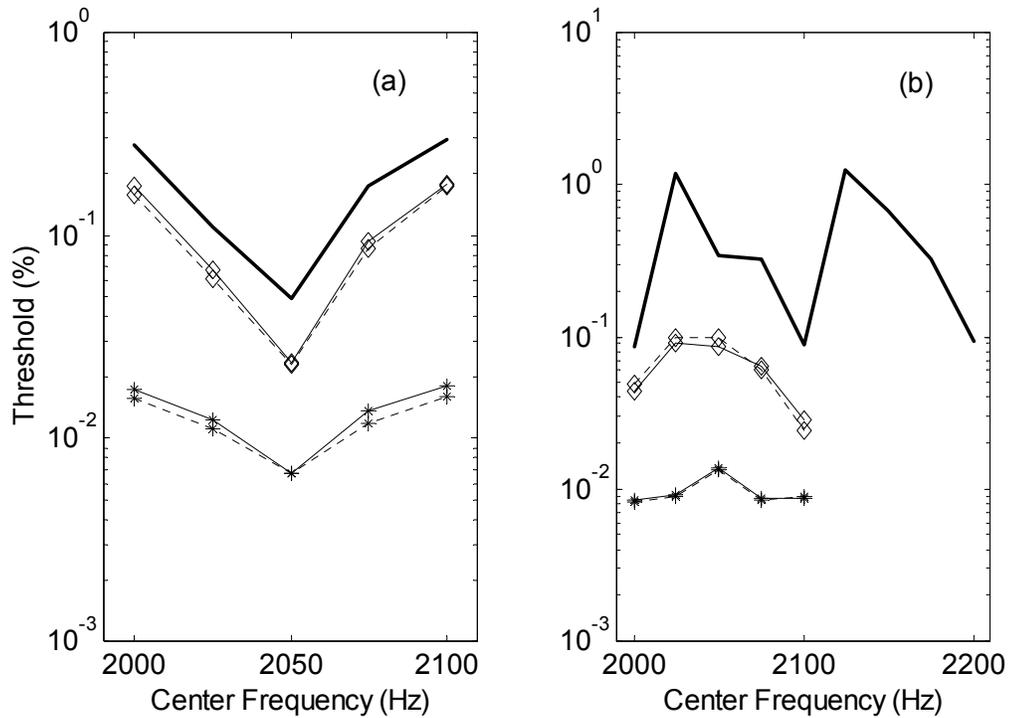
Figure 5-2 Compare predicted thresholds based on the temporal information of the response patterns of linear AN models (dashed lines) with that of nonlinear AN models (solid lines with markers) for (a) triangular spectrum and (b) trapezoidal spectrum. The thick solid line is human performance. The line with asterisks is based on all the AN model fibers with CF between 1500 Hz and 3000 Hz. The line with diamonds is based on the AN model fiber with CF equal to 2077 Hz.
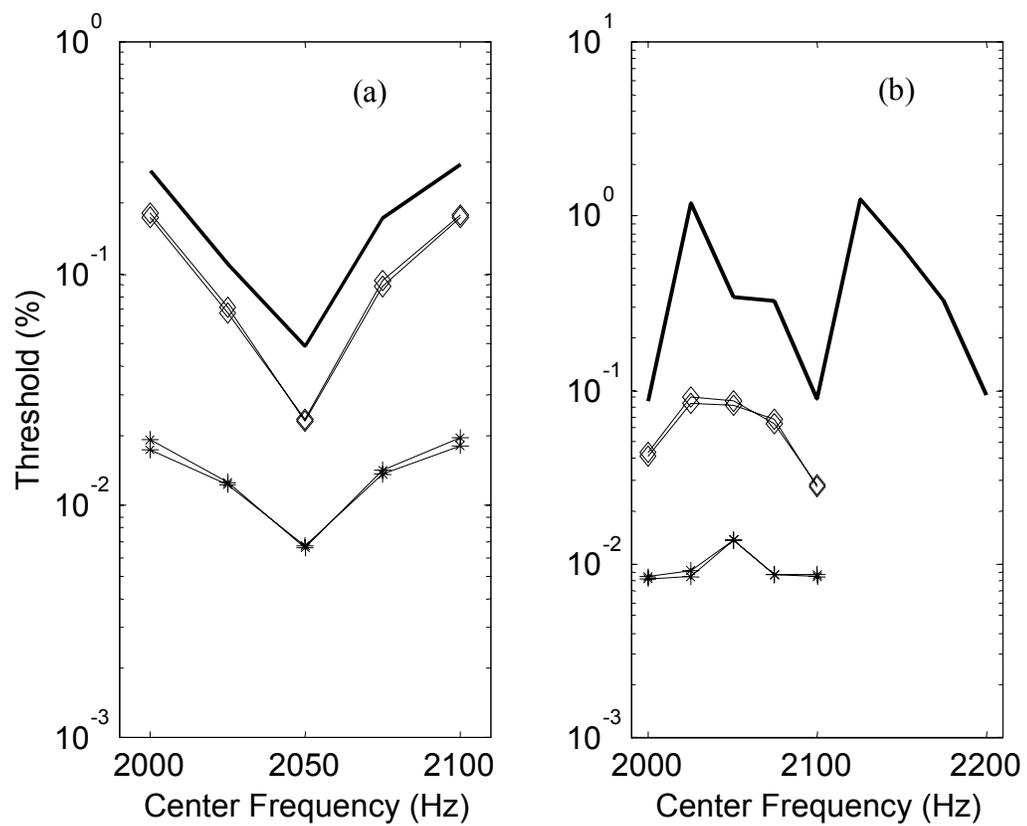
Figure 5-3 Effect of removing the frequency glide of the AN model for (a)
Triangular spectrum and (b) Trapezoidal spectrum. The dashed lines are predicted
thresholds based on the temporal information of the response patterns of the AN
models without frequency glides and the solid lines with markers are that for the
original AN models. The meanings of the markers are same as in Fig. 5-2.

linear bandpass filter. In general, both the trend and the absolute threshold of the linear model's performance are similar to that of the original nonlinear AN model. However, for the results based on the larger population (marked by asterisks in Fig. 5-2), the absolute threshold is slightly lower for the linear model than that for the nonlinear AN model. This is very likely due to the fact that a) the bandwidth of the nonlinear model is generally wider than the linear model and thus the nonlinear model's frequency resolution is lower, and b) the gain of the nonlinear model keeps going down as the input sound intensity increases, and thus the nonlinear model is less sensitive to the sound intensity change than the linear version. For the results based on both rate and timing information of only one model fiber [marked by diamonds in Fig. 5-2 (b)], the performance of the linear model looks like a shifted version (to the low frequency side) compared with the performance of the nonlinear model. This might be the result of best-frequency shift in the nonlinear model [i.e., the best frequency of the nonlinear model shifts to a lower frequency as the input sound pressure level increases (Tan 2000; Appendix A)].

The effect of the frequency glide on center-frequency discrimination predictions is illustrated in Fig. 5-3. The frequency glide was achieved by manipulating the locations of the poles in the control space for the bandpass filter of the original AN model's signal-path (Tan 2000; Appendix A). By placing all the poles together at the same location and carefully adjusting the bandwidth to match the original bandwidth, the bandpass filter becomes similar to a high-order gammatone filter and thus there is no frequency glide in the impulse response or in the model's revcor function. There is no clear difference

90

between the threshold of the model with no frequency glide and the threshold of the original model.

5.4 Limitations and future work

The simulations of the AN fiber responses in this study were all based on a nonlinear AN model (Tan, 2000; Appendix A). Only AN fibers with high spontaneous rate were considered. The parameters of the AN model were based on physiological data of cat and gerbil. They may not best represent the response properties of human AN fibers.

Optimal decision theory estimates the performance limits based on an ideal central processor. It requires that the ideal central processor has ideal memory, exact timing referenced to the onset, and accurate computational ability if temporal information is considered for the prediction work. In general such predicted thresholds are lower than human thresholds because none of the above requirements for the ideal processor is physiologically realistic.

The coincidence-detection mechanism was tested in this work and the results indicated that the coincidence-detection prediction could not account for the psychophysical threshold trend better than the predictions directly based on AN model responses. This suggested that temporal information decoders other than the coincidence-detection mechanisms should be considered in the future work. The search for such decoding mechanisms can be guided by the temporal cues discussed here (e.g., the 180-degree phase transition).

For simplicity, the coincidence detection mechanism assumes that each coincidence detector receives inputs from only two AN model fibers and each AN model fiber only synapses with one coincidence detector. This might be the reason that the thresholds based on the model coincidence detectors were always higher than human performance. Another ongoing project in our laboratory has been involved in developing modeling strategies of more complicated versions of the coincidence detectors, which may be able to better predict the psychophysical thresholds.

The suggestion that a small population of AN model fibers could account for the psychophysical performance should be tested at other frequencies, especially at lower frequencies. Physiological data (Carney *et al.*, 1999) showed that the low-frequency AN fibers (CF<1000 Hz) have downward frequency glides in their revcor functions and this downward frequency glide is associated with an increase in the best frequency as the input sound pressure level increases. Based on this, one would hypothesize that the AN model that could best predict the threshold trend at low frequencies (<1000 Hz) would have a CF on the low-frequency side of the narrow-band signal.

Appendix A

The appendix is a manuscript that has been submitted for publication in the Journal of the

Acoustical Society of America (JASA). It's currently in revision. Thus, the format of the

appendix is that of a JASA manuscript, with it's own Introduction, Methods, Results,

Discussion, and References sections.

A phenomenological model for the responses of auditory-nerve fibers: II.

Nonlinear tuning with a frequency glide

Qing Tan [a) and Laurel H. Carney [a) b)

[a) Boston University Hearing Research Center, Department of Biomedical

Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts

02215, USA

[b) Department of Bioengineering & Neuroscience, Institute for Sensory Research,

621 Skytop Road, Syracuse University, Syracuse, New York 13244.

Address for correspondence:

Laurel H. Carney, Department of Bioengineering & Neuroscience, Institute for

Sensory Research, 621 Skytop Road, Syracuse University, Syracuse, New York

13244. Electronic-mail: Lacarney@syr.edu

Abstract

A computational model was developed to simulate the responses of auditory-nerve (AN) fibers in cat. The model's signal path consisted of a time-varying bandpass filter; a nonlinear feed-forward control path changed the bandwidth and gain of the signal path. This model produced realistic response features to several stimuli, including pure tones, two-tone combinations, wideband noise, and clicks. Instantaneous frequency glides in the reverse-correlation function of the model's response to broadband noise were achieved by carefully restricting the locations of the poles and zeros of the bandpass filter. The pole locations were continuously varied by the control signal to change the gain and bandwidth of the signal path, but the instantaneous frequency profile was independent of sound pressure level, consistent with physiological data. In addition, the control path introduced other important properties, such as nonlinear compression, two-tone suppression, and reasonable Q_{10} values for tuning curves. The inclusion of both the level-independent frequency glide and the level-dependent compressive nonlinearity was the primary focus of this work. The ability of this model to process arbitrary sound inputs makes it a useful tool for studying peripheral auditory processing.

I. INTRODUCTION

The auditory nerve (AN) transfers the information of sound stimuli from the cochlea to the cochlear nucleus, which projects to higher levels of the auditory nervous system. Detailed knowledge of the firing pattern of AN fibers is necessary to understand how sounds are encoded at the input stage of the auditory system. The goal of this study was to improve a previous nonlinear model for the response patterns of AN fibers to different sound inputs. The computational AN model presented here includes the level-independent instantaneous frequency glide and the level-dependent compressive nonlinearity. These properties influence both the rate and timing of AN responses. This model is a useful tool for the study of sound encoding in the peripheral auditory system, and it provides realistic responses that can be used as inputs to models of higher levels of the auditory system.

A system's impulse response can be estimated by the cross-correlation of the response of the system to a wideband noise with the noise input waveform. This technique has been used as an indirect estimate of the basilar membrane (BM) impulse responses while the click response is a direct estimate (deBoer and Nuttall, 1997). The reverse-correlation (revcor) function is an extension of the cross-correlation method and is used as an indirect estimate of the AN impulse response (deBoer and de Jongh, 1978). A frequency modulation, or "glide" in the instantaneous frequency, has been reported in the impulse responses of BM (Robles *et al*., 1976; de Boer and Nuttall, 1997; Recio *et al*., 1997) and AN fibers (Lin and Guinan, 2000; Carney *et al*., 1999). An upward frequency glide indicates that the early part of the impulse response is dominated by lower frequency components and the later part is dominated by higher frequency

components (i.e., frequency increasing). And a downward glide indicates the opposite trend (i.e., frequency decreasing).

Upward frequency glides have been observed in BM and AN responses with relatively high characteristic frequencies (CF > 1500 Hz), constant frequency glides were seen in AN fibers with medium CFs (CF = 750-1500 Hz) and downward frequency glides were seen in low-CF AN fibers (CF < 750 Hz). These frequency glides are consistent with the level-dependent peak-frequency shifts observed in auditory peripheral transfer functions [AN: Moller, 1977; inner hair cell (IHC): Cheatham and Dallos, 1999]. Shera (2001a, 2001b) explored instantaneous frequency glides in BM click responses and suggested that the slope of the normalized instantaneous frequency is independent of cochlear location for CFs above 1.5 kHz and strongly dependent on cochlear location for lower CFs.

The frequency glide pattern not only affects the fine structure of AN response in the time domain but also is related to the best-frequency[1] (BF) shift as a function of sound pressure level (SPL). This level-dependent BF shift can be qualitatively explained by the combination of the instantaneous frequency trend in the impulse response and the change in the shape of the impulse response envelope at various input SPLs (Carney, 1999). The instantaneous frequency tends to have an upward glide in fibers with CFs higher than 1500 Hz, which means that the beginning of the response has a relatively lower instantaneous frequency. When the input SPL is higher, the impulse response has a shorter duration in the time domain, which means that the early part of the impulse response is emphasized and the best frequency at high SPLs is relatively lower.

The glide in the impulse response of a filter is reflected by the asymmetry of its transfer function. This asymmetry indicates that there are poles with different damping coefficients within

the filter. The difference in damping coefficients is associated with a frequency glide in the filter's impulse response (see section IIB for more detail). Because the middle ear affects the asymmetry of cochlear filters (Cheatham and Dallos, 2001), a simple middle-ear function consisting of a linear band-pass filter (Rosowski, 1996) was used to model this aspect of the frequency glide. The contribution of the middle-ear filter to the frequency glide is most important at low CFs.

One goal of this study was to simulate the frequency glide phenomenon in the AN fiber's impulse response. Another focus was the inclusion of the compressive nonlinearity, which is the decrease in the gain of BM response for mid- to high-level sound inputs (Rhode, 1971; Ruggero et al., 1997). The compressive nonlinearity causes broadened tuning of AN responses and shifted phase responses with increased SPLs. Two-tone suppression, defined as the reduction of the response to a tone at CF when a second tone is presented at a frequency other than CF, is associated with the compressive nonlinearity (Ruggero and Rich, 1991).

Many recent reports of phenomenological AN models focus on different aspects of fiber responses. The responses of the auditory periphery, whether recorded from single AN fibers or single sites on the basilar membrane, are characterized as level-dependent band-pass filters. A nonlinear AN model that was developed by Carney (1993) and extended by Zhang *et al.*, (2001) and Heinz *et al.*, (2001) includes a fourth-order gamma-tone filter with level-dependent bandwidth and gain. The most recent versions of this model included level-dependent phase responses, compression, and two-tone suppression; however, the frequency glide observed by Carney *et al.* (1999) in the reverse-correlation (revcor) functions of cat AN fibers was not included.

Meddis and colleagues' (2001) BM model consists of a dual resonance nonlinear (DRNL) filter with two parallel branches, one linear and the other nonlinear. This model and Goldstein's (1990, 1995) multiple bandpass nonlinear (MBNL) model are extensions of Pfeiffer's (1970) bandpass nonlinear (BPNL) model. These models successfully reproduce many physiological phenomena related to basilar membrane motion. However, they do not address the level-independence of the instantaneous frequency glide, which is an important focus of the study presented here. Meddis *et al*. (2001) qualitatively describe the instantaneous frequency glide in the impulse response of their model but do not quantify the frequency glide or demonstrate its level-independence. The level-dependency of this instantaneous frequency in the impulse response of the DRNL model will be compared to that of the model presented here.

Irino and Patterson (1997, 2001) proposed a gammachirp auditory filter to account for peripheral auditory processing. The gammachirp filter is an extension of the gammatone filter and was the first model to include the frequency glide property in its impulse response. Although this model includes frequency glides in the impulse responses, the trends of the best frequency (BF) shifts with SPL are not physiologically accurate in Irino and Patterson's (2001) model. Physiological data (Anderson *et al*., 1971, their Fig. 8) show that the BF of auditory periphery tuning shifts to lower frequencies with increased SPLs for fibers with high CF. However, the BF of one example in Irino and Patterson (2001, their Fig. 7) shifts to higher frequencies for mid-level sounds and back to lower frequencies for high-level sounds. The BF of another example (their Fig. 10, CF=1800Hz) shifts to higher frequencies with increased SPLs.

Robert and Eriksson's (1999) cochlear model is based on a filter bank of all-pole gamma-tone filters (APGFs). Each branch of the filter bank consists of a passive and an active (nonlinear) bandpass filter in series. These two filters are tuned to different center frequencies, and therefore, a best-frequency shift can be observed when input SPL changes. This level-dependency of the

best frequency suggests that their filter-bank model may have an instantaneous frequency glide in the impulse response. However, Robert and Eriksson (1999) did not show the instantaneous frequency profile or test its level-independence. In addition, Robert and Eriksson's (1999) gammatone filters do not include any zeros, which are important for producing downward instantaneous-frequency glides (see Model Description below for detail). These downward glides are observed in low-CF AN fibers (Carney *et al.*, 1999).

The present study combined the level-independent frequency glide with the level-dependent features (e.g., the gain and bandwidth of peripheral tuning) in a simple manner to create a model that can process arbitrary stimuli. Previous studies (Shekhter and Carney, 1997; Tan and Carney, 1999) showed that careful selection of the locations of poles and zeros in the complex plane made it possible to design filters with realistic instantaneous frequency glides in the impulse responses of the filter. The model described here extended the model of Tan and Carney (1999) by combining the pole-zero approach with a feed-forward control path, thereby modeling the compressive nonlinearity of the auditory periphery.

II. MODEL DESCRIPTION

A. Model Overview

The basic model components are shown in the block diagram in Fig. 1. The model consisted of four parts: a middle-ear model, a time-varying bandpass filter as the signal path, a nonlinear control path, and an IHC and synapse model. The middle-ear model was a linear bandpass filter based on the middle-ear frequency response properties described by Rosowski (1996). This linear bandpass filter had two pairs of poles and one second-order zero in control space. The locations of the poles and zeros are specified in Table 1. The low-frequency zeros of

the middle-ear filter improved the downward frequency glide in the model's impulse response at low CFs.

Basilar membrane tuning was modeled with a time-varying bandpass filter, and the compressive nonlinearity of the BM was achieved with the nonlinear control path. The IHC and synapse model was based on that in Zhang *et al.* (2001). A 0.5 ms delay was added to the model output to match model and neural latencies. The output of this model was the instantaneous firing rate as a function of time.

This section describes the major components of the signal path and the control path of the model. This is followed by a description of how the parameters for model fibers across a range of CFs were estimated from AN recordings. The values of all model parameters are listed in Table 1.

B. The Signal Path

The signal path was configured to produce a frequency glide in its impulse response. To illustrate mathematically how manipulation of pole-zero locations generates frequency glides in impulse responses, we consider a fourth-order linear filter with two complex-conjugate pole pairs[2] at p_1 (-x_1, $2\pi f_1$), p_2 (-x_2, -$2\pi f_1$), p_3 (-x_1, $2\pi f_2$), and p_4 (-x_2, -$2\pi f_2$), where $x_1 > 0$ and $x_2 > 0$. The transfer function of this simplified linear filter is:

$$H(s) = \frac{a}{(s - p_1)(s - p_2)(s - p_3)(s - p_4)} \tag{1}$$

The right side of Eq. (1) can be transformed into the sum of four first-order fractions:

$$H(s) = \frac{a_1}{(s - p_1)} + \frac{a_1}{(s - p_2)} + \frac{a_2}{(s - p_3)} + \frac{a_2}{(s - p_4)} \tag{2}$$

where a1 and a2 are gains derived by factoring the right side of Eq. (1).

In the time domain, the impulse response of this linear filter (the inverse Laplace transform of Eq. 2) is:

$$h(t) = 2a_1 e^{-x_1 t} \sin(2\pi f_1 t) + 2a_2 e^{-x_2 t} \sin(2\pi f_2 t), \qquad \text{for } t \geq 0 \qquad (3)$$

x_1, and x_2 determine how quickly the envelopes of the first and second terms in Eq. (3) reach their peak values, respectively. If it is assumed that $x_1 > x_2$ and the values of a1 and a2 are carefully adjusted (a1 > a2), then the first term has a larger amplitude and dominates at the beginning of the impulse response h(t). The second term dominates the latter part of h(t) because the first term decays faster than the second term. Thus, the instantaneous frequency is closer to f_1 at the beginning of the impulse response and is closer to f_2 at the end of the impulse response. Addition of poles to the filter provides increased control over the frequency shifts as a function of time in h(t).

A fifth-order zero was placed on the real axis in the complex plane. For lower CFs, the zero is pushed closer to the origin, which makes the low-frequency side of the filter transfer function steeper and the high-frequency side shallower than that for high CFs.

The combined influence of zeros and poles on the frequency glide can be illustrated as follows: For a simple system with one pair of poles (in conjugate) and one zero on the negative real axis:

$$H(s) = \frac{(s+c)}{(s+a)^2 + b^2} \qquad (4)$$

For the convenience of the inverse Laplace transform, Eq. (4) can be rewritten as:

$$H(s) = \frac{(s+a)}{(s+a)^2 + b^2} + \frac{(c-a)}{(s+a)^2 + b^2} \qquad (5)$$

In the time domain, the impulse response is:

$$h(t) = e^{-at}[\cos(bt) + \frac{(c-a)}{b}\sin(bt)] \quad , \qquad \text{for } t \geq 0 \qquad (6)$$

or more conveniently,

$$h(t) = e^{-at}\sqrt{1 + \left(\frac{(c-a)}{b}\right)^2}\cos(bt - \arctan(\frac{(c-a)}{b})) \quad , \qquad \text{for } t \geq 0 \qquad (7)$$

For a slightly more complicated system with two pairs of poles and two zeros, as shown in Fig. 2, the poles and zeros can be divided into two groups, each having one zero and one pair of conjugate poles. The location of the zeros affects the coefficient in Eq. (7).

The envelope ratio between the two groups of poles and zeros, i.e.,

$$R = \frac{e^{-a_1 t}\sqrt{1 + \left(\frac{(c-a_1)}{b1}\right)^2}}{e^{-a_2 t}\sqrt{1 + \left(\frac{(c-a_2)}{b_2}\right)^2}} \qquad (8)$$

determines the relative dominance of each group of poles when determining the instantaneous frequency at time t.

The model presented here had ten pairs of conjugate poles. The relative locations of the poles and the locations of ten zeros determined the instantaneous frequency glide (Fig. 3). The locations of the poles and zeros were set to be functions of model CF based on fitting revcor functions for a population of AN fibers, as described later. The model-CF dependence of the pole locations is described as:

$$\log_{10}(Pa) = \log_{10}(CF) \times 1.0230 + 0.1607; \qquad (9)$$

103

$\log_{10}(\text{Pb}+1000) = \log_{10}(\text{CF}) \times 1.4292 - 1.1550;$ (10)

$\log_{10}(\sigma_0) = \log_{10}(\text{CF}) \times 0.4 + 1.9;$ (11)

$\text{P}\omega = 1.0854 \times \text{CF} - 106.0034;$ (12)

All zeros were at the same location on the real axis, X_{zero}. The distance between the zeros and the origin was a function of CF on a log-log scale:

$$\log_{10}(X_{zero}) = Z_1 \log_{10}(\text{CF}) + Z_0 \quad\quad (13)$$

The zeros move away from the origin to negative infinity as CF increases. This definition of X_{zero} emphasizes the dominance of the poles with higher frequency at the beginning of the impulse response, especially for low CFs.

The signal-path filter had two eighth-order poles and one fourth-order pole, their complex conjugates, and a tenth-order zero on the real axis. This was the minimum number of poles and zeros required to generate realistic frequency glides in the time domain and realistic sharpness of tuning in the frequency domain.

C. The Control Path

The compressive nonlinearity is an important property of cochlear tuning in the healthy ear. This property was achieved by including the control path, which continuously changed the bandwidth and gain of the signal path. The control path included four segments in series (Fig. 1).

A nonlinear wideband filter determined the frequency range of the stimulus that affects the bandwidth and gain of the signal path. The bandwidth of the control-path wideband filter was set to twice the bandwidth of the signal path when the input was zero. The center frequency of the wideband filter was set to a frequency corresponding to the place on the BM approximately 1.2

mm basal to the place that corresponded to the model CF. The bandwidth and the basal shift of the control path were chosen to achieve the appropriate shape of AN suppression tuning curves (e.g., Sachs and Kiang, 1968; Arthur *et al.*, 1971; Delgutte, 1990). The gain of the wide-band filter was normalized to one at the model CF. A feedback signal derived from the output of the control path increased the bandwidth of the wideband filter with larger input sound intensity. This bandwidth control and the normalization of the gain resulted in the different slopes in the two-tone suppression growth function for suppressor frequencies above or below model CF (Fig. 13).

A symmetric nonlinear function adopted from Zhang *et al.* (2001), followed the wideband filter

$$X_2(t) = \text{sgn}[X_1(t)]B_{cp} \log(1 + A_{cp}|X_1(t)|^{C_{cp}}), \tag{14}$$

In Eq. (14), $X_1(t)$ is the output of the wideband filter in Pascals and $X_2(t)$ is the output of the symmetric nonlinear function. This compressive function made it easier to control the shape of the BM velocity-intensity function (see below, Fig. 7).

An asymmetric second-order Boltzmann function followed the symmetric logarithmic nonlinear function. This Boltzmann function corresponded to the membrane potential-displacement function of the outer hair cell, as suggested by Mountain and Hubbard (1996):

$$Y[X_2(t)] = B[X_2(t)] - B(0) \tag{15}$$

where x_2 was the output of the asymmetric function described by Eq. (14) and B(x) was the second-order Boltzmann function:

$$B[x(t)] = \frac{1}{1 + \exp[\frac{(T_0 - x(t))}{S_0}] \times (1 + \exp[\frac{(T_1 - x(t))}{S_1}])}$$

B(0) was subtracted from B[x(t)] to guarantee that Y(0) was zero.

The parameters T_0, T_1, S_0, and S_1 were chosen such that the asymmetry of this control-path nonlinearity had a 7:1 ratio, as suggested by the responses of outer hair cells (OHCs) (Mountain and Hubbard, 1996).

The last component of the control path was a second-order lowpass filter with an 800-Hz cutoff frequency. The cutoff frequency of this filter was estimated from the results of Recio *et al.*, (1998), which showed that the time course of the onset of compression has a time constant of approximately 0.2 ms, which corresponds to an 800-Hz cutoff frequency. This filter was chosen to be second-order for simplicity; the effect of filter order will be explored in future work.

The control signal (the output of the control path), $\sigma_c(t)$, changes the real part of the locations of the poles of the band-pass filter in control space (i.e., a positive $\sigma_c(t)$ makes the pole locations move further away from the imaginary axis and a negative $\sigma_c(t)$ makes the pole locations move closer to the imaginary axis):

$$\sigma_i(t) = \sigma_{i0} + \sigma_c(t) \qquad (17)$$

where $\sigma_i(t)$ is the damping coefficient of the i-th pole, σ_{i0} is the damping coefficient of the i-th pole when the input is zero, and $\sigma_c(t)$ is the control signal.

C. The IHC and Synapse Model

The IHC and synapse models were the same as in Zhang *et al*. (2001). The IHC model consisted of a logarithmic saturating function followed by a seventh-order lowpass filter. The IHC-AN synapse model was a time-varying three-store diffusion model (Westerman and Smith, 1988; adapted into a time-varying model by Carney, 1993, and Zhang *et al*., 2001). The model described here included only AN fibers with high spontaneous rates.

D. Parameter Estimation

The relative positions of the poles in the signal path (Pa and Pb in Fig. 3) were estimated by fitting the model response to revcor functions of low-frequency cat AN fibers (Carney and Yin, 1988). The damping coefficient of the poles, σ_{80} (the average value of σ in 80 dB SPL noise) was initially estimated on the basis of cat revcor functions computed for responses to 80 dB SPL (rms) noise stimuli. Using 80 dB SPL responses for the initial parameter estimation had two advantages over using lower-level data. First, when the input SPL is high, the signal-to-noise ratio in the revcor function is relatively high. Second, revcor functions for 80 dB SPL responses were available for most fibers in the data set used (Carney and Yin, 1988). A linearized model (with the control signal set to zero and thus with level-independent pole locations) was first used to fit the 80 dB SPL revcor functions. The imaginary part of the pole that was closest to the imaginary axis was first set to the peak frequency of the revcor function's magnitude spectrum. The Marquardt (1963) method was then used to estimate the locations of the poles and zeros in the control space. For simplicity, the zeros were set so that they were always on the

real axis. The target function of the parameter estimation was to minimize the RMS value

of the difference between revcor data and model revcor functions. An example of

optimized locations for poles and zeros is shown in Fig. 3. (The poles for the high-SPL

filter are on the solid short line and are P11, P12, and P13) The parameters estimated with

the high-level AN revcor functions were Pa, Pb, Pω and σ_{80}, where Pa and Pb were the

relative positions of the poles as illustrated in Fig. 3; Pω was the imaginary part of the

pole with the largest imaginary part; and $\sigma80$ was the real part of the poles with the

largest imaginary part and corresponds to the average value of σ when the stimulus is

white noise at 80dB SPL. Figure 4 shows an example of a model revcor function fit to cat

revcor data (fiber U15-C86166 from Carney *et al.*, 1999). This AN fiber with CF of

650Hz has a downward frequency glide in its impulse response.

Revcor data (Carney and Yin, 1988) are available mostly for high SPLs (>40dB).

This makes it hard to estimate the tuning properties of AN fiber for low SPLs for a wide

CF range based on revcor data (Carney and Yin, 1988). Q10 data (Miller *et al.*, 1997) are

based on the tuning curves, which are measuring the thresholds of AN fiber responses.

Thus Q10 data can be used to estimate the tuning properties of AN fibers at low SPLs.

The Q10 data set was used to determine the locations of the poles in the resting state

(when there was no sound input), specifically the value of σ_0 (σ in quiet). The gain of the

control signal was adjusted such that an input of 80 dB SPL noise resulted in an average

control signal equal to the difference in the real parts of P11 and P01. The Q10 value was

then measured for the nonlinear model and P01, P02, P03, and the gain of the control

signal were adjusted until the model's Q10 value matched experimental Q10 data for the

fiber's CF. Because the Marquardt method is sensitive to local minima, several values were used as the initial imaginary part of the pole closest to the imaginary axis so that the fitting performance would not be limited by chance selection of a value.

After σ_0 and σ_{80} were set as functions of CF, the control signal was adjusted to produce an appropriate value to control the locations of the poles in quiet and in 80 dB SPL noise.

The fitting results to each fiber's revcor function were pooled for each parameter (Fig. 5). Simple expressions for the values of these parameters were established as functions of CF, as shown in the text above (Eq. 9-13), enabling the model to simulate AN fiber responses at any CF. A total of 139 cat revcor functions from Carney and Yin (1988) were used for parameter estimation.

III. RESULTS

This section illustrates several response properties of the model to tones and other stimuli. It begins with a description of the model's frequency glide, since that was the primary goal in the development of this model. Other fundamental response properties to tones at CF and to other stimuli are then shown. Nonlinear aspects of average rate and temporal response properties were of particular interest and are discussed below.

A. Instantaneous Frequency Glide

The primary goal of this effort was to incorporate a glide in the instantaneous frequency (IF) of the model's impulse response. The IF glide in this model's revcor function was very close to that reported in the data. An important feature of the IF glide is the constant slope at different noise levels. The model possesses this property because the

relative positions of the poles do not change at different sound intensities (i.e., all the poles move in the same direction and have the same amount of displacement). Since the IF glide is determined by the relative positions of the poles, this model had a level-independent IF profile in its revcor function.

Figure 6(a) shows the model's revcor functions (CF = 2000 Hz) at three noise levels: 40 dB, 60 dB, and 80 dB SPL. The zero-crossing points of the revcor functions were almost identical at different SPLs; therefore, the revcor function's instantaneous frequency is independent of input SPL. Figure 6(b) shows the instantaneous frequencies for three model fibers (CF= 3000 Hz, 2200 Hz, and 550 Hz from top to bottom) at three SPLs (40 dB, 60 dB, and 80 dB). To calculate the IF of a revcor function, the envelope of the revcor function was calculated by taking the absolute value of the Hilbert transform of the revcor function. IF was then calculated over the time period where the envelope was more than one quarter of the peak value, using the zero-crossing method (see Appendix for details about calculation of instantaneous frequency). The overlap of the IF trajectories for the same revcor function at different levels verified the level-independence of IF. Generally, the slopes of the IF trajectories increased as a function of CF. The slopes were usually positive (upward) for CFs greater than 1.5 kHz and were negative (downward) for CFs less than 0.75 kHz (Carney *et al*., 1999). For the same AN fiber, or for the model at the same CF, the duration of the IF trajectory was shorter at higher levels because the time duration of the revcor function was shorter at higher levels. This phenomenon reflects the increasing bandwidth of the revcor function at higher levels.

B. Rate-level Curves

At low sound intensity, the control signal of this model was small (i.e., σ_c is almost zero and σ is near σ_0) and the filter in the signal path behaved like a linear bandpass filter with relatively narrow bandwidth and high gain. The filter output was compressed by the nonlinear control mechanism (i.e., σ_c and σ are larger than their values corresponding to sound intensity below threshold) when CF-tone levels were greater than 20 dB SPL. At very high SPLs, the control signal was nearly saturated, which made the filter output behave more linearly.

The compressive nonlinearity is illustrated in Fig. 7, which shows the root mean square (RMS) value of the signal path output (F_{out} in Fig. 1) as a function of the input sound pressure level for several CFs. The input is a 50 ms pure tone at the model's CF with 2.5ms onset and offset times. The compressive nonlinearity is stronger for higher CFs (Rhode and Cooper, 1996; Ruggero *et al.*, 1997). Figure 8 illustrates (a) the level-dependent onset rate, (b) sustained rate, and (c) synchronization coefficient of the model's responses to a pure tone at CF. (The left and right columns are results for model CFs at 1100 Hz and 4000 Hz respectively.) Both the onset rate and the sustained rate increase as the input SPL increases. The dynamic range of the onset rate is larger than that of the sustained rate (about 40 dB), which is appropriate for AN fibers (Smith, 1988).

C. Tuning Curves and Q10 Values

Tuning curves represent the excitation threshold of an AN fiber to tones at different frequencies (Kiang *et al.*, 1965) and thus quantify the relative sensitivity of the AN fiber to various tone frequencies. The stimulus used to measure the model tuning

111

curves was a 50-ms tone followed by 50 ms of silence. The threshold is defined as the

sound pressure level at which the average discharge rate during the 50-ms tone is 10

spikes/second greater than the average discharge rate during silence. Figure 9 (a) shows

tuning curves for the AN model at different CFs. The thresholds at CF were set between

0 and 10 dB by adjusting the gain in the IHC model. This threshold can also be adjusted

by changing the gain in the middle-ear model. The width of model tuning curves depends

on the bandwidth of the bandpass filter in the signal path. The model tuning curves lacks

explicit "tails" on the low frequency side as observed in physiological data (Kiang and

Moxon, 1974; Kiang, 1975; Liberman, 1978). However this model is relatively sensitive

to frequencies near sub-harmonics of CF.

Q10 value (CF divided by the bandwidth 10 dB above threshold) is a standard

measurement of the sharpness of the tuning curves. Model Q10s were comparable with

Q10 values measured from normal cat AN fibers (Miller *et al.*, 1997). The tuning of this

AN model was adjusted to match model Q10 values to AN fibers with tuning sharper

than 75% of the population data.

D. Response Areas with Phase Responses

The upper panel of Fig. 10 shows average discharge rate for an AN model fiber

(CF = 2200 Hz) with tone stimuli at several frequencies and levels. Each curve

corresponds to responses to tones at a constant sound pressure level (iso-level contours).

The peak of the curve shifts to lower frequencies as SPL increases. This peak shift is also

seen in physiological data (e.g., Anderson *et al.*, 1971) and illustrates the change in best

frequency as a function of sound pressure level. This peak shift is not seen in Zhang *et al.*

(2001, their Fig. 9) because the gammatone filter is essentially a symmetric filter in frequency domain.

The lower panel of Fig. 10 shows the level-dependent phase shift for fiber responses to pure tones with frequencies above and below model CF, referenced to the phase in response to tones at 90 dB SPL. Thus, any nonzero relative phase indicates that phase changes with level. The opposite phase change above and below CF is consistent with physiological data (Anderson *et al.,* 1971). However, the maximum negative phase change at frequencies below model CF is about $\pi/4$, which is smaller than the value of $\pi/2$ seen in AN fibers (Anderson *et al.*, 1971). This is a limitation of this model, as discussed below.

E. Synchronization Coefficient of Model AN Fibers to CF Tones as a Function of CF.

Synchronization coefficient measures how well the AN response synchronized to the input pure tone in time domain. A synchronization coefficient of one means that the AN response is perfectly tuned (phase-locked) to the input pure tone. Figure 11 illustrates the maximum synchronization coefficient of this model's response to pure tones as a function of CF. The model response's synchronization is an important temporal property that indicates how well the AN response preserves the fine structure of input sound in the time domain. This ability of the AN model to phase lock to the fine structure of input sound is limited by the lowpass filtering in the IHC model. The parameters of the lowpass filter in the IHC and synapse model (Zhang *et al.*, 2001) were chosen to achieve the lowpass roll-off seen in data (Johnson, 1980). The model's synchronization coefficient is slightly smaller compared with data (Fig. 11) from cat (Johnson, 1980), due to the

113

limitation of the synapse model. This limitation is also shown in Zhang *et al.* (2000), which has the same synapse model.

F. Two-tone Suppression and Suppression Growth Functions

Two-tone suppression (Nomoto *et al.*, 1964; Sachs and Kiang, 1968; Delgutte, 1990) is a nonlinear phenomenon of AN fiber responses, in which a stimulus away from CF can act to reduce the response to a stimulus near CF. It was included in the model by making the bandwidth of the control path wider than that of the signal path, as suggested by Geisler and Sinex (1980). The suppressor passes through the relatively wideband control path and decreases the gain of the signal path. This feature is especially important for the simulation of AN responses to complex sounds, such as speech signals, where more than one frequency component is present. This suppression can be quantified by the suppression tuning curve (Fig. 12), which measures suppression threshold as a function of suppressor frequency. The suppression threshold is defined as the SPL of the suppressor when the response to a CF tone is reduced by ten spikes/second (Delgutte, 1990). The tip of the suppression tuning curve was shifted toward the high frequency side of the model tuning curve (Delgutte, 1990). To implement this tip shift, the center frequency of the wideband filter in the control path was set to be higher than the model CF. This selection of the wideband filter's center frequency (a higher value than CF) is also in agreement with the suggestion that the outer hair cells responsible for the

nonlinearity of a group of IHCs are located basally (i.e., tuned to a higher frequency) from the IHCs (Patuzzi, 1996).

While the suppression tuning curves describe the frequency tuning of the onset of suppression, it is also important to examine how suppression grows with level for frequencies above and below CF. The suppression growth function measures the amount of suppression at a particular suppressor frequency as a function of the SPL of the suppressor. The SPL of the CF tone was adjusted such that the response was constant as the suppressor SPL increased. Different slopes are shown for suppression growth functions at different suppressor frequencies for the same CF (Fig. 13): The suppressors with frequencies higher than CF show slower growth (shallower slope) than suppressors with frequencies below CF do. This asymmetrical growth in two-tone suppression has been observed previously in physiological experiment (Delgutte, 1990). For example, in Fig. 13 (CF = 3500 Hz) the slope for growth of suppression by a 1550 Hz tone is 0.8 dB/dB, which is smaller than the averaged physiological data (about 1.3 for suppressor frequency/CF = 0.44, Fig. 9 of Delgutte, 1990). The slope for growth of suppression by a 4400 Hz tone is 0.4 dB/dB, which agrees with the slope observed physiologically for a suppressor frequency 1.26 times CF (Delgutte, 1990).

IV. DISCUSSION

This report describes a computational AN model that has a level-independent frequency glide and nonlinear compression. AN models based on gamma-tone filters (Carney, 1993; Zhang *et al*., 2001) do not have frequency glides in their impulse responses because the frequency response of the gamma-tone filter is symmetrical. Pfeiffer's (1970) BPNL model and its extension, Goldstein's (1990, 1995) MBNL model, did not address the level-independence of the instantaneous frequency glide. Meddis *et*

al. (2001) qualitatively described the frequency glide in their model's impulse responses, but they did not quantify this response property or examine its level dependence. We explored the level dependence of the frequency glide in the DRNL model by producing impulse responses for a 2000 Hz CF fiber at several input sound pressure levels (Fig. 14). The instantaneous frequency profile of the DRNL model changes considerably as a function of level, which is inconsistent with the level-independent frequency glide reported for AN fibers (e.g., Fig 3 in Carney *et al*., 1999, and cf. Fig. 6 for the model presented here). The instantaneous frequency of the DRNL model has the same downward glide at the two lowest levels tested (44 and 64 dB SPL) and has an upward frequency glide at higher SPLs.

Irino and Patterson (2001) demonstrated an instantaneous frequency glide in the impulse response of their gammachirp-based model. However, the trend of this model's best-frequency shift, which is a feature associated with the frequency glide in the impulse response, is not consistent with physiological results. In Fig. 7 of Irino and Patterson (2001), best frequency (CF = 2000Hz) first shifts to a relatively higher frequency and then shifts down as the input SPL increases from 30 dB to 80 dB. In Fig. 10 of Irino and Patterson (2001), the best frequency increases as the input SPL changes from 30 dB to 60 dB. However, Anderson *et al*. (1971, their Fig. 8) show that the best frequency in responses of AN fibers with mid- to high-frequency CFs shifts to lower frequencies as the input SPL increases. The best-frequency shift with level change can be qualitatively explained by the combination of the instantaneous frequency trend in the impulse response and the change in the shape of the impulse response envelope at various input

SPLs (Carney *et al.*, 1999). The instantaneous frequency tends to have an upward glide in fibers with CFs higher than 1500 Hz, which means that the beginning of the response has a relatively lower instantaneous frequency. When the input SPL is higher, the impulse response has a shorter duration in the time domain, which means that the early part of the impulse response is emphasized and the best frequency at high SPLs is relatively lower.

A. Limitations

The goal of this study was to provide a computational phenomenological AN model with more complete response features than those of previous AN models, and efforts focused on modeling the level-independent frequency glide. The instantaneous frequency glide and compressive nonlinearity were successfully included in this model. This model does have some limitations, however. Only AN fibers with high spontaneous rate have been implemented to date; other spontaneous rate fibers will be considered in future work. The major changes related to modeling low and medium spontaneous rate fibers would be mostly within the IHC and synapse model and it is anticipated that the configuration of the signal path and the control path described here will not be changed.

The values of this model's parameters were estimated on the basis of revcor functions recorded from low-frequency AN fibers (Carney *et al.*, 1999, below 4000 Hz), which limits the application of this model in processing signals with relative higher frequency components. Revcor data based on BM measurements are available for higher CFs (de Boer and Nuttall, 1997). These BM revcor data could be used for parameter estimation at high CFs by calculating model revcor functions from the signal path output and then fitting these model revcor functions to BM data.

117

Another limitation is the relatively small phase shift below CF in the intensity-dependent phase responses (Fig. 10). The limited phase shift is due to the limited number of poles in the signal-path filter and may also be due to the level-independence of the zeros' locations. These limitations can be overcome with a more complicated AN model; however, the results presented here focused on a relatively simple model structure.

ACKNOWLEDGMENTS

APPENDIX

Calculation of instantaneous frequency

Two methods were used to calculate the instantaneous frequency of a model or data revcor function, s(t). The first method was based on the Hilbert transform. A complex signal is made by taking s(t) as the real part and the Hilbert transform of s(t) as the imaginary part. The time-domain derivative of the phase of this complex signal is the instantaneous frequency of the revcor function s(t). The second method was based on the zero-crossing points in the revcor functions. Each pair of consecutive zero-crossing points in a revcor function s(t) are assumed to have a time difference of one half of the instantaneous period. (The instantaneous frequency is the reciprocal of the instantaneous period.) Before estimating the instantaneous frequency with either method, a three-point average of the original revcor function was performed to smooth the revcor function. Also, the instantaneous frequency profile was only estimated over a time interval for which the envelope amplitude was at least 25% of the maximum value (de Boer and Nuttall, 1997), which avoided noisy fluctuations where the revcor function amplitude was too small to make accurate estimates. In general, these two methods gave similar results; however, the method based on the Hilbert transform was more sensitive to noise in the revcor function and therefore showed more oscillation in the results. The results presented in Figs. 6 and 14 are therefore based on the zero-crossing method.

[1]Best frequency (BF) is the frequency at which the fiber response is strongest at a certain SPL (peak frequency of the response area curve, see Fig. 10). The characteristic frequency (CF) is the frequency at which the threshold of the fiber is lowest.

[2]To get real responses in the time domain after the inverse Laplace transform, poles were arranged in conjugate pairs in the Laplace domain.

REFERENCES:

Anderson, D. J., Rose, J. E., Hind, J.E., and Brugge, J.F. (1971). "Temporal position of

    discharges in single auditory nerve fibers within the cycle of a sine-wave

    stimulus: Frequency and intensity effects," J. Acoust. Soc. Am. 49, 1131-1139.

Arthur, M. A., Pfeiffer, R. R., and Suga, N. (1970). "Properties of 'Two-tone inhibition' in

    primary auditory neurons," J. Physiol. 212, 593-1139.

Carney, L.H. (1993). "A model for the responses of low-frequency auditory nerve fibers in cat,"

    J. Acoust. Soc. Am. 93:401-417.

Carney, L. H., McDuffy, M. J., and Shekhter, I. (1999). "Frequency glides in the impulse

    responses of auditory-nerve fibers," J. Acoust. Soc. Am. 105, 2384-2391.

Carney, L.H. and T.C.T. Yin (1988). "Temporal coding of resonances by low-frequency auditory

    nerve fibers: Single fiber responses and a population model," J. Neurophysiol. 60:1653-

    1677.

Cheatham, M. A., and Dallos, P. (1999). "Response phase: A view from the inner hair cell," J.

    Acoust. Soc. Am. 105, 799-810.

Cheatham, M. A., and Dallos, P. (2001). "Inner hair cell response patterns: implications

    for low-frequency hearing," J. Acoust. Soc. Am. 110, 2034-2044.

de Boer, E., and Nuttall, A. L. (1997). "The mechanical waveform of the basilar membrane. I.

    Frequency modulations ("glides") in impulse responses and cross-correlation functions,"

    J. Acoust. Soc. Am. 101, 3583-3592.

Delgutte, B. (1990). "Two-tone rate suppression in auditory-nerve fibers: Dependence on

    suppressor frequency and level," Hear. Res. 49, 225-246.

Geisler, C. D., and Sinex, D. G. (1980). "Responses of primary auditory fibers to combined noise and tonal stimuli," Hear. Res. 3, 317–334.

Goldstein, J. L. (1990). "Modeling rapid wave form compression on the basilar membrane as multiple-band-pass-nonlinearity filtering," Hear. Res. 49, 39-60.

Goldstein, J. L. (1995). "Relations among compression, suppression, and combination tones in mechanical responses of the basilar membrane: data and MBPNL model," Hear. Res. 89, 52-68.

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

Irino, T. and Patterson, R.D. "A time-domain, level-dependent auditory filter: the gammachirp," J. Acoust. Soc. Am. 101, 412-419, 1997.

Irino, T. and Patterson, R.D. "A compressive gammachirp auditory filter for both physiological and psychophysical data," J. Acoust. Soc. Am. 109 (5), 2008-2022, 2001.

Johnson, D. H. (1980). "The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones," J. Acoust. Soc. Am. 68,1115-1122.

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 Research Monograph No. 35 (MIT, Cambridge, MA).

Kiang, N.Y.S., and Moxon, E.C. (1974). "Tails of tuning curves of auditory-nerve fibers," J. Acoust. Soc. Am. 55, 620-630.

Liberman, M. C. (1978). "Auditory-nerve responses from cats raised in a low-noise chamber," J. Acoust. Soc. Am. 63, 442-455.

Lin, T. and Guinan, J.J. (2000). "Auditory-nerve-fiber responses to high-level clicks: Interference patterns indicate that excitation is due to the combination of multiple drives," J. Acoust. Soc. Am. 107, 2615-2630.

Marquardt, D. W. (1963). "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," J. Soc. for Industrial and Applied Mathematics, 11, 431-441.

Meddis, R., O'Mard, L.P., Lopez-Poveda, E.A. (2001). "A computational algorithm for computing nonlinear auditory frequency selectivity," J. Acoust. Soc. Am. 109, 2852-2861.

Moller, A. R. (1977). "Frequency selectivity of single auditory-nerve fibers in response to broadband noise stimuli," J. Acoust. Soc. Am. 62, 135-142.

Mountain, D.C., Hubbard, A. E. (1996). "Computational analysis of hair cell and auditory nerve processes," in Auditory Computation, edited by H.L. Hawkins, T.A. McMullen, A.N. Popper, and R.R. Fay (Springer-Verlag, New York), pp. 121-156.

Nomoto, M., Suga, N., and Katsuki, Y. (1964). "Discharge pattern and inhibition of primary auditory nerve fibers in the monkey," J. Neurophysiol. 27, 768-787.

Patuzzi, R. (1996). "Cochlear micromechanics and macromechanics," in The Cochlea, edited by P. Dallos, A.N. Popper, and R.R. Fay (Springer-Verlag, New York), pp. 186-257.

Pfeiffer, R.R. (1970). "A model for two-tone inhibition of single cochlear-nerve fibers,"

J. Acoust. Soc. Am. 48, 1373.

Recio, A., Narayan, S. S., and Ruggero, M. A. (1996). "Wiener-kernel analysis of basilar membrane response to noise," in Diversity in Auditory Mechanics, edited by E. R. Lewis, G. R. Long, R. F. Lyon, P. M. Narins, C. R. Steele, and E. Hecht-Poinar (World Scientific, Singapore), pp. 325-331.

Recio, A., Narayan, S. S., and Ruggero, M. A. (1998). "Basilar-membrane responses to clicks at the base of the chinchilla cochlea," J. Acoust. Soc. Am. 103, 1972-1989.

Rhode, W.S.(1971). "Observations of the vibration of the basilar membrane in squirrel monkeys using the Mossbauer technique," J. Acoust. Soc. Am. 49, 1218-1231.

Rhode, W. S., and Cooper, N.P. (1996). "Nonlinear mechanics in the apical turn of the chinchilla," Aud. Neurosci. 3, 101-120.

Robert, A., Eriksson, J. L. (1999). "A composite model of the auditory periphery for simulating responses to complex sounds," J. Acoust. Soc. Am. 106, 1852-1864.

Robles, L., Rhode, W. S., and Geisler, C. D. (1976). "Transient response of the basilar membrane measured in squirrel monkeys using the Mossbauer effect," J. Acoust. Soc. Am. 59, 926-939.

Rosowski, J.J. (1996). "Models of External- and Middle-Ear Function," in Auditory Computation, edited by H.L. Hawkins, T.A. McMullen, A.N. Popper, and R.R. Fay (Springer-Verlag, New York), pp. 15-61.

Ruggero, M. A., Rich, N. C. (1991). "Furosemide alters organ of corti mechanics: evidence for feedback of outer hair cells upon the basilar membrane," J. Neurosci. 11, 1057-1067.

Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (1997). "Basilar-membrane responses to tones at the base of the chinchilla cochlea," J. Acoust. Soc. Am. 101, 2151-2163.

Sachs, M.B., and Kiang, N.Y.S. (1968). "Two-tone inhibition in auditory-nerve fibers," J. Acoust. Soc. Am. 43, 1120-1128.

Shekhter, I. and Carney, L.H. (1997). "A nonlinear auditory nerve model for CF-dependent shifts in tuning with sound level," Assoc. for Res. In Otolaryngol. 20:617.

Shera, C.A. (2001a) "Frequency glides in click responses of the basilar membrane and auditory nerve: Their scaling behavior and origin in traveling-wave dispersion," J. Acoust. Soc. Am. 109, 2023-2034.

Shera, C.A. (2001b) "Intensity-invariance of fine time structure in basilar-membrane click responses: Implications for cochlear mechanics," J. Acoust. Soc. Am. 110, 332-348.

Smith, R. L. (1988). "Encoding of sound intensity by auditory neurons," in *Auditory Function: Neurobiological Bases of Hearing*, edited by G.M.Edelman, WE.Gall, and W.M.Cowan (Wiley, New York), pp. 243-274.

Tan, Q and Carney, L.H. (1999). "A phenomenological model for auditory nerve responses: Including the frequency glide in the impulse response." Proc. IEEE 25th Annual. Northeast Bioengineering Conference, pp. 23-24.

Westerman, L.A., Smith, R.L. (1988). "A diffusion model of the transient response of the cochlear inner hair cell synapse," J. Acoust. Soc. Am. 83, 2266-2276.

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," J. Acoust. Soc. Am. 109, 648-670.

Table 1. Parameter Values

Parameter	Description	value
CF	Characteristic frequency	
Middle ear model		
Pm1_real	Real part of pole 1 in middle ear model (Hz)	-250
Pm1_img	Imaginary part of pole 1 in middle ear model (Hz)	400
Pm2_real	Real part of pole 2 in middle ear model (Hz)	-2000
Pm2_img	Imaginary part of pole 2 in middle ear model (Hz)	6000
Zm	Location of the zero on real axis for middle ear model (rad/s)	-200
Control path		
F_{cwb}	Center frequency of the wide-band filter (Hz)	1.2mm basal to fiber CF
BW_{wb}	Bandwidth of the wide-band filter (Hz)	CF/4
Acp	Parameter in the first nonlinear function	100
Bcp	Parameter in the first nonlinear function	2.5
Ccp	Parameter in the first nonlinear function	0.6
S_0	Parameter in the second nonlinear function	8.0
S_1	Parameter in the second nonlinear function	3.0
T_0	Parameter in the second nonlinear function	0.85
T_1	Parameter in the second nonlinear function	5.0
Fclp	Cut-off frequency of the low-pass filter (Hz)	800
σ_c	Output of the control path	
Signal path		
Pa	Relative locations of poles, real part in the bandpath filter of the signal path	See Eq. 9
Pb	Relative locations of poles, imaginary part	See Eq. 10
$P\omega$	Imaginary part of the pole closest to the imaginary axis	See Eq. 12
σ	Real part of the pole closest to the imaginary axis	$\sigma = \sigma_0 - \sigma_c$
σ_0	Real part of the pole closest to the imaginary axis at quiet	See Eq. 11
Z_0	Parameter for the location of the zeros in signal path	0.9
Z_1	Parameter for the location of the zeros in signal path	-1.5

Inner hair cell and Synapse model
See Zhang *et al*. (2001)

Figure captions:

Figure 1. Schematic diagram of the AN model. The model included a middle-ear model, a signal path, a control path, the IHC and synapse model, and a time delay. See text for more detail.

Figure 2. A simple example of pole-zero locations for a bandpass filter, which has two pairs of poles (Pa1/Pa2 and their conjugates) and one zero (z) in control space. The relative locations of the poles and zeros affect the trend of the instantaneous frequency profile in the filter's impulse response. In this example, Pa2 has a larger damping coefficient and smaller resonance frequency than Pa1 does. Therefore, the beginning part of the impulse response is dominated by relatively lower frequency (b2) and the later part is dominated by higher frequency (b1).

Figure 3. Pole-zero locations for the bandpass filter in the model's signal path. Ten pairs of poles (P11 and P13 were fourth-order poles and P12 was a second-order pole) and a tenth-order zero were included. P01, P02, and P03 are the pole locations when the input sound intensity is zero for the poles P11, P12, and P13 respectively.

Figure 4. An example of the model revcor fitted to cat data. The thin line is a revcor function from cat [unit 15 from cat 86166 (Carney *et al.*, 1999)] and the bold line is the corresponding model result. The CF of this fiber is 650 Hz. A downward frequency glide is apparent in these impulse responses, as the zero crossings are increasingly separated at later times in the responses.

Figure 5. Parameter estimation results for the pole and zero locations in the signal path for: (a) Pω (b) Pa (c) Pb and (d) Xzero (see Fig. 3 for a description of the parameters).

The fitting results for each fiber's revcor function are shown as stars. Simple expressions for each of the parameters, illustrated by the black lines, were set up and fit to the population results.

Figure 6. Revcor functions and instantaneous frequencies. (a) Model revcor functions (CF = 2200 Hz) at three SPLs (40, 60, and 80 dB). The revcor functions were scaled by a factor of 20 at 40 dB SPL and by 5 at 60 dB SPL so that the amplitudes are comparable. (b) Instantaneous frequencies calculated based on zero-crossings from model revcor functions at CFs equal to 550 Hz, 2200 Hz and 3000 Hz, each with three SPLs (40, 60, and 80 dB) of noise stimuli. The overlap of the instantaneous frequency profiles indicates that the trends of the instantaneous frequency are level independent.

Figure 7. RMS value of the signal path output (at steady state) in response to CF tones at different SPLs for model CFs at 500, 1100, 2000 and 4000 Hz, respectively. The response patterns demonstrate the compressive nonlinear nature of the signal path. RMS decreases as the model CF increases at a certain tone level, indicating a greater compression with higher CF.

Figure 8. Response rate and synchronization coefficient to CF-tone input with CF of 1100 Hz (left column) and 4000 Hz (right column). Onset rate is the maximum discharge rate during the first 10 ms and is calculated using 0.5 ms bins. The sustained rate and the synchronization coefficient were calculated in the 10 to 45 ms time window for 400 repetitions.

Figure 9. (a) Model tuning curves for different CFs. The model threshold is defined as the pure-tone SPL that results in a rate response that is 10 spike/sec greater than the

129

spontaneous rate. (b) Q10 values measured from model tuning curves and compared with physiological data from Miller *et al*. (1997, their Fig. 3). Q10 data were used to set the locations of the poles of the band-pass filter in the signal path. Q10 values quantitatively described the sharpness of the model's tuning curves as a function of CF.

Figure 10. Response areas with model CF at 2200 Hz. (a) Rate response (b) Phase response. Each line corresponds to an input tone SPL as indicated in the figure. The phase responses were referenced to the phase in response to that frequency at 90 dB SPL. The center of the model's rate response showed a shift toward the low frequency side as the input SPL increases. This shift is also shown in physiological data (Anderson et al., 1971).

Figure 11. Maximum synchronization coefficient. The results were measured by taking the maximum synchronization coefficient of model response to CF-tone inputs with SPLs from 0 dB to 100 dB (i.e., the maximum value from the curves in the bottom panels in Fig. 8). The stimuli were the same as in Fig. 8. Physiological data (Johnson, 1980) from a population of cat AN fibers are indicated with crosses.

Figure 12. Suppression threshold. The solid line illustrates the model's tuning curve with CF at 4000 Hz. The stars indicate the suppression threshold, which is defined as the suppressor tone SPL that decreases the response to CF tone by 10 spike/sec.

Figure 13. Suppression growth functions measured for a CF of 3500 Hz, with suppressors at 1550 Hz (below CF) or 4400 Hz (above CF). The CF tone SPL was adjusted to maintain a response rate that was two thirds of the maximum response rate at each suppressor SPL. The dotted line indicates a growth with slope of 1 (dB/dB).

Figure 14. Instantaneous frequencies calculated for the DRNL model's (Meddis *et al*., 2001) impulse responses for a CF of 2000 Hz at different sound pressure levels (40 to 100dB in 20dB steps) based on the zero-crossing method (see Appendix). A level-dependency in the instantaneous profile is demonstrated by the changes in the direction of the glides at different levels. Note that the instantaneous profiles for the two lowest levels are nearly identical, so the curves lie atop one another.
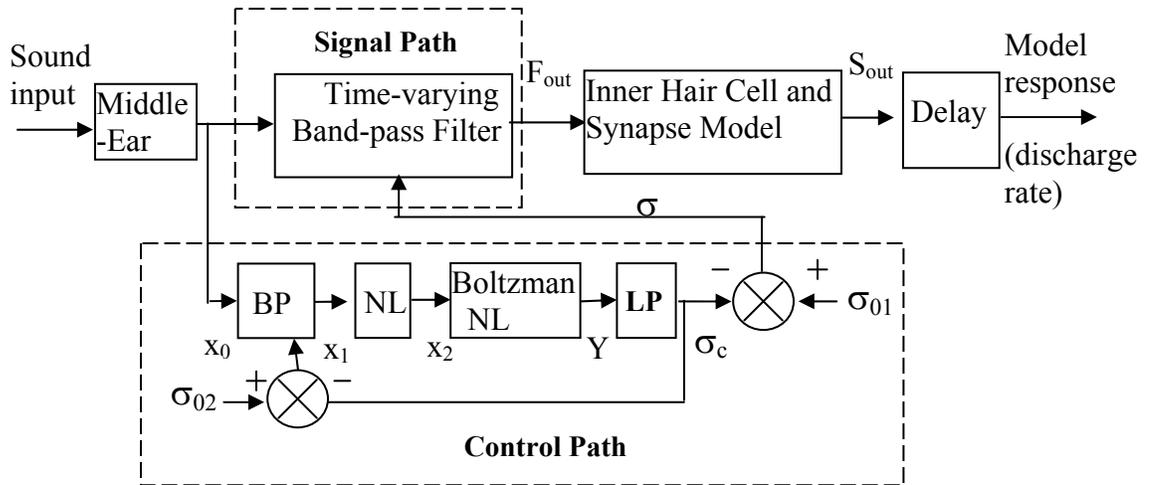
Figure 1
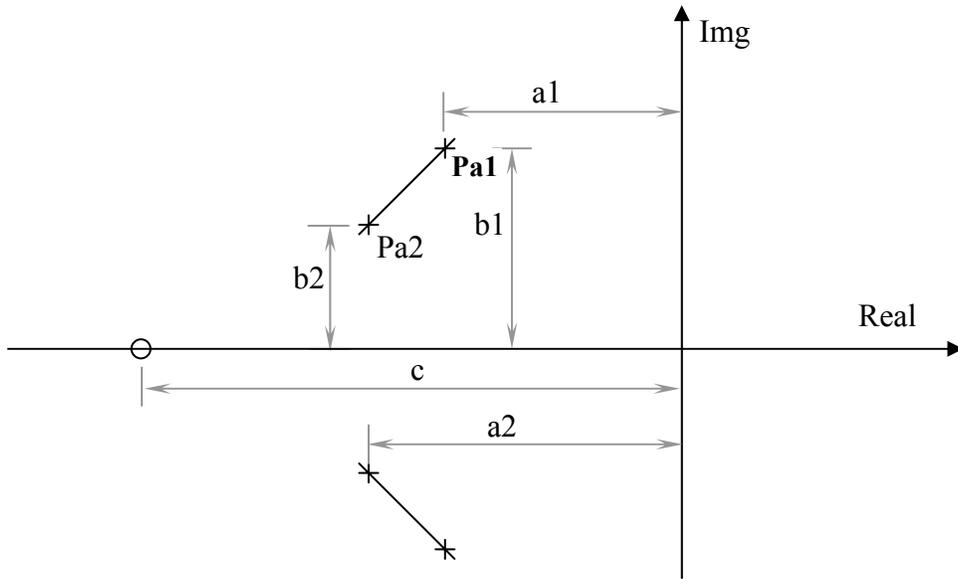
Figure 2
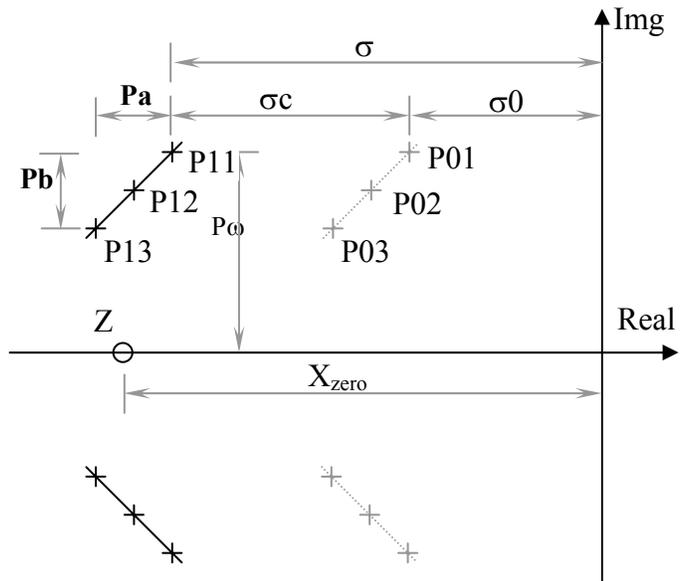
Figure 3
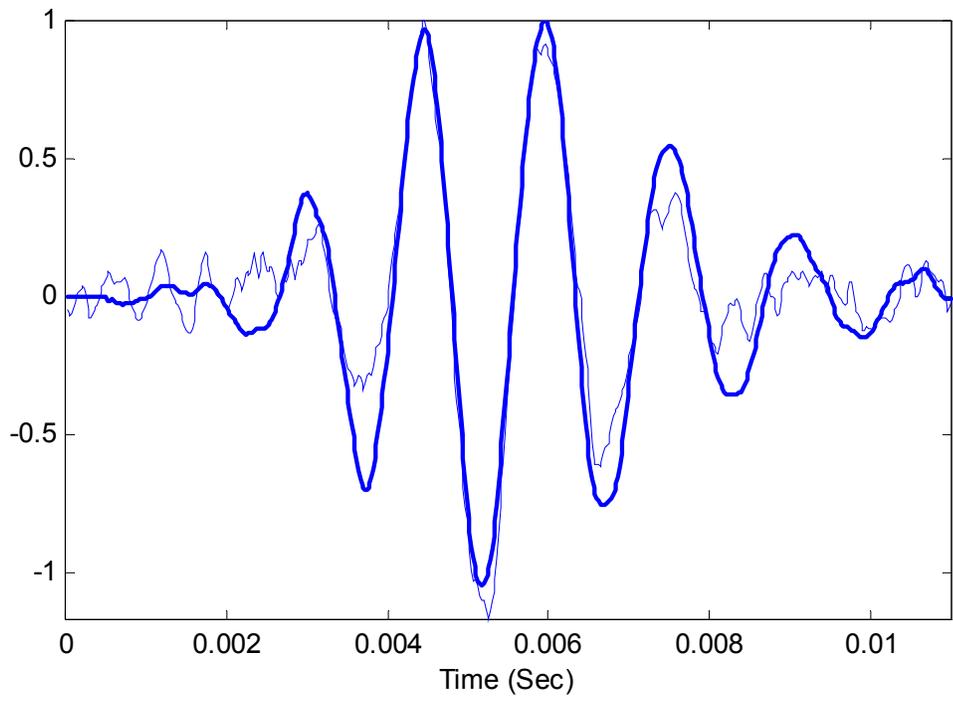
Figure 4
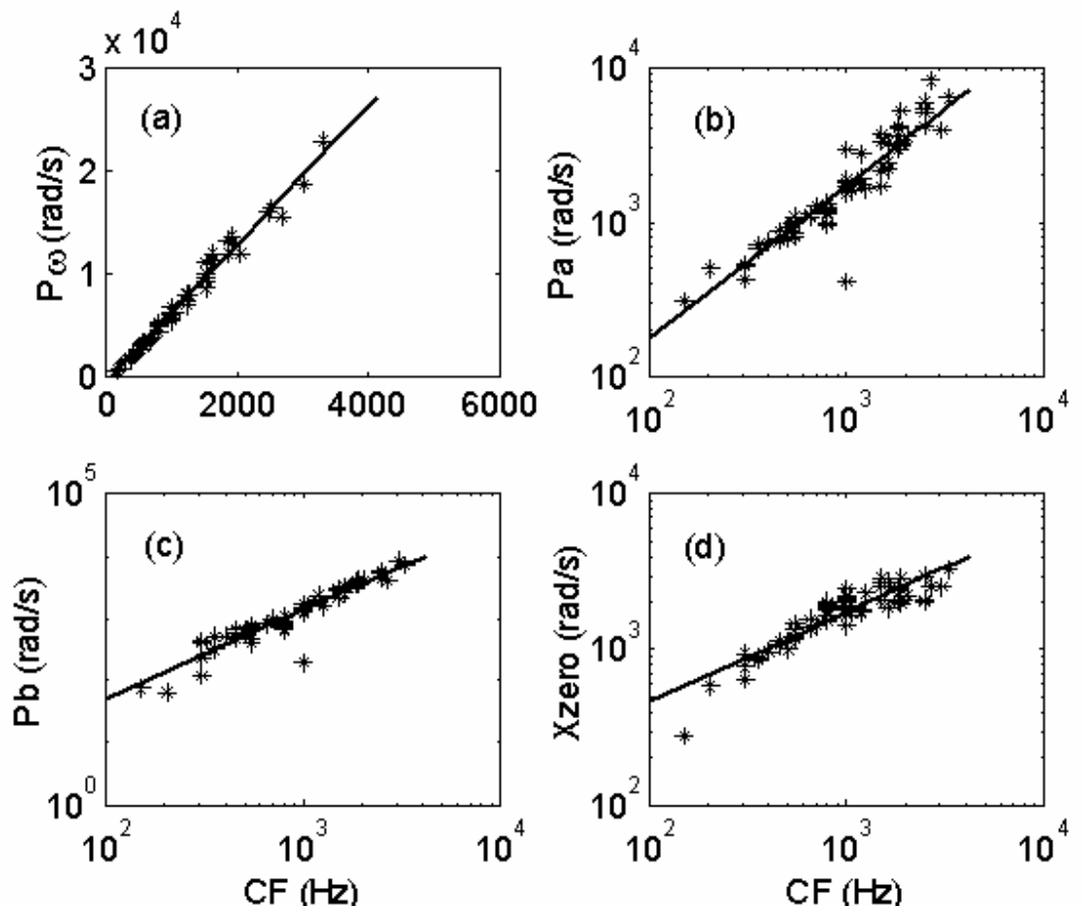
Figure 5

(a)

(b)


Figure 6

137
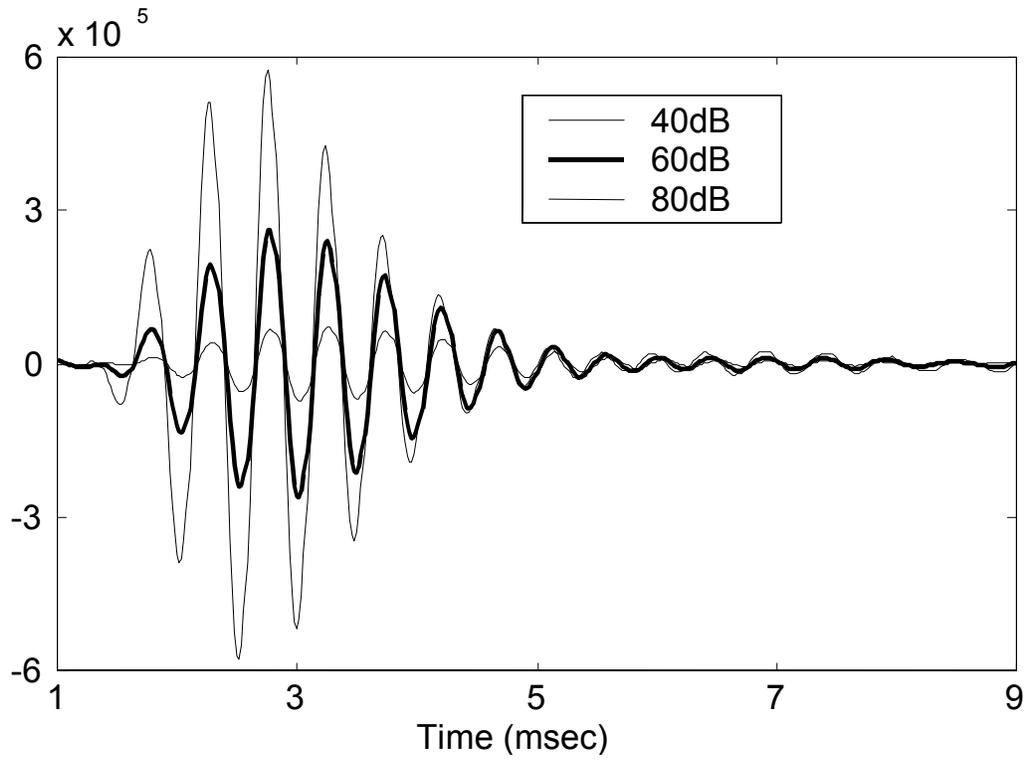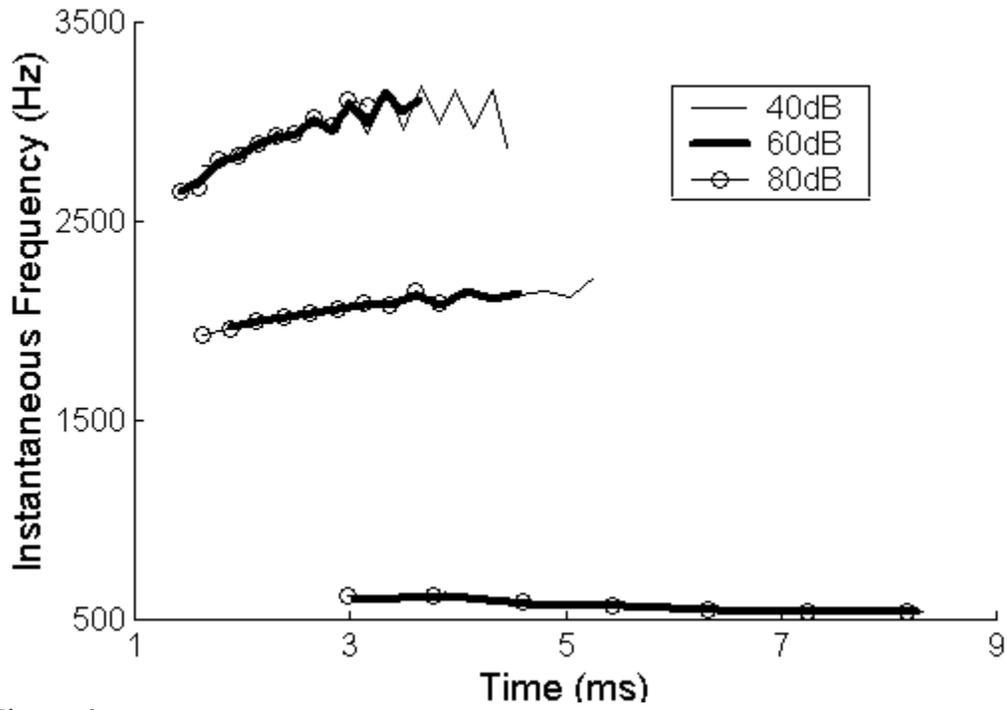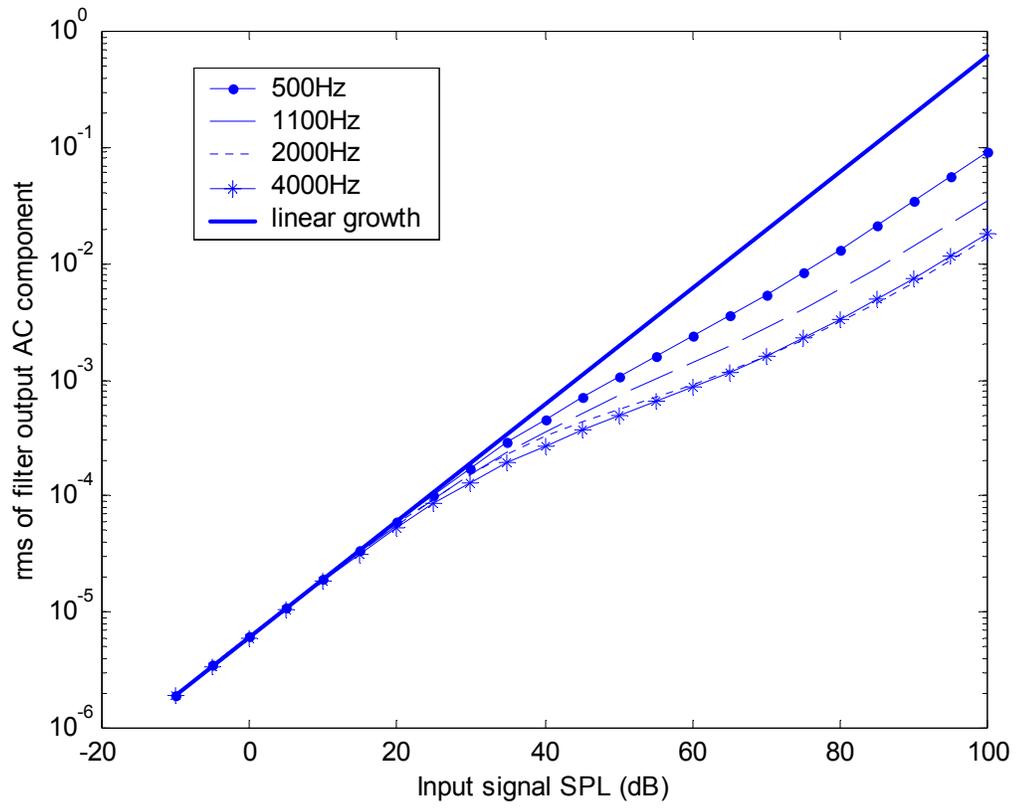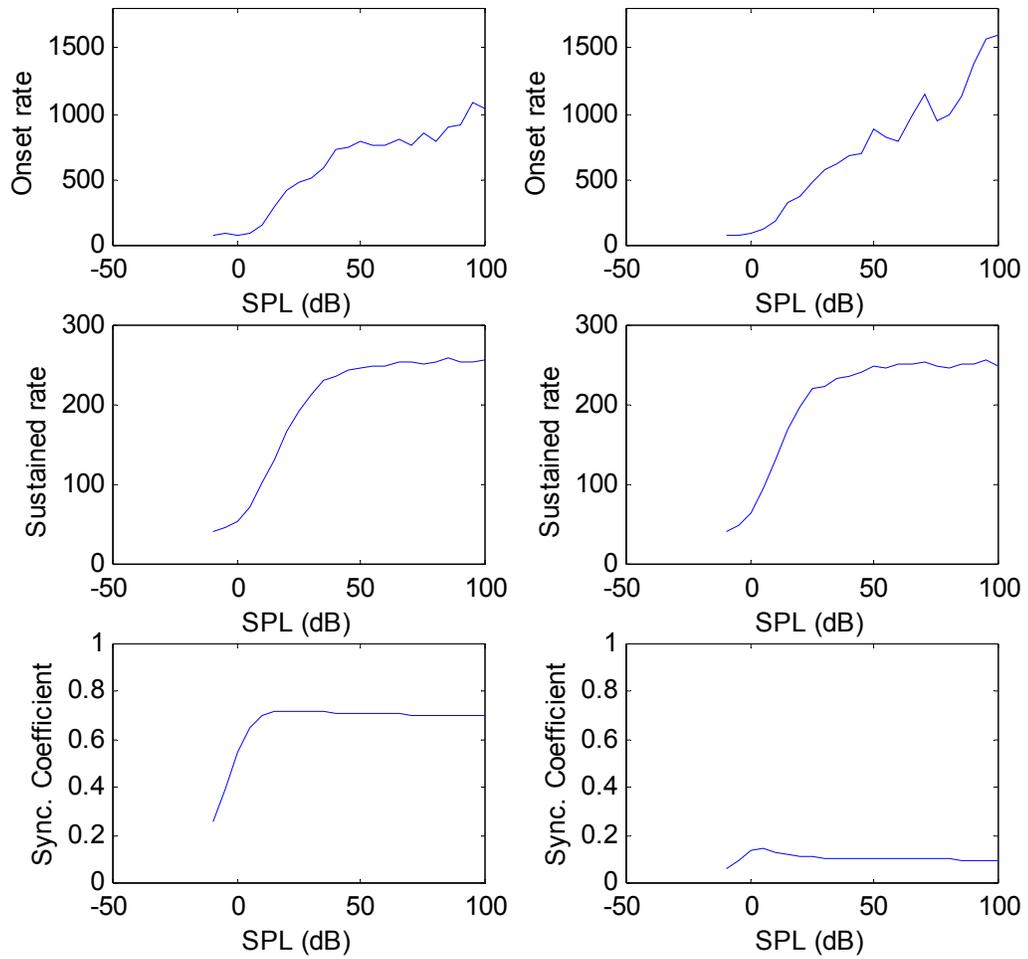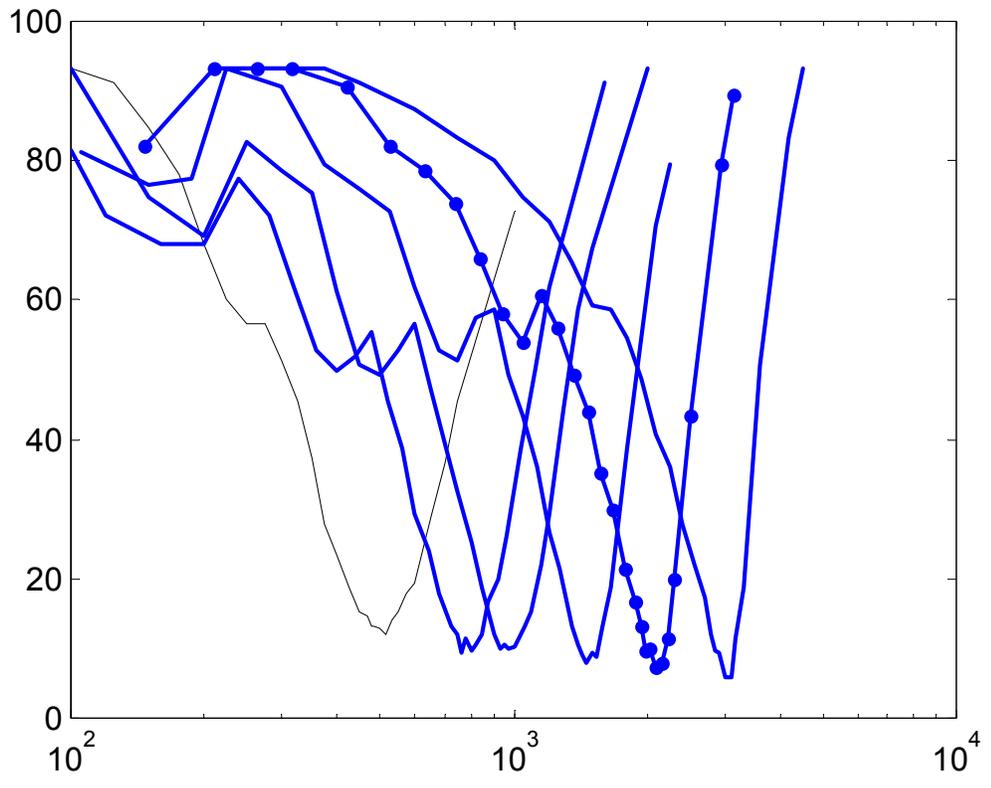
Figure 7

Figure 8.

Figure 9 (a)
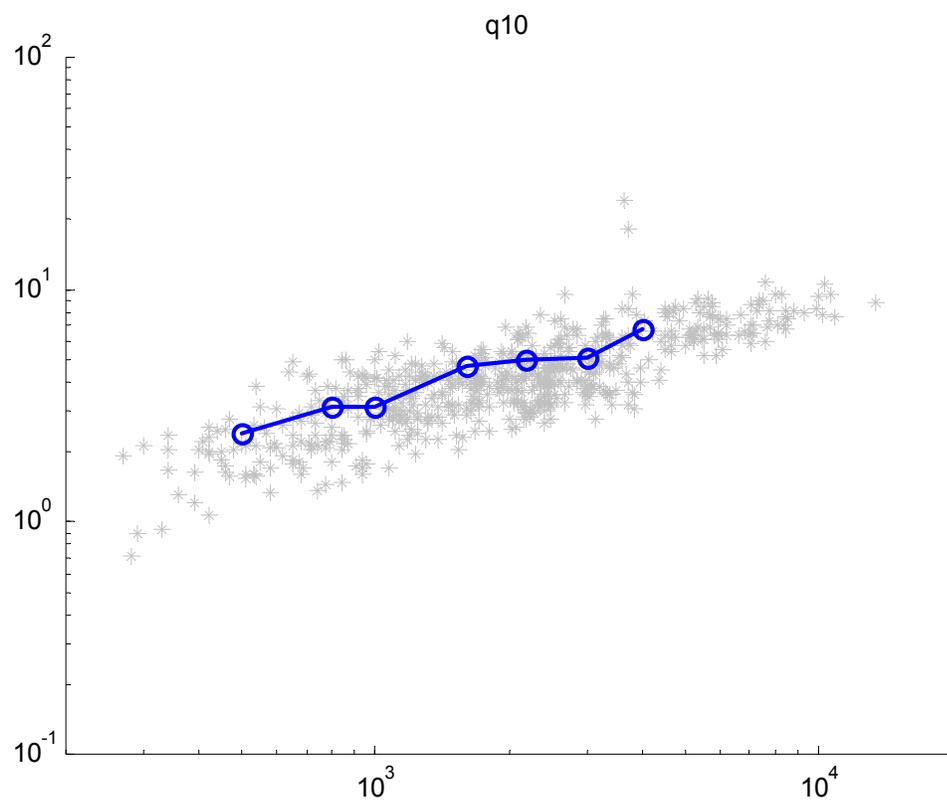
Figure 9 (b)
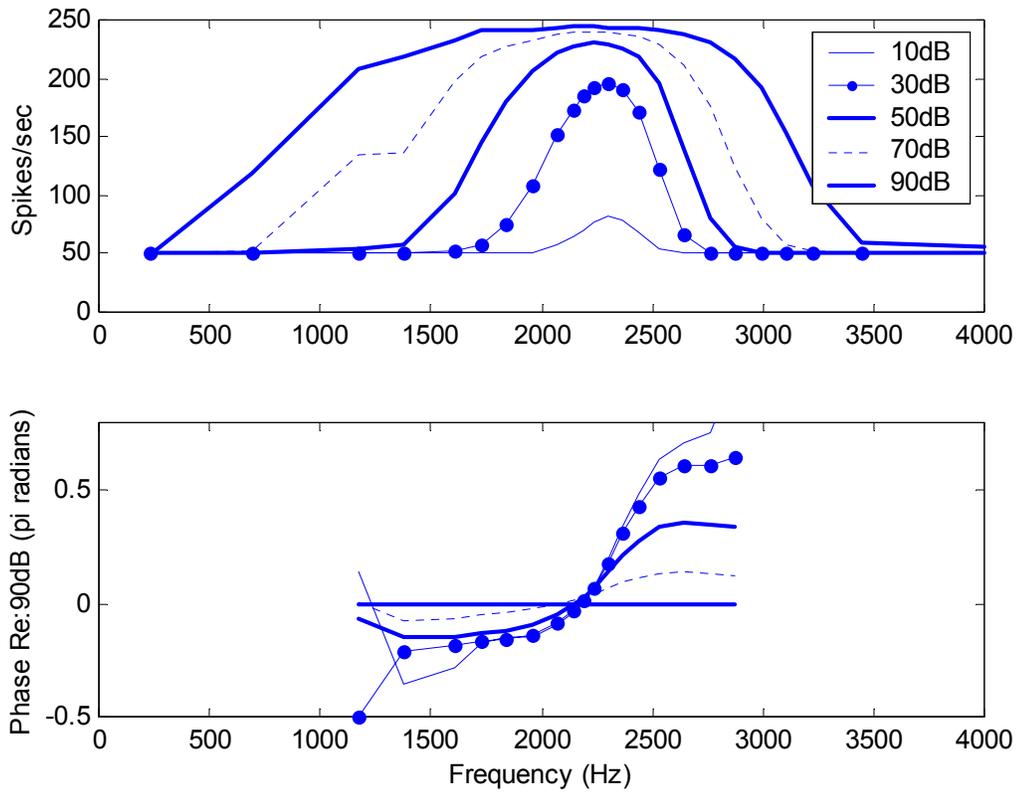
Figure 10.
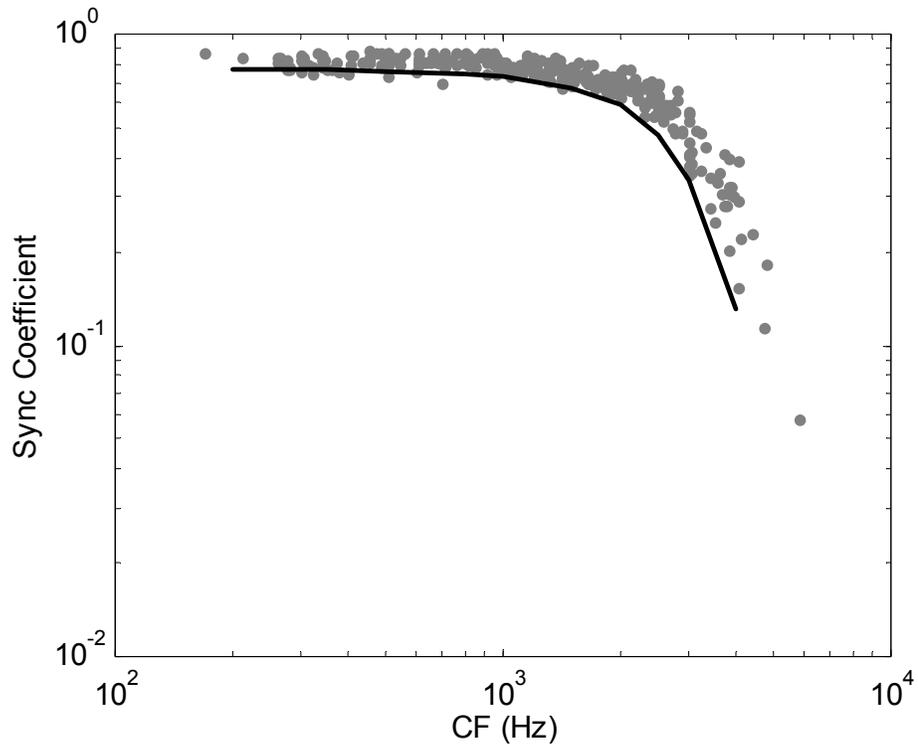
Figure 11
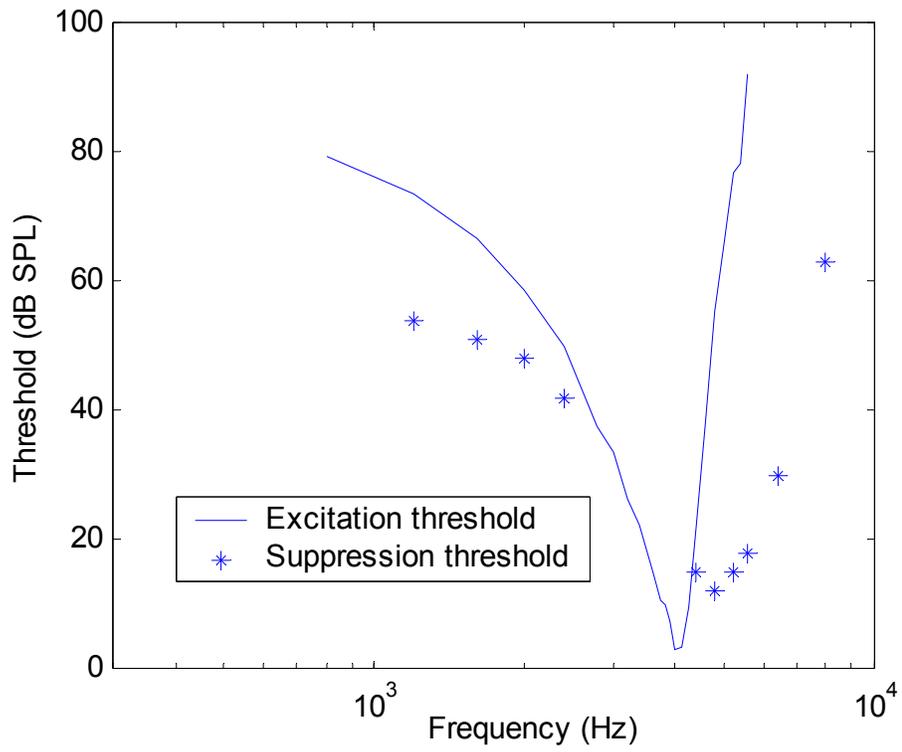
Figure 12
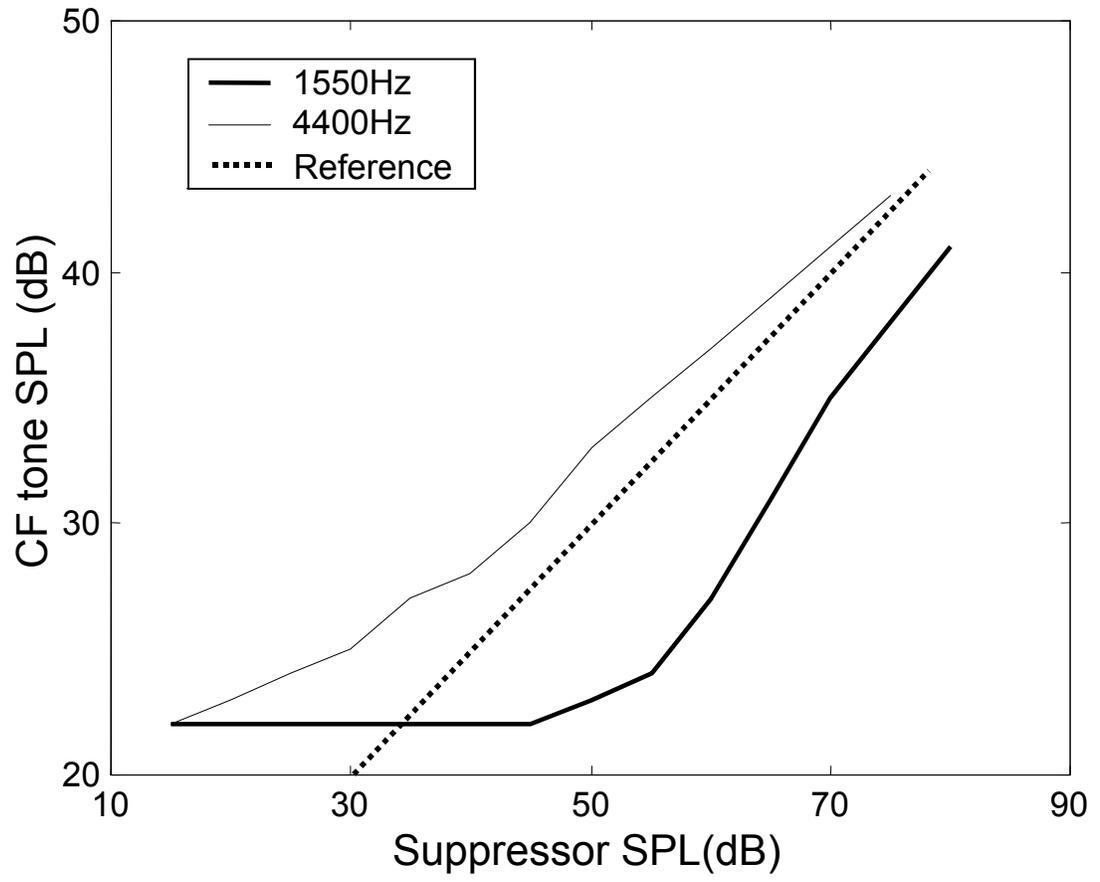
Figure 13
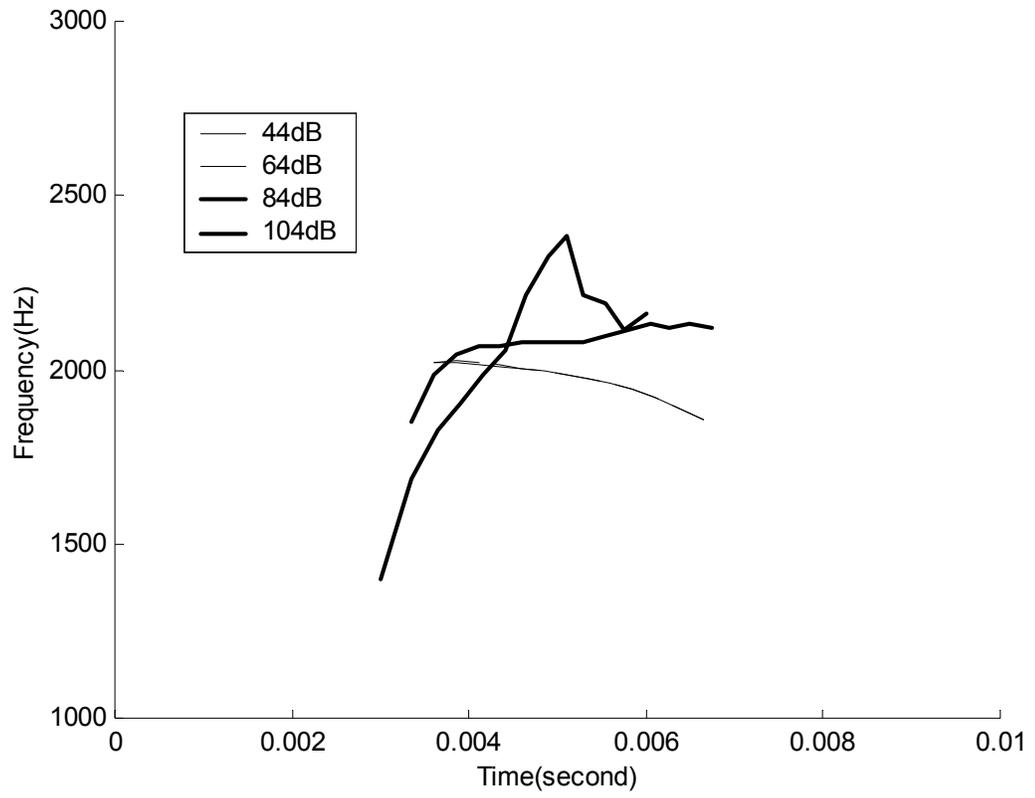
Figure 14

REFERENCES:

ABBREVIATION   FULL NAME

Biophys. J.   Biophysical Journal

Hear. Res.   Hearing Research

J. Acoust. Soc. Am.   Journal of the Acoustical Society of America

J. Neurosci.   Journal of Neuroscience

Neural Comput.   Neural Computation

Percept. Psychophys.   Perception and Psychophysics

Boer, E. de, and Nuttall, A.L. (1997). "The mechanical waveform of the basilar membrane. I. Frequency modulations ("glides") in impulse responses and cross-correlation functions," J. Acoust. Soc. Am. **101**, 3583-3591.

Burock, M. A., and Carney, L. H. (1995). "Neural encoding of level in the auditory nerve and anteroventral cochlear nucleus: A study of neural models and physiological responses," J. Acoust. Soc. Am. **97**, 3281.

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," Acustica – acta acustica **88**, 334-347.

Carney, L. H., McDuffy, M.J. and Shekhter, I. (1999). "Frequency glides in the impulse responses of auditory-nerve fibers," J. Acoust. Soc. Am. **105**, 2384-2391.

Colburn, H.S. (1969). "Some physiological limitations on binaural performance," Ph.D. Dissertatin, Massachusetts Institute of Technology, Cambridge, MA.

Colburn, H.S. (1973). "Theory of binaural interaction based on auditory-nerve data. I. General strategy and preliminary results on interaural discrimination," J. Acoust. Soc. Am. **54**, 1458-1470.

Colburn, H.S. (1977). "Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise," J. Acoust. Soc. Am. **61**, 525-533.

Cramér, H. (1951). *Mathematical Methods of Statistics* (Princeton University Press, Princeton, NJ), Chapter 32.

Feth, L. L. (1974). "Frequency discrimination of complex periodic tones," Percept. Psychophys., **15**, 375-379.

Flanagan, J.L. (1995). "A difference limen for vowel formant frequency," J. Acoust. Soc. Am. **27**, 613-617.

Greenwood, D.D. (1990). "A cochlear frequency-position function for several species – 29 years later," J. Acoust. Soc. Am. **87**, 2592-2605.

Heinz, M. G. (2000). "Quantifying the effects of the cochlear amplifier on temporal and average-rate information in the auditory nerve," Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Heinz, M. G., Colburn, H.S., Carney L.H. (2001a). "Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve," Neural Comput. **13**(10):2273-316.

Heinz, M. G., Colburn, H.S. and Carney, L.H. (2001b). "Rate and timing cues associated with the cochlear amplifier: Level discrimination based on monaural cross-frequency coincidence detection," J. Acoust. Soc. Am. **110**, 2065-2084.

Hienz, R. D., Stiles, P., and May, B. J. (1998). "Effects of bilateral olivocochlear lesions on vowel formant discrimination in cats," Hear. Res. **116**, 10-20.

Johnson, D. H., and Kiang, N. Y. S. (1976). "Analysis of discharges recorded simultaneously from pairs of auditory-nerve fibers," Biophys. J. **16**, 719-734.

Kewley-Port, D., and Watson, C.S. (1994). "Formant-frequency discrimination for isolated English vowels," J. Acoust. Soc. Am. **95**, 485-496.

Klatt, D. H. (1980). "Software for a cascade/parallel formant synthesizer." J. Acoust. Soc. Am. **67**, 971-995.

Lyzenga, J. and Horst, J. W., (1995). "Frequency discrimination of bandlimited harmonic complexes related to vowel formants," J. Acoust. Soc. Am. **98**, 1943-1955.

Lyzenga, J. and Horst, J. W., (1997). "Frequency discrimination of stylized synthetic vowels with a single formant," J. Acoust. Soc. Am. **102**, 1755-1767.

Mermelstein, P. (1978). "Difference limens for formant frequencies of steady-state and consonant-bound vowels," J. Acoust. Soc. Am. **63**, 572-580.

Patuzzi, R.B., Yates, G.K., and Johnstone, B.M. (1989). "Outer hair recepter currents and sensorineural hearing loss," Hear. Res. **42**, 47-72.

Rabiner, L.R. and Schafer, R.W. (1978). *Digital Processing of Speech Signals* (Printice-Hall Inc., Upper Saddle River, New Jersey).

Rasmussen, G.L. (1940). "Studies of the VIIIth cranial nerve in man," Laryngoscope **50**, 67-83.

Recio, A, Narayan, S.S., and Ruggero, M.A. (1997). "Weiner-kernel analysis of basilar-membrane responses to white noise," in *Diversity in Auditory Mechanics*, edited by E.R. Lewis, G.R. Long, R.F. Lyon, P.M. Narins, C.R. Steele, and E. Hecht-Poinar (World Scientific, Singapore), pp. 325-331.

Ruggero, M.A., and Rich, N.C. (1991). "Furosemide alters organ of corti mechanics: Evidence for feedback of outer hair cells upon the basilar membrane," J. Neurosci. **11**, 1057-1067.

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

Sinnott, J. M., and Kreiter, N.A. (1991). "Differential sensitivity to vowel continua in Old World monkeys (Macaca) and humans," J. Acoust. Soc. Am. **89**, 2421-2429.

Tan, Q (2000). "A nonlinear auditory-nerve model with an instantaneous frequency glide," M.S. Thesis, Boston University, Boston, MA.

Van Trees, H. L. (1968). *Detection, Estimation, and Modulation Theory: part I* (Wiley, New York), Chapter 2.

Van Zanten, G.A. (1980). "Temporal modulation transfer functions for intensity modulated noise bands," *Psychophysical, Physiological and Behavioural Studies in Hearing,* (Delft University Press), pp. 206-209.

Wier, C.C., Jesteadt, W., and Green, D.M. (1977). "Frequency discrimination as a function of frequency and sensation level," J. Acoust. Soc. Am. **61**, 178-184.

Vita

Qing Tan

Born: 1973, Tianjin, People's Republic of China

Education:

Attended Tsinghua University from September 1992 – June 1997

Received B.S. in Electrical Engineering, June 1997

Received B.E. in Economics, June 1997

Attended Boston University from September 1997 – May 2003

Received M.S. in Biomedical Engineering, May 2000

Received Ph.D. in Biomedical Engineering, May 2003