Quantifying the effects of the cochlear amplifier on temporal and average-rate information in the auditory nerve

by

Michael Gregory Heinz

Sc.B., Brown University (1992)
M.S.E., Johns Hopkins University (1994)

Submitted to the Division of Health Sciences and Technology
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Speech and Hearing Sciences

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2000

Author .....
Division of Health Sciences and Technology
May 18, 2000

Certified by .....
Laurel H. Carney, Ph.D.
Associate Professor of Biomedical Engineering, Boston University
Thesis Supervisor

Accepted by .....
Martha L. Gray, Ph.D.
Edward Hood Taplin Professor of Medical and Electrical Engineering
Co-director, Harvard-M.I.T. Division of Health Sciences and Technology

Quantifying the effects of the cochlear amplifier

on temporal and average-rate information in the auditory nerve

by

Michael Gregory Heinz

Submitted to the Division of Health Sciences and Technology on May 18, 2000,
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Speech and Hearing Sciences

ABSTRACT

An active mechanism referred to as the cochlear amplifier is believed to be responsible for sharp tuning and excellent sensitivity in the normal auditory system, and is impaired or absent in many common forms of sensorineural hearing loss. The cochlear amplifier is thought to benefit normal-hearing listeners, especially in complex listening environments in which hearing-impaired listeners have difficulty even with hearing aids.

A modeling approach was developed to relate nonlinear physiological response properties associated with the cochlear amplifier to human psychophysical performance. Quantitative methods combined analytical and computational population models of the auditory-nerve (AN) with statistical decision theory to evaluate performance limits imposed by the random nature of AN discharges (modeled by a nonstationary Poisson process). A new theoretical approach was developed to predict performance for psychophysical tasks that use random-noise stimuli to mask signal information. The ability of temporal and average-rate information in the AN to account for human performance was evaluated for several listening tasks for which the cochlear amplifier has been suggested to be significant.

The benefit of the cochlear amplifier for extending the auditory system's dynamic range was evaluated in terms of AN information available for encoding changes in stimulus level. An analytical model included nonlinear gain, level-dependent phase, and high-, medium- and low-spontaneous-rate AN fibers. Both nonlinear-phase and average-rate information were required to encode levels up to 120 dB SPL based on a narrow range of AN characteristic frequencies. A physiologically realistic mechanism to decode nonlinear gain and phase cues is monaural, across-frequency coincidence detection. Level-discrimination performance of a coincidence-counter population matched human performance across the entire dynamic range of hearing at both low and high frequencies. Results suggest that the cochlear amplifier is beneficial for encoding sound level within narrow frequency regions, and has only a small influence on simple listening tasks in quiet.

The cochlear amplifier alters tuning within the normal auditory system based on the spectral and temporal configuration of the stimulus. The influences of compression and suppression on psychophysical measures of auditory frequency selectivity were evaluated. Implications for the interpretation of psychophysical methods for estimating auditory filters are discussed.

Thesis Supervisor: Laurel H. Carney, Ph.D.
Title: Associate Professor of Biomedical Engineering, Boston University

Acknowledgments

There are many people to whom I owe thanks for their contributions to my dissertation work. I would like to thank the members of my thesis committee for extremely valuable and enjoyable interactions throughout the course of this dissertation. Their continual encouragement to focus this work on fundamental issues of broad interest has provided an important lesson to me, and has made it possible for me to finish in a reasonable amount of time.

Don Eddington's influence on my work began when I was a 4th-semester student in his research-methods class. He emphasized the benefits of formulating a good thesis topic and proposal, the importance of considering relevant clinical issues, and the often forgotten fact that working on a dissertation should be an enjoyable experience.

Lou Braida, the chairman of the committee, also made an early and profound contribution to this work through his auditory-perception class, which introduced me to the method of combining statistical decision theory with auditory-nerve models to evaluate psychophysical performance limits. Lou has also helped to make me aware of specific conditions in which hearing-impaired listeners have the most difficulty, even with hearing aids.

Andrew Oxenham provided valuable input to this work by continually reminding me of the important contributions that have been made to understanding the significance of cochlear nonlinearity through psychophysics, and of the many experiments for which hearing-impaired listeners demonstrate near-normal performance.

Bertrand Delgutte made an extremely valuable contribution through his critical evaluations of this work, both in committee meetings as well as in conversations with me. He continually reminded me of the limitations of the auditory-nerve models used in this dissertation, as well as the new contributions of this work. I have greatly appreciated his willingness to critically read manuscripts, which were always too lengthy.

I consider myself to be extremely fortunate to have had the opportunity to work so closely with both Steve Colburn and Laurel Carney during this dissertation. Steve's willingness to meet with Laurel and me almost every week is a primary reason that this work has progressed to its current state. Steve is equally as kind as he is brilliant. His intuition about mathematics and the auditory system has strongly influenced my way of thinking, and the importance of his previous work to this dissertation is immeasurable.

Finally, it is difficult to describe the profound influence Laurel Carney has had on this dissertation, as well as on me as a person. She has contributed invaluable guidance to every aspect of this dissertation, while giving me unrestricted freedom to work on issues that were of interest to me. Through her willingness to meet with me far more frequently than she had time for, she pushed me to continually move forward. Laurel's encouragement to finish the dissertation quickly while maintaining rigor has provided valuable lessons on the practical issues of doing quality research. The ability to combine her enthusiasm for both learning and teaching with a genuine concern for her students has made her a great role model. Laurel's advice and friendship in both academic and personal settings has made working with her an experience that I will never forget.

In addition to the committee, I would like to thank Craig Formby and Moise Goldstein, my advisors at Johns Hopkins University, for introducing me to the field of hearing, and for giving me support and guidance at the beginning of my graduate-school experience.

This dissertation would not have been possible without financial support from the National Institute on Deafness and Other Communication Disorders in the form of a training grant to the Harvard-MIT Speech and Hearing Sciences Program. Additional support was provided by the Office of Naval Research and the National Science Foundation.

The Speech and Hearing Sciences (SHS) Ph.D. Program has provided the ideal multidisciplinary environment for me as a graduate student. I thank Susan Voss, a friend since our days in Brown Engineering, for introducing me to the SHS program. I am grateful to Nelson Kiang for his role in founding this program, and for our thought-provoking conversations. Bill Peake provided early encouragement and advice for my interest in carrying out my research with Laurel Carney at Boston University. He also maintained an interest in my work, which included his willingness to read a lengthy manuscript. John Rosowski, who also provided early encouragement for my working with Laurel, was a great role model during my time as a teaching assistant, continually demonstrating the importance of focusing on the students' perspective. Larry Frishkopf provided much valuable advice on academic and career issues in his role as my academic advisor at MIT. I thank Bill Siebert for his willingness to talk with me about my work, which would clearly not have been possible without his pioneering work in this area. Jason Smith and Geoff Meltzner were instrumental in making the first two years of classes in the SHS program a fun, productive, and memorable experience. In addition, A.J. Aranyosi, Harold Cheyne, Brad Cranston, Greg Huang, John Iversen, Sridhar Kalluri, Courtney Lane, Chris Long, Martin McKinney, Janet Slifka, and Jason Sroka have helped to make graduate school an enjoyable experience beyond all the hard work.

I have truly enjoyed working in the Boston University Hearing Research Center (HRC), which is where the day-to-day research for this dissertation was conducted. The breadth of research areas and the pleasant atmosphere provided an excellent location to work on hearing research. I have benefitted greatly from conversations with Domenica Karavitaki, a fellow SHS student in the HRC, about research and logistical issues related to working in the HRC as an MIT student. I have had many valuable discussions with Chris Mason, and her help with reviewing manuscripts and presentations of the material from this dissertation has been extremely valuable. Susan Moscynski has also provided much help with reading drafts of papers, preparing posters, and listening to many long practice talks. In addition, Mary Evilsizer, Qing Tan, Xuedong (Frank) Zhang, and Ling Zheng have made working in the lab an enjoyable and productive experience.

I have been extremely fortunate to have had the opportunity to spend the last year getting to know Torsten Dau, even sharing an office with him for part of the year. I have enjoyed our many stimulating conversations about all aspects of the auditory field, his never-ending curiosity and enthusiasm, and especially his friendship.

Finally, I would like to give special thanks to Karen Hull for her never-ending support, patience, love, friendship, and inspiration. She has put up with me through many stressful times over the past five years, and has given me a balance in my life, without which completion of this dissertation would not have been possible. My mother, Jane Arnold Heinz, and my father, John Heinz, have supported me from the beginning in many ways, giving me advice all along the way, but without ever an ounce of pressure. Only through the opportunities that they have provided for me have I been able to arrive where I am today. Jan Turner has also given me great support throughout my graduate education, and I appreciate her kindness and sincerity in everything she does.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The excellent sensitivity and exquisite frequency resolution of the normal auditory system are thought to result from an active mechanism within the inner ear that is often referred to as the cochlear amplifier (Yates, 1995; Moore, 1995b). Many nonlinear response properties observed in the auditory periphery have been shown to be related to a single physiologically vulnerable mechanism, and to correspond to a reduction in basilar-membrane (BM) gain as stimulus level is increased. The observation that these nonlinear response properties are reduced or absent in physiologically compromised cochleae suggests that damage or loss of the underlying active mechanism is related to sensorineural hearing impairment (Patuzzi *et al.*, 1989; Ruggero and Rich, 1991; Ruggero, 1992; Ruggero *et al.*, 1997). These findings raise the question of whether the cochlear amplifier benefits normal hearing listeners, especially in complex listening environments, such as speech in the presence of background sounds, for which hearing-impaired listeners have difficulty even with hearing aids (Moore, 1995b; Peters *et al.*, 1998; Moore *et al.*, 1999a). In addition to evaluating the potential benefit of the cochlear active process for complex tasks, it is also important to address why it is often difficult to measure deficits in psychophysical performance of impaired listeners in simple tasks. A sufficient understanding now exists of how many of the basic nonlinear properties of the auditory periphery relate to one another in terms of a single physiologically vulnerable mechanism to justify a quantitative evaluation of the psychophysical significance of the cochlear amplifier.

The modeling study described here seeks to quantitatively relate many of the nonlinear

physiological response properties associated with the cochlear amplifier to psychophysical performance in several tasks based on the information in auditory-nerve (AN) discharge patterns. The five chapters that follow this introduction are written in the form of journal papers, with individual introduction and discussion sections. This introduction provides a brief summary of previous physiological and psychophysical work that has characterized cochlear nonlinearity, as well as previous studies that have quantitatively related physiological response properties to psychophysical performance. Each of these topics is also discussed in several excellent review chapters; however, because the current work relies heavily on this previous work, a brief overview of these areas is warranted. This introduction concludes with a brief overview of the general approach taken in the present work, and a description of how each chapter relates to the overall goals of this dissertation.

1.1 Background

1.1.1 Physiological measures of cochlear nonlinearity

Gold (1948) proposed an active mechanism based on local feedback within the inner ear that supplied additional energy to BM vibration. However, it was not until 1971 that BM motion was observed to be nonlinear (Rhode, 1971; see Ruggero, 1992, and Yates, 1995, for reviews). The gain of the BM response to tones at the characteristic frequency (CF, the frequency of maximal low-level response) of basal (high-CF) locations on the BM is reduced by 60-70 dB as tone level is increased from low to high levels (guinea pig: Nuttall and Dolan, 1996; chinchilla: Ruggero *et al.*, 1997). A similar reduction in gain is observed in the same animals as a result of death. The strength of cochlear nonlinearity has been reported to be significantly less in the apex of the cochlea (i.e., low CF's), e.g., Cooper and Rhode (1997) observed less than 20 dB of gain for a 500-Hz CF in the chinchilla. The BM response to near-CF tones is linear at low levels ($\leq\simeq$ 20 dB SPL) and is compressive at higher stimulus levels (Nuttall and Dolan, 1996; Ruggero *et al.*, 1997). Many reports have suggested that near-CF responses become linear at levels above 80-90 dB SPL (e.g., Patuzzi *et al.*, 1989; Ruggero and Rich, 1991; Ruggero *et al.*, 1992; Goldstein, 1995), but Ruggero *et al.* (1997)

have recently demonstrated that the BM response is still compressive at levels above 100 dB SPL in the most sensitive cochleae they studied. Normalized BM magnitude responses at different levels demonstrate that while the response to tones near CF is highly nonlinear, responses to tones at frequencies more than roughly half an octave away from CF are linear at all levels (e.g., Ruggero *et al.*, 1997). Thus, tuning on the BM broadens as stimulus level is increased in the normal auditory system. In addition to compressive magnitude responses, level-dependent phase shifts that are consistent with broadened tuning have been observed over a wide dynamic range in BM (Ruggero *et al.*, 1997), inner-hair-cell (IHC, Cheatham and Dallos, 1998), and AN responses (Anderson *et al.*, 1971). Several forms of cochlear nonlinearity have been suggested to provide a benefit to the normal auditory system by extending the dynamic range. Compressive BM magnitude responses allow a wide input dynamic range to be encoded over a much smaller output dynamic range on the BM and in the IHC (e.g., Yates, 1995). In addition, level-dependent phase shifts result in spatio-temporal discharge patterns across the AN population that vary systematically with level over a wide dynamic range (Carney, 1994). The relative contributions of nonlinear gain and level-dependent phase cues to extending the psychophysical dynamic range of the auditory system are compared quantitatively in this dissertation.

The nonlinear BM responses to tones described above are consistent with the idea that BM tuning varies with stimulus level, i.e., as stimulus level increases, broadened tuning results in a shallower phase-versus-frequency response and a decrease in gain near CF. Studies using complex stimuli, such as multiple tones or bandlimited noise, suggest that BM tuning is controlled by stimulus energy in a broader frequency region than the excitatory tuning curve. Studies of two-tone suppression in the AN indicate that the response to a tone at CF can be reduced by the presence of an off-CF tone (Sachs and Kiang, 1968; Arthur *et al.*, 1971; Delgutte, 1990b), with suppressive regions flanking and lying below the excitatory tuning curve. Similar findings have been observed in BM responses, and indicate that two-tone suppression results from the same mechanism at the level of the BM that is responsible for the nonlinear gain and phase responses to CF tones (Ruggero *et al.*, 1992). The ability of off-CF stimulus energy to influence responses to near-CF energy is not limited to tonal stimuli. Auditory-nerve discharge rate in response to narrowband noises that are centered at CF

increases as the noise bandwidth is increased initially, but decreases as the noise bandwidth is increased further such that part of the noise energy falls outside the excitatory tuning curve and within the suppressive regions (Ruggero, 1973; Gilbert and Pickles, 1980; Schalk and Sachs, 1980). Basilar-membrane responses to transient stimuli indicate that changes in BM tuning are extremely fast (a time constant on the order of 200 μs, Recio *et al.*, 1998), consistent with the fast onset of suppression (Arthur *et al.*, 1971). Thus, peripheral tuning in the normal auditory system varies based on the spectral and temporal configuration of the stimulus. Several psychophysical consequences of stimulus-dependent tuning in the normal auditory system are evaluated and discussed in this dissertation.

Further evidence that nonlinear gain, level-dependent phase, and suppression response properties originate from the same physiologically vulnerable mechanism comes from observations that many types of cochlear damage (e.g., due to acoustic trauma, ototoxic drugs, death) result in simultaneous loss of these nonlinear properties (see Ruggero, 1992 for review). The most prevalent view in the literature is that these nonlinear response properties originate from a "cochlear-amplifier" mechanism that provides additional energy to BM vibration for low-level stimuli (e.g., Yates, 1995; Ruggero *et al.*, 1997; Recio *et al.*, 1998). In this view, the gain of the cochlear amplifier is reduced as stimulus level is increased, resulting in compressive magnitude responses near CF, broadened tuning, level-dependent phase shifts, and suppressive effects for complex stimuli. The term *cochlear amplifier* is used in this dissertation to represent the underlying physiologically vulnerable mechanism that is responsible for these nonlinear phenomena, and is not intended to depend on the biophysical basis for this mechanism. There is ongoing debate concerning the biophysical basis for the underlying mechanism (see Yates, 1995, and Patuzzi, 1996, for reviews), including whether an amplification mechanism is required to account for these nonlinear properties (e.g., Allen and Fahey, 1992; Allen, 1997).

The role of the outer hair cells (OHC's) in this physiologically vulnerable mechanism is well established. Ruggero and Rich (1991) demonstrated a strong correlation between the normal function of the OHC's and nonlinear BM responses. Further support for the role of OHC's comes from the reduction in BM response that results from stimulation of the medial-olivary-complex efferent fibers (Murugasu and Russell, 1996), which are known to innervate

OHC's (see Wiederhold, 1986, Ryugo, 1992, and Guinan, 1996, for reviews). Physiological studies of damaged cochleae have demonstrated that different functional consequences occur based on the proportion of IHC and OHC damage, and the effects are consistent with the proposed role of OHC's in modifying BM mechanics (Kiang *et al.*, 1970; Liberman and Dodds, 1984; Liberman and Kiang, 1984; Liberman *et al.*, 1986). Many different types of cochlear damage result in a reduction in OHC function, suggesting that impairment or absence of the cochlear amplifier may occur in many common types of sensorineural hearing loss in humans (Pickles, 1988; Patuzzi *et al.*, 1989; Ruggero and Rich, 1991; van Tasell, 1993; Moore, 1995b). This idea is supported by the observation that otoacoustic emissions (OAE's), which depend on the normal function of OHC's in a variety of ways (Shera and Guinan, 1999a), are present in most normal-hearing human ears and are absent in most hearing-impaired ears (Probst *et al.*, 1991; Allen and Lonsbury-Martin, 1993). Despite the clear relation of the cochlear amplifier to normal hearing, it has been difficult to determine how loss of the cochlear amplifier relates to the near-normal performance of hearing-impaired listeners in simple listening tasks and to the large deficits in hearing-impaired performance in complex listening environments even with hearing aids. A quantitative evaluation of the psychophysical significance of the physiological response properties associated with the cochlear amplifier may lead to a better understanding of this important problem.

1.1.2 Psychophysical measures of cochlear nonlinearity

Many psychophysical studies have sought ways to measure the effects of cochlear nonlinearity in humans, with the goal of finding simple performance measures that may be correlated with the difficulties experienced by hearing-impaired listeners in complex tasks. The most fundamental physiological nonlinear property that has been evaluated psychophysically is broadened tuning with increases in stimulus level. Many psychophysical masking methods that are used to estimate auditory tuning demonstrate reduced frequency selectivity as stimulus level is increased (e.g., Moore and Glasberg, 1987; Glasberg and Moore, 1990; Rosen and Baker, 1994; Rosen *et al.*, 1998; see Moore, 1995a for a review); however, there is much debate over the appropriate method to evaluate the dependence of filter shape on stimulus level, given the underlying physiology (summarized by Rosen *et al.*, 1998). Many recent psy-

chophysical studies have suggested that compression and suppression have significant effects on masking in normal-hearing listeners (Oxenham and Moore, 1994; Oxenham *et al.*, 1997; Moore and Vickers, 1997; Plack and Oxenham, 1998; Oxenham and Plack, 1998; Bacon *et al.*, 1999; Gifford and Bacon, 2000), thereby raising questions about the interpretation of psychophysical methods for estimating frequency selectivity that do not consider these effects. Several psychophysical methods have been developed to estimate BM compression based on assumptions of compressive responses to tones at CF and linear responses to tones well away from CF (Oxenham and Plack, 1997; Plack and Oxenham, 2000). These methods often measure compression that is comparable to that observed in BM responses. Psychophysical methods that are used to evaluate various aspects of cochlear nonlinearity often give results that are generally consistent with physiological response properties measured in experimental animals. Many of these psychophysical measures have been shown to be related to one another (Hicks and Bacon, 1999a,b; Moore *et al.*, 1999b), consistent with the idea that the underlying physiological response properties result from the same mechanism. However, a thorough understanding of how the set of physiological response properties associated with the cochlear amplifier influence psychophysical measures is lacking.

The interpretation of the psychophysical consequences of cochlear damage has often been based on isolated response properties of the BM that are associated with the cochlear amplifier, such as increased thresholds, broadened tuning, or lack of compression (see Moore, 1995b, and Moore and Oxenham, 1998, for reviews). However, damage to the cochlear amplifier affects these nonlinear response properties simultaneously, in addition to the effects of suppression and level-dependent phase shifts. Many of these BM response properties are significantly transformed in the AN (see Patuzzi and Robertson, 1988 for a review) due to the limited dynamic range of AN fibers (Sachs and Abbas, 1974), the reduced ability of AN fibers to phase-lock to the fine-time structure of high-frequency stimuli (Johnson, 1980), and the random nature of AN discharges (Kiang *et al.*, 1965). Thus, it is desirable to evaluate the consequences of cochlear damage based on the effects of the cochlear-amplifier mechanism as a whole on both temporal and average-rate information in the AN, rather than based on individual BM response properties. The phenomenological AN models used in the present study allow each of the nonlinear response properties associated with the cochlear amplifier

to be studied together or individually. These physiological response properties can then be quantitatively related to psychophysical performance.

1.1.3 Quantitatively relating physiology to psychophysical performance

Siebert (1965, 1968, 1970) was the first to evaluate the limits on psychophysical performance imposed by the random nature of AN discharges by combining models of peripheral auditory signal processing with statistical decision theory (SDT). This approach allows the psychophysical significance of physiological response properties to be evaluated quantitatively (see Delgutte, 1996 for a review). The fundamental requirement of this approach is a statistical description of the variation in the random nature of AN responses as a function of stimulus parameters. This approach is used in the present dissertation to evaluate the effects of the cochlear amplifier on both temporal and average-rate information in the AN.

Many studies have used this approach to evaluate the ability of the population of AN discharge counts (i.e., rate-place models) to account for human performance (e.g., Siebert, 1965, 1968; Delgutte, 1987; Viemeister, 1988a,b; Winslow and Sachs, 1988; Winter and Palmer, 1991). The number of discharges observed in a small number of AN fibers is often able to account for human performance levels; however, the trends in predicted performance as a function of stimulus parameters often do not match the trends in human performance. Siebert (1970) and Colburn (1969, 1973, 1981) extended the analytical AN models and analysis to include a description of the ability of AN fibers to phase-lock to tonal stimuli. Siebert (1970) evaluated the ability of temporal and rate-place information to account for human frequency discrimination performance. Similarly, Colburn (1981) extended Siebert's (1968) rate-place analysis of monaural level discrimination to include information from phase-locking cues. Colburn (1969, 1973, 1977a,b) also evaluated a variety of binaural discrimination and detection tasks based on the average-rate and temporal information available in the AN. The amount of temporal information in AN responses is often found to be much greater than rate-place information. Predicted performance based on temporal information is often orders of magnitude better than human performance; however, trends

in predicted performance based on temporal information often match trends in human performance as a function of stimulus parameters (see Chapters 2 and 3). Several studies have demonstrated restricted (sub-optimal) temporal processors that are better able to account for human performance than optimal use of AN discharge times (Colburn, 1969, 1973, 1977a,b; Goldstein and Srulovicz, 1977).

1.2 General approach used in this dissertation

The work presented in this dissertation is a direct extension of the approach taken by Siebert and Colburn to evaluate the effects of physiological response properties on temporal and average-rate information in the AN. This approach has been extended in several ways in order to make the approach more general, and therefore better able to address the overall goals of this dissertation. Several new methods have been developed using statistical decision theory that allow the application of this approach to more complex psychophysical tasks. Another significant extension from Siebert and Colburn's work is the inclusion of more realistic descriptions of auditory peripheral signal processing. The most significant nonlinear response properties associated with the cochlear amplifier have been included in phenomenological analytical and computational AN models. Comparisons of performance predictions based on various nonlinear and linear versions of the AN models allow for the evaluation of the relative significance of individual response properties of the cochlear amplifier that may not be separable in psychophysical experiments. Many of these model versions are unrealistic models of auditory processing (e.g., nonlinear gain with linear phase), but are useful in gaining insight into the contributions of specific response properties to psychophysical performance.

Performance predictions are made throughout this dissertation based on two models that differ in the type of AN information that is assumed to be used by an optimal processor. The rate-place model assumes that the processor uses the distribution of discharge counts across the AN population to make decisions in a psychophysical task, i.e., no temporal information is used. The all-information model assumes that the set of discharge times from each AN fiber are processed optimally to make a decision, i.e., both temporal and average-rate information

is used. Comparisons of optimal performance based on the all-information and rate-place models to human data from the literature are made in terms of both trends and absolute performance levels. These comparisons provide an indication of the significance of different types of information for explaining human performance. The methods developed in this dissertation were used to make predictions for several psychophysical tasks for which the cochlear amplifier has been suggested to be significant.

1.3 Overview of the dissertation document

Chapter 2 (Heinz *et al.*, 2000a) describes a method that generalizes Siebert and Colburn's approach to allow any AN model to be used, either analytical or computational, for any one-parameter discrimination experiment. Much of the previous work in this area has been tied to specific analytical AN models, and therefore has been limited in the ability to evaluate new physiological response properties or stimuli other than pure tones. This general extension is particularly important for evaluating hearing impairment because tasks with complex stimuli are typically required in order to demonstrate a significant impairment. The method developed in Chapter 2 was verified by comparing frequency- and level-discrimination performance predictions based on a simple computational AN model to previous predictions from an analytical AN model (Siebert, 1968, 1970).

Chapter 3 (Heinz *et al.*, 2000b) presents an extension of the statistical decision theory analysis to allow a more general set of psychophysical tasks to be evaluated, specifically discrimination experiments in which a secondary parameter is randomized in order to limit the cues available to a listener. Many psychophysical experiments have eliminated level cues by randomizing the overall stimulus level in each observation interval. For example, the class of profile-analysis experiments developed by Green (1988) and colleagues have provided new constraints on how to interpret the ability of listeners to discern changes in the spectral shape of complex stimuli in terms of basic auditory level-discrimination performance. The method developed in Chapter 3 was used to evaluate the effects of random level variation on frequency-discrimination performance predicted based on the rate-place and all-information models.

Chapter 4 describes an evaluation of the benefit of the cochlear amplifier for extending the dynamic range of the auditory system. An analytical AN model was developed to include nonlinear gain, level-dependent phase, and high-, medium-, and low-spontaneous-rate AN-fibers. The relative contributions of average-rate and nonlinear-phase information were evaluated in terms of the ability of AN fibers with a narrow range of CF's to encode changes in level. Both nonlinear-phase and average-rate information are shown to be required to encode changes in level over the entire dynamic range of hearing based on a narrow range of CF's. A realistic processor for decoding the level-dependent phase and nonlinear-gain cues provided by the cochlear amplifier is monaural coincidence detection across adjacent AN fibers (Carney, 1994). This processor is attractive because several neural response types in the antero-ventral cochlear nucleus (the first stage of processing in the auditory pathway after the AN) have been shown to have response properties consistent with monaural coincidence detection (Carney, 1990; Joris *et al.*, 1994). Predictions based on a population of coincidence counters are shown to be consistent with both the absolute performance levels and performance trends in human data across the entire dynamic range of hearing for both low and high frequencies. A comparison of human level-discrimination performance to predicted performance based on the entire AN population suggests that the cochlear amplifier has only a small effect on overall performance in quiet; however, several subtle but significant aspects of the human data [e.g., the "near-miss" to Weber's Law at low frequencies and the "mid-level bump" at high frequencies (Florentine *et al.*, 1987)] appear to be consistent with nonlinear response properties associated with the cochlear amplifier.

The level-discrimination predictions in Chapter 4 suggest that the cochlear amplifier is beneficial for encoding changes in sound level within AN fibers with a narrow range of CF's; however, the cochlear amplifier has only a small influence on level-discrimination performance in quiet, i.e., when all CF's in the AN population are considered. Chapters 5 and 6 address this issue by first developing the theoretical analyses required to study noise-masking experiments, and then applying this approach to test a common psychophysical measure of auditory frequency selectivity.

Chapter 5 presents a new theoretical approach that allows the significance of physiological response properties to be evaluated in psychophysical masking experiments that use random-

noise stimuli. Noise stimuli have been used extensively in psychophysical experiments as maskers and signals, and many theories of auditory processing are based on the results of these experiments. However, many of these theories are based on the assumption that a noise masker simply eliminates the information in CF's that are within the noise spectrum. The complex effects of the cochlear amplifier on auditory tuning, e.g., compression and suppression, raise questions about the validity of many of these assumptions. In addition, random-noise stimuli are variable, and thus stimulus fluctuations from noise waveform to noise waveform can limit psychophysical performance in addition to the random nature of AN discharges. The general approach developed in Chapter 5 extends the analyses in Chapters 2 and 3 to quantify the relative contributions of stochastic AN responses and random stimulus fluctuations for limiting performance based on temporal and average-rate information in the AN.

Chapter 6 describes an evaluation of the significance of the cochlear amplifier for a commonly used psychophysical method for estimating auditory frequency selectivity. The theoretical approach developed in Chapter 5 was used to quantify AN information in discharge counts and in discharge times for the detection of a tone in notched-noise maskers. Different versions of a computational AN model (Appendix A) were used to evaluate the effects of compression and suppression on estimates of auditory filters using the notched-noise method (Patterson, 1976; Glasberg and Moore, 1990). Both compression and suppression are shown to be significant for psychophysical methods of estimating auditory frequency selectivity, based on the dependence of the cochlear amplifier on the spectral and temporal configuration of the stimulus. Implications of these results for the interpretation of psychophysical methods for estimating auditory filters are discussed.

The analysis in Chapter 6 represents one example of the applicability of the general approach developed in this dissertation to complex psychophysical experiments. In general, this approach can be used to evaluate many types of complex masking phenomena that are important to theories of normal and impaired auditory processing. The last chapter summarizes the significance of the work presented in this dissertation, provides some brief comments on potential implications of this work for evaluating hearing impairment, and suggests future work.

Chapter 2

Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory-nerve

2.1 Abstract

A new method for the calculation of psychophysical performance limits based on neural responses is introduced and compared to previous analytical methods for evaluating auditory discrimination of tone frequency and level. The new method uses optimal decision theory (ODT) and a computational model for a population of auditory-nerve (AN) fiber responses. The use of computational AN models allows predictions to be made over a wider parameter range and with more complete descriptions of AN responses than in analytical models. Performance based on AN discharge *times* (*all information*) is compared to performance based only on discharge *counts* (*rate-place*). After the new method is verified over the range of parameters for which previous analytical models are applicable, the parameter space can then be extended. For example, a computational model of AN activity that extends to high frequencies can be used to explore the idea that rate-place information is responsible for frequency encoding at high frequencies because the rolloff in AN phase locking above 2 kHz is thought to eliminate temporal information at high frequencies. Contrary to this belief, results of the new analysis show that rate-place predictions for frequency discrimination are inconsistent with human performance in the dependence on frequency for high frequencies, and that there is significant temporal information up to at least 10

[0]A similar version of this chapter has been submitted to *Neural Computation* (Heinz *et al.*, 2000a).

kHz. In contrast, all-information predictions match the trends in human performance across all frequencies, although optimal performance is much better than humans. The use of computational AN models in the present study provides new constraints on hypotheses of neural encoding of frequency in the auditory system; however, the present method is limited to simple tasks with deterministic stimuli. A companion paper describes an extension of this approach to more complex tasks that include random variation of one parameter, e.g., random level variation, which is often used in psychophysics to test neural encoding hypotheses.

2.2 Introduction

One of the challenges in understanding sensory perception is to test, quantitatively, hypotheses for the physiological basis of psychophysical performance. The random nature of neural responses (i.e., that two presentations of an identical stimulus produce different discharge patterns) imposes limitations on performance. The fundamental requirement for relating neural encoding to psychophysical performance is a statistical description of neural discharge patterns in response to the relevant stimuli. In general, if the statistics of the response change significantly as a stimulus parameter is varied, then accurate discrimination of that parameter is possible.

Attempts to relate neural responses to psychophysical performance often involve the question of whether a small set of individual neurons can statistically account for performance, but can also involve studies of encoding by an entire population of neurons (see review by Parker and Newsome, 1998). Implicit in any hypothesis focused on a small set of neurons is the assumption that the brain ignores statistically significant information in other neurons within the population. To avoid this assumption, the total amount of information in a neural population can be quantified. The amount of total information, and the trends in information as stimulus parameters are varied, can be used to test hypotheses of neural encoding and to suggest neural information-processing mechanisms. These studies, which require modeling approaches due to the inability to record from all neurons within a population, have typically been limited to simple psychophysical tasks with simple stimuli that can be studied with analytical models. The present paper describes a new method that allows computational models for neural responses to be used in studies that relate physiological

properties to psychophysical performance. A companion paper describes a further generalization of this approach to include psychophysical tasks that randomly vary one stimulus parameter in order to limit the cues available to a subject (Chapter 3).

The auditory system provides a useful example of the benefits that can result from quantifying the total information in a neural population. The auditory nerve (AN) is an obligatory pathway between the cochlea (inner ear) and the central nervous system (CNS), and thus contains all of the information about an auditory stimulus that can be processed by a listener. Siebert (1965, 1968, 1970) was the first to evaluate limits on psychophysical performance in auditory discrimination tasks by combining methods from optimal decision theory (ODT) with models of peripheral auditory signal processing [1] [see Delgutte (1996) for a review of Siebert's approach and many subsequent studies]. The questions addressed by the ODT approach were fundamental, such as whether the physiological noise inherent in peripheral auditory processing can account for human performance, and whether the central nervous system (CNS) makes use of temporal information in AN discharge patterns to encode frequency. In addressing these questions, simple analytical AN models were used that included only those response properties considered to be of primary importance: bandpass frequency tuning, saturating rate-level curves, phase-locking to tonal stimuli, a population of 30,000 AN fibers with logarithmic characteristic frequency (CF) spacing, and nonstationary Poisson statistics of AN discharges (Kiang *et al.*, 1965). These studies were limited to simple psychophysical tasks, such as pure-tone level and frequency discrimination, because of their relatively simple interpretations and the limits of the analytical AN models. Valuable insight into the encoding of frequency and level in the peripheral auditory system was garnered from these previous studies; however, fundamental questions are still debated, such as the frequency ranges over which average-rate and temporal information are used to encode frequency, and how a system in which the majority of AN fibers have a small dynamic range can encode sound level over a broad dynamic range.

The physiological mechanism of frequency encoding has been a disputed issue since (at least) the 1940s, when spectral (place) cues (von Helmholtz, 1863) were challenged by tem-

[1] Fitzhugh (1958) used this approach to investigate visual psychophysical performance based on the random nature of optic-nerve responses.

poral cues (Schouten, 1940; Wever, 1949). The task of pure-tone frequency discrimination has been widely used to test rate-place and temporal hypotheses for frequency encoding by measuring the just-noticeable-difference (JND) in frequency between two tones [see Moore (1989) and Houtsma (1995) for reviews of human performance, also see Fig. 2-4]. Rate-place schemes are based on the frequency selectivity (or tuning) of AN fibers (Kiang *et al.*, 1965; Patuzzi and Robertson, 1988). Two tones of differing frequency excite different sets of AN fibers, and thus the listener can discriminate two tones that differ in frequency by detecting differences in the sets of AN fibers that are activated. Temporal schemes are based on the phase-locked responses of AN fibers to tones (Kiang *et al.*, 1965; Johnson, 1980; Joris *et al.*, 1994).

It is generally believed that rate-place information must be used at high frequencies (at least above 4-5 kHz, e.g., Wever, 1949; Moore, 1973, 1989; Dye and Hafter, 1980; Wakefield and Nelson, 1985; Javel and Mott, 1988; Pickles, 1988; Moore and Glasberg, 1989), because of the rolloff in phase-locking above 2-3 kHz [Johnson, 1980; Joris *et al.*, 1994; or see Fig. 2-1(c)]. For low frequencies, there is continued debate as to whether rate-place or temporal information is used to encode frequency (e.g., Pickles, 1988; Javel and Mott, 1988; Moore, 1989; Cariani and Delgutte, 1996a,b; Kaernbach and Demany, 1998). The assumption that there is no temporal information above 4-5 kHz has been used to interpret numerous psychophysical experiments and to develop theories for the encoding of sound (e.g., Moore, 1973; Viemeister, 1983). However, no study has accurately quantified the total amount of temporal information at high frequencies, nor compared the trends in rate-place and temporal information versus frequency at high frequencies using the same AN model. Siebert (1970) used an analytical AN model and ODT analysis to evaluate the ability of temporal and rate-place information to account for human frequency discrimination. His model included a nonstationary Poisson process with a time-varying discharge rate that described the AN fibers' phase-locking to low-frequency tonal stimuli; however, the model did not include rolloff in phase-locking above 2-3 kHz, and thus Siebert's predictions were limited to low frequencies. Siebert found that temporal information significantly outperformed rate-place information at low frequencies, but that optimal performance based on rate-place is more similar to levels of human performance. Goldstein and Srulovicz (1977) found that when a

simple description of the rolloff in phase-locking at high frequencies was included in Siebert's (1970) AN model, predicted performance based on temporal information (in terms of $\Delta f/f$) became worse as frequency increased above 1 kHz, similar to human performance (Moore, 1973). However, the description of the lowpass rolloff in phase-locking used by Goldstein and Srulovicz (1977) had too low a cutoff frequency (600 Hz rather than 2500 Hz), and too shallow a slope (40 dB/dec rather than 100 dB/dec) when compared to more recent data from cat [see Fig. 2-1(c) and Johnson, 1980; Weiss and Rose, 1988]. In addition, Goldstein and Srulovicz (1977) only quantified performance based on a small set of AN fibers with CF equal to the tone frequency, and thus it is not possible from their predictions to know whether there is significant temporal information at high frequencies.

Another significant result from Siebert (1970) was that optimal performance based on all information improved much too quickly as a function of duration compared to human performance, at least for long durations. Siebert concluded that human listeners are severely limited in their ability to use the temporal information in AN discharges, and that there is sufficient rate-place information to account for human performance if this information were used efficiently. Goldstein and Srulovicz (1977) demonstrated that a restricted temporal scheme based on only the intervals between adjacent discharges (first-order intervals) corrected the duration dependence at long durations, and concluded that temporal schemes should not be ruled out. However, both of these analytical models are based on steady-state responses, and therefore are limited to long-duration stimuli due to the absence of onset and offset responses as well as neural adaptation.[2] The trends in human frequency discrimination performance versus duration have been shown to differ at long and short duration (Moore, 1973; Freyman and Nelson, 1986).

The encoding of level in the auditory system is an interesting problem given the discrepancy between the wide dynamic range of human hearing [at least 120 dB (Viemeister and Bacon, 1988)] and the limited dynamic range of the majority of AN fibers [less than 30 dB (May and Sachs, 1992)] (see Viemeister, 1988a,b for reviews of the "dynamic-range

[2]An analytical model that included transient responses was used by Duifhuis (1973) to evaluate the consequences of peripheral filtering on nonsimultaneous masking; however, his model did not include neural adaptation.

problem"). For many types of stimuli, the JND in level is essentially constant across a wide range of levels. The observation of constant performance across level (i.e., that the smallest detectable change in intensity is proportional to intensity) is referred to as Weber's Law. Weber's Law is observed for broadband noise (Miller, 1947) and for narrow-band signals in the presence of notched noise (Viemeister, 1974, 1983; Carlyon and Moore, 1984). However, for narrow-band signals in quiet where spread of excitation across the population of AN fibers is possible, performance improves slightly as a function of level, an effect referred to as the "near-miss" to Weber's Law (e.g., McGill and Goldberg, 1968; Rabinowitz *et al.*, 1976; Jesteadt *et al.*, 1977; Florentine *et al.*, 1987). Siebert (1965, 1968) demonstrated that a rate-place encoding scheme based on a population of AN fibers with limited dynamic range can produce Weber's Law for pure-tone level discrimination through spread of excitation. However, Siebert's model did not predict the near-miss to Weber's law for pure tones, and would not be expected to predict Weber's law for broadband noise or for narrowband stimuli in notched noise. Other studies have implicated the role of low-spontaneous rate, high-threshold AN fibers in producing Weber's law for conditions in which spread of excitation is assumed to be restricted (Colburn, 1981; Delgutte, 1987; Viemeister, 1988a,b; Winslow and Sachs, 1988; Winter and Palmer, 1991). However, when models have been used to quantify the total information available in a restricted CF-region with physiologically realistic distributions of the spontaneous-rate groups, performance degrades as level increases above 40 dB SPL (Colburn, 1981; Delgutte, 1987), inconsistent with Weber's law and with trends in human performance. In addition, no model has accurately quantified the effect of noise maskers on the spread of rate or temporal information across the population of AN fibers; instead simple assumptions for the influence of the noise maskers have been used.

There are now computational AN models (e.g., Payton, 1988; Carney, 1993; Robert and Eriksson, 1999) that provide more accurate physiological responses over a wider range of stimulus parameters than the analytical models described above. In addition, computational models can simulate AN responses to arbitrary stimuli, and thus can be used to study a wider range of psychophysical tasks, such as those involving noise maskers. The present paper focuses on extension of the ODT approach to allow the use of computational AN models to predict psychophysical performance limits for auditory discrimination. The

potential of *rate-place* and *all-information* (including rate-place and temporal information) encoding schemes to account for human performance is characterized for both frequency and level discrimination. The adjective *all-information* is used to emphasize the fact that this encoding scheme makes use of all the information in the distribution of average rates, in addition to the times of all discharges across the population of AN fibers. The contribution of temporal information in the responses can be inferred by comparing the predictions of the rate-place and all-information models. Note that the all-information model does not assume any specific forms of temporal processing, such as calculating synchrony coefficients or creating interval histograms. All-information predictions thus provide an absolute limit on achievable performance given the total information available in the AN.

It is critical, before proceeding to more complex computational AN models, to verify the computational method by comparing it to previous analytical model predictions (e.g., Siebert, 1965, 1968, 1970). Thus, optimal performance predictions were made for pure-tone frequency- and level-discrimination as a function of frequency, level, and duration based on a relatively simple computational AN model. The general agreement between computational and analytical predictions provides confidence that the computational method is valid. The present study quantifies that there is significant temporal information for frequency discrimination up to at least 10 kHz, and that all-information performance demonstrates the same trends versus frequency as human performance, whereas rate-place does not. These new predictions are important to theories of frequency encoding in the auditory system because they contradict the generally held belief that the rolloff in phase-locking eliminates all temporal information at high frequencies.

Previous studies have typically focused on either frequency or level discrimination, have often only examined a limited parameter range, and have used different AN models. By using *the same AN model* to make *both* rate-place and all-information predictions for both frequency and level discrimination versus frequency, level, and duration, the present study provides a unified description of previously suggested constraints on rate-place and temporal models. Thus, the present study demonstrates several benefits of using computational neural models to predict psychophysical performance limits. This approach is applicable to predicting performance limits for any one-parameter discrimination experiment based on any

neural population of any sensory system for which there are accurate models of the discharge statistics.

2.3 General Methods

2.3.1 Computational auditory-nerve model

The type of computational AN model for which the current study was designed can process an arbitrary stimulus and produce a time-varying discharge rate that can be used with a nonstationary point process (e.g., Poisson) to simulate AN discharge times. Examples of this type of model were described by Payton (1988) and Carney (1993). A simplified version of the Carney (1993) model was used in the present study to provide a better comparison to the AN model used by Siebert (1970), and to provide a simple comparison for future predictions with more complex AN models.[3]

Figure 2-1 shows pure-tone response properties of the computational AN model that are important for the present study. Table 2.1 provides details of the implementation of the model. A linear fourth-order gamma-tone filter bank was used to represent the frequency selectivity of AN fibers [see population response in Fig. 2-1(a)]. Model filter bandwidths were based on estimates of human bandwidths from the psychophysical notched-noise method for estimating auditory filter shapes [Glasberg and Moore (1990), see Table 2.1]. Psychophysically measured filters have been shown to match AN frequency tuning when both were measured in guinea pig (Evans *et al.*, 1992). The bandpass filter was followed by a memoryless, asymmetric, saturating nonlinearity (implemented as an arctan with a 3:1 asymmetry), which represents the mechano-electric transduction of the inner hair cell (IHC). The saturating nonlinearity contributes to the limited dynamic range of AN fibers [Fig. 2-1(b)]. [Note that the dynamic range is also affected by the IHC-AN synapse, see below and Patuzzi and Robertson (1988).] All AN model fibers had a rate threshold of roughly 0 dB SPL, a spontaneous rate of 50 spikes/s, and a maximum sustained rate of roughly 200 spikes/s. The model dynamic range for sustained rate was roughly 20-30 dB, while the dynamic range for onset

[3]Code for the AN model used in the present study is available at http://earlab.bu.edu/.

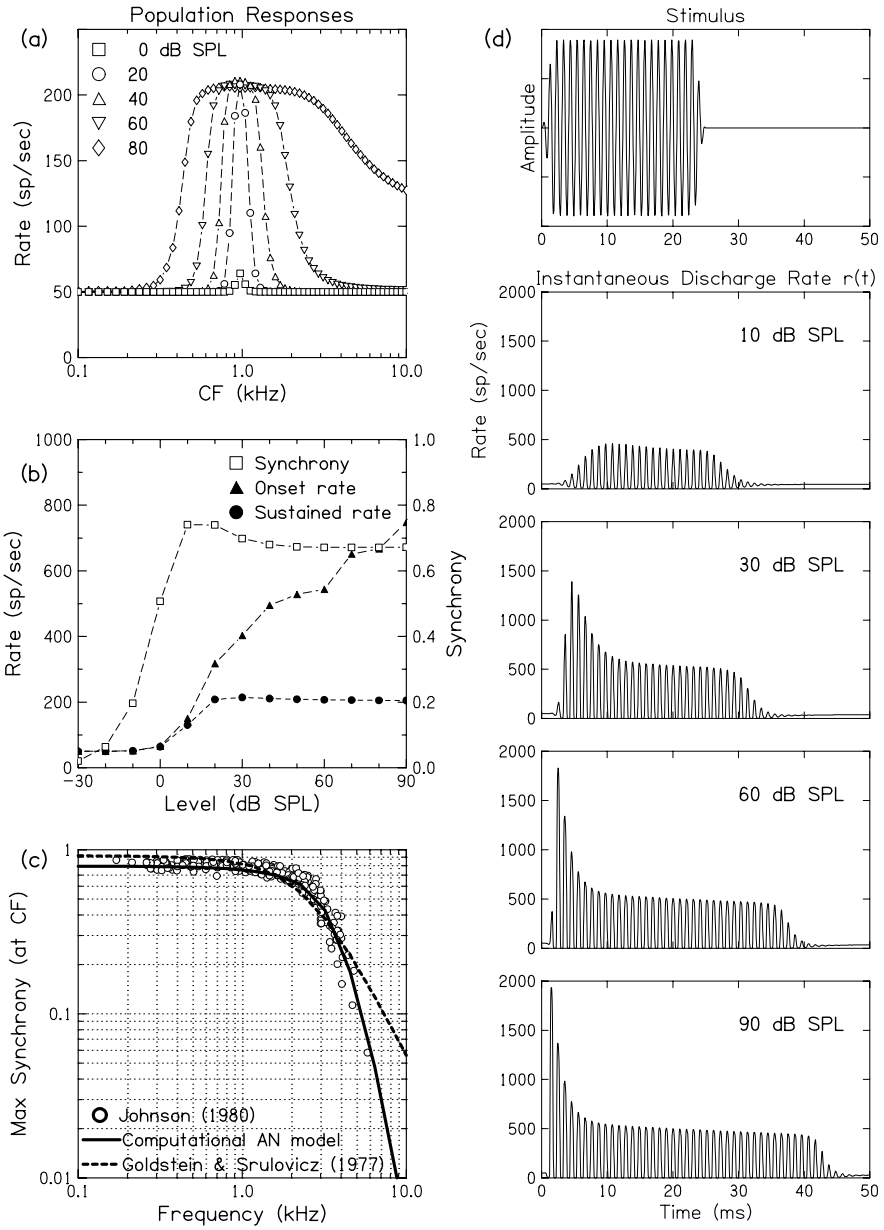

Figure 2-1: Basic pure-tone response properties of the computational AN model. (a) Sustained-rate population responses for the 60 AN-model CF's used in this study to pure tones at several levels. Tones were 970 Hz, 62 ms (10-ms rise/fall), with levels from 0 to 80 dB SPL. Sustained rate (calculated as the average rate over all full stimulus cycles within a temporal window from 10 to 52 ms) is plotted versus characteristic frequency (CF) of each model fiber. (b) Onset rate, sustained rate, and synchrony for a 970-Hz model fiber responding to a 970-Hz, 62-ms tone as a function of level. Onset rate was calculated as the maximum average rate over one stimulus cycle. The synchrony coefficient (units on right axis), represents the vector strength (Johnson, 1980) of the model response calculated over one cycle beginning at 40 ms. (c) Maximum synchrony coefficient over level for tones at CF as a function of frequency. Circles are data from cat (Johnson, 1980). Dashed line is from analytical model used by Goldstein and Srulovicz (1977). (d) Stimulus waveform and instantaneous discharge rates $r(t, f, L)$ for a $f = 970$-Hz fiber in response to 970-Hz, 25-ms (2-ms rise/fall) tones over a range of levels (L).

Table 2.1: Equations and parameters used to implement computational AN model.

Symbol	Description (units)	Equations/values
Human cochlear map[a]		
x	distance from apex (mm)	
$f(x)$	frequency corresponding to a position x (Hz)	$= 165.4(10^{0.06x} - 0.88)$
Gamma-tone filters[b]		
CF	characteristic frequency (kHz)	
ERB	equivalent rectangular bandwidth (Hz)	$= 24.7(4.37CF + 1)$
τ	Time constant of gamma-tone filter (s)	$= [2\pi(1.019)ERB]^{-1}$
$gtf[k]$	output of gamma-tone filter	
Inner-hair-cell		
$ihc[k]$	output of saturating nonlinearity	
	$ihc[k] = \{\arctan(K \cdot gtf[k] + \beta) - \arctan(\beta)\} / \{\pi/2 - \arctan(\beta)\}$	
K	controls sensitivity	1225
β	sets 3:1 asymmetric bias	-1
$ihc_L[k]$	lowpass-filtered inner-hair-cell output (see text)	
Neural adaptation model[c]		
T_s	sampling period (s) (see text)	
r_o	spontaneous discharge rate (spikes/s)	50
V_I	immediate "volume"	0.0005
V_L	local "volume"	0.005
P_G	global permeability ("volume"/s)	0.03
P_L	local permeability ("volume"/s)	0.06
PI_{rest}	resting immediate permeability ("volume"/s)	0.012
PI_{max}	maximum immediate permeability ("volume"/s)	0.6
C_G	global concentration ("spikes/volume")	
	$C_G = C_L[0](1 + P_L/P_G) - C_I[0]P_L/P_G =$	6666.67
$P_I[k]$	immediate permeability ("volume"/s)	
	$P_I[k] = 0.0173 \ln\{1 + \exp(34.657 \cdot ihc_L[k])\}$	
$C_I[k]$	immediate concentration ("spikes/volume")	
	$C_I[k+1] = C_I[k] + (T_s/V_I)\{-P_I[k]C_I[k] + P_L[k](C_L[k] - C_I[k])\}$	
	$C_I[0] = r_o/PI_{rest} =$	4166.67
$C_L[k]$	local concentration ("spikes/volume")	
	$C_L[k+1] = C_L[k] + (T_s/V_L)\{-P_L[k](C_L[k] - C_I[k]) + P_G(C_G - C_L[k])\}$	
	$C_L[0] = C_I[0](PI_{rest} + P_L)/P_L =$	5000.00
$r[k]$	instantaneous discharge rate (spikes/s)	$= P_I[k]C_I[k]$

[a] Greenwood (1990).

[b] see Patterson *et al.* (1987); Glasberg and Moore (1990); Carney (1993).

[c] see Westerman and Smith (1988); Carney (1993).

rate was much larger. The synchrony-level curves had a threshold roughly 20 dB below rate threshold, a maximum at a level that was just above rate threshold, and a slight decrease in synchrony as level was increased further [comparable to Johnson (1980) and Joris *et al.* (1994)].

An important property for the present study is the rolloff in phase-locking as frequency increases above 2-3 kHz [Fig. 2-1(c); Johnson, 1980; Joris *et al.*, 1994]. Weiss and Rose (1988) compared synchrony versus frequency in five species on a log-log scale and reported that the data from all species were well described by a lowpass filter with roughly 100 dB/decade rolloff (the only difference across species was the 3-dB cutoff frequency, e.g., $f_c = 2.5$ kHz for cat, and $f_c = 1.1$ kHz for guinea pig). To achieve the proper rolloff in synchrony (for all species) and cutoff frequency (for cat), seven first-order lowpass filters were used, each with a first-order cutoff frequency of 4800 Hz. The resulting filter had a 3-dB cutoff frequency near 2500 Hz and $\sim$ 100 dB/decade rolloff in the frequency range 4-6 kHz [Fig. 2-1(c)]. The model synchrony coefficients above 5 kHz are a simple extrapolation of the physiological data, consistent with the lowpass shape and slope across many species reported by Weiss and Rose (1988). For comparison, a typical description of synchrony rolloff used in analytical AN models is also shown (Goldstein and Srulovicz, 1977).

Neural adaptation was introduced through a simple three-stage diffusion model based on data from Westerman and Smith (1988) for the IHC-AN synapse. The continuous-time version of this adaptation model used by Carney (1993) was simplified by using fixed values for the immediate and local volumes and the local and global permeabilities [Lin and Goldstein (1995); and see Fig. 3 of Westerman and Smith (1988)]. The immediate permeability was a function of the input amplitude, where the relation was essentially linear above the resting permeability, and exponentially decayed to zero for negative inputs (Table 2.1). The model's time constants for rapid adaptation (1.3 ms at high levels) and short-term adaptation (63 ms at high levels) are consistent with those reported for AN fibers (e.g., Westerman and Smith, 1988). The output of the AN model [Fig. 2-1(d)] represents the instantaneous discharge rate $r(t, f, L)$ of an individual high-spontaneous-rate, low-threshold AN fiber.

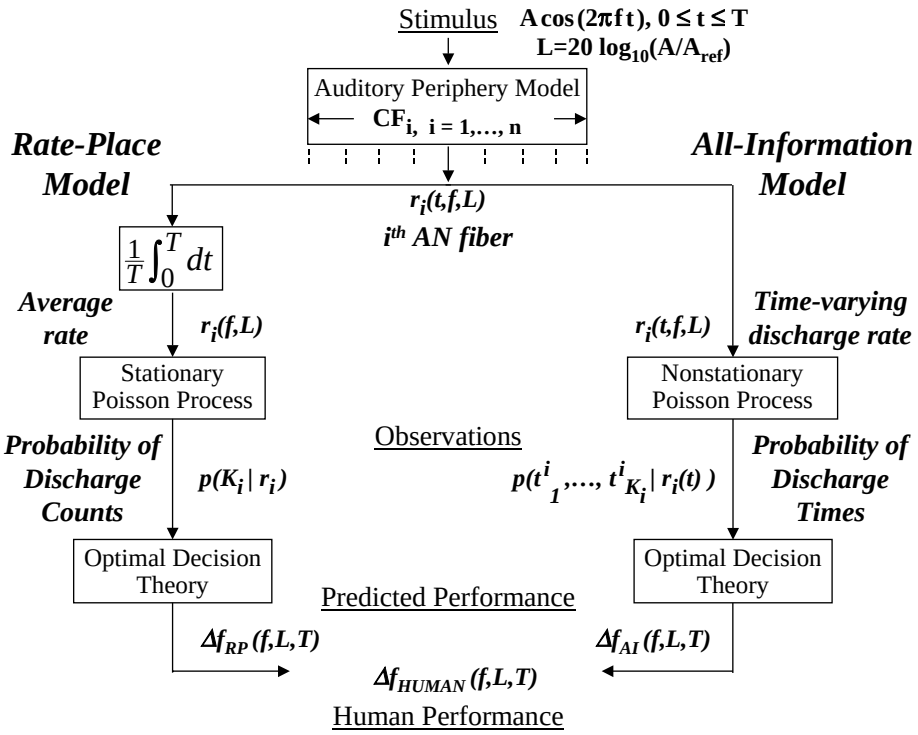

Figure 2-2: Overview of optimal decision theory (ODT) used with models of the auditory periphery to evaluate fundamental limits of performance on a psychophysical task (frequency discrimination shown).

2.3.2 Optimal decision theory

The application of ODT to evaluate auditory discrimination in the present study is an extension of Siebert's (1970) study of frequency discrimination (Fig. 2-2). A model of the auditory periphery describes the time-varying discharge rate $r_i(t, f, L)$ of the i^{th} AN fiber in response to a tonal stimulus of level L, frequency f, and duration T. A model population response is produced from a range of characteristic frequencies (CF's). A Poisson process provides a good description of the random nature of AN discharges [e.g., exponential interval distribution, and independent successive intervals (Colburn, 1969, 1973; Siebert, 1970)]. Two types of analyses are considered in the present study. *Rate-place* predictions are based on the assumption that the set of AN discharge counts over the duration of the stimulus, $\{K_i\}$, is the only information used by the listener. Thus, the rate-place analysis (left branch of Fig. 2-2), is based on a set of independent, stationary Poisson processes with rates equal to

the average discharge rates $r_i(f, L)$ of the AN model fibers (Siebert, 1968). In contrast, the *all-information* predictions (right branch of Fig. 2-2) are based on the assumption that the set of observations are the population of discharge times and counts $\{t_1^i, \cdots, t_{K_i}^i\}$ produced by a set of independent, nonstationary Poisson processes with time-varying rates $r_i(t, f, L)$ provided by the AN model (Siebert, 1970). Descriptions of the probability density of the observations, $p(t_1^i, \cdots, t_{K_i}^i | r_i)$, are used to evaluate the optimal performance achievable by any decision process (see Appendix in Section 2.10). Optimal performance based on rate-place, $\Delta f_{RP}(f, L, T)$, and all-information, $\Delta f_{AI}(f, L, T)$, is compared directly to human performance measured psychophysically, $\Delta f_{HUMAN}(f, L, T)$.

2.4 General Analytical Results: One-Parameter Discrimination

The results in this section are presented in terms of discrimination of a single stimulus parameter α, which is either frequency f or level L in the present study. A measure closely related to the sensitivity measure d' of psychophysics, $(\delta'_\alpha[CF])^2$, is used to represent the information from an individual AN fiber with characteristic frequency CF, where the *sensitivity per unit* α is represented by δ'_α (see Appendix in Section 2.10).[4] Assuming independent AN fibers (Johnson and Kiang, 1976), the total information from the population is the sum of the information from individual AN fibers. The just-noticeable difference in α, $\Delta\alpha_{JND}$, is inversely related to the amount of information available from the AN population, and can be calculated as

$$\begin{aligned} \Delta\alpha_{JND} &= \left\{ \sum_i \left(\delta'_\alpha \left[CF_i\right]\right)^2 \right\}^{-\frac{1}{2}} \\ &= \left\{ \sum_i \int_0^T \frac{1}{r_i(t,\alpha)} \left[\frac{\partial r_i(t,\alpha)}{\partial \alpha}\right]^2 dt \right\}^{-\frac{1}{2}}, \end{aligned} \tag{2.1}$$

[4]Durlach and Braida (1969) used the metric δ' to represent *sensitivity per bel* for intensity discrimination [i.e., $\delta' = d'(I_1, I_2)/\log_{10}(I_2/I_1)$] (see also Braida and Durlach, 1988). In the present study, δ'_α is used to represent *sensitivity per unit* α, where α can be level L in dB, or frequency f in Hz. Thus, the normalized sensitivity metric δ'_L differs from that used by Durlach and Braida (1969) only in the unit of level used (i.e., δ'_L represents *sensitivity per decibel*).

as described in the Appendix in Section 2.10. Equation 2.1 describes the optimal performance possible from all information available in the AN nerve in terms of the time-varying discharge rates $r_i(t, \alpha)$ of the population of AN fibers. Optimal rate-place performance is calculated by using the average discharge rate $r_i(\alpha)$ in Eq. 2.1, rather than $r_i(t, \alpha)$. Equation 2.1 is extremely general because it is applicable to any single-parameter discrimination experiment, and it can be used with any AN model (e.g., analytical or computational) that assumes Poisson, statistically independent fiber responses.

2.5 Computational Methods: Use of Auditory-Nerve Models

The time-varying discharge rates $r_i(t, \alpha)$ from the computational AN model can be substituted into Eq. 2.1 to evaluate optimal performance for various stimulus conditions. However, the partial derivative of the rate waveform with respect to α must be calculated as a function of time. The partial derivative was approximated from the responses of a computational AN model, e.g.,

$$\frac{\partial r_i(t, \alpha)}{\partial \alpha} \simeq \frac{r_i(t, \alpha + \Delta\alpha) - r_i(t, \alpha)}{\Delta\alpha}. \tag{2.2}$$

An approximation of the partial derivative with respect to frequency using $\Delta f = 0.0001$ Hz is illustrated in Fig. 2-3. The top panel shows the two stimuli (although only one waveform is apparent due to the small Δf), while the middle panel illustrates the two time-varying discharge rate waveforms from the AN model fiber. The bottom panel represents the approximation of the partial derivative squared as a function of time. The distinct growth of the derivative-squared as time increases up to 30 ms (bottom panel of Fig. 2-3) is a result of the increasing difference in the phase between the two tones as a function of time. Even if a random phase were used, as in some psychophysical experiments, the ability to distinguish between a difference in phase and a difference in frequency increases with duration. The bimodal shape within each stimulus period of the derivative-squared arises because the two rate waveforms differ both in the upward- and downward-going portions of each response

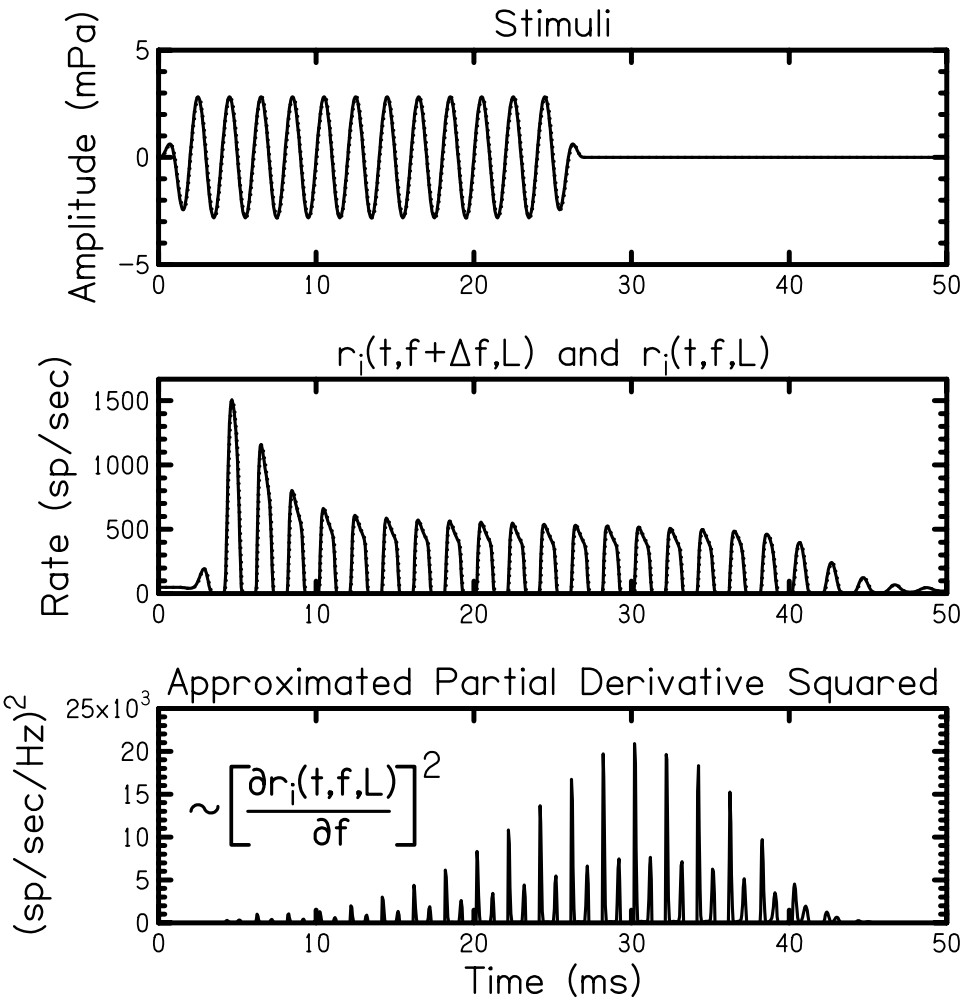

Figure 2-3: Approximation of the partial derivative in Eq. 2.1 for evaluating optimal frequency-discrimination performance for the stimulus condition: $L = 40$ dB SPL, $f = 500$ Hz, and $T = 25$ ms (2-ms rise/fall). Two tonal stimuli differing in frequency by $\Delta f = 0.0001$ Hz are shown in the top panel. The middle panel shows the two time-varying rate waveforms from the computational AN model ($CF = 500$ Hz) in response to the two stimuli. The two functions cannot be distinguished visually in the top and middle panels. The bottom panel illustrates the partial derivative squared, as approximated by Eq. 2.2.

period. Values of $\Delta f = 0.0001$ Hz and $\Delta L = 0.0001$ dB were used to approximate the partial derivatives in Eq. 2.2. These values were chosen by verifying that the derivatives for a variety of conditions were unaffected by increases in Δf or ΔL by an order of magnitude.

All predictions from the computational AN model are based on 60 model CF's ranging from 100 Hz to 10 kHz and uniformly spaced in location according to a human cochlear map (Greenwood, 1990; see Table 2.1). The choice of 60 model CF's represents a (somewhat arbitrary) compromise between a dense sampling of the frequency range and reduced computation time. The model CF range represents two orders of magnitude, while the commonly accepted human CF range [20 Hz - 20 kHz, Greenwood (1990)] represents three orders of magnitude. Thus, roughly two-thirds of the 30,000 total AN fibers in human (Rasmussen, 1940) are represented by the model. Also, the model only represents high-spontaneous-rate, low-threshold fibers, which make up 61% of the population in cat (Liberman, 1978). Thus, predictions in this study are based on a population of 12,200 total AN fibers, where each of the 60 model responses represent roughly 200 independent AN fibers.

Stimulus conditions were chosen based on psychophysical experiments of interest,[5] and a sampling rate of 500 kHz was used.[6] Stimulus duration was defined to be the duration between half-amplitude points on the stimulus envelope. All rise/fall ramps were generated from raised-cosine functions. The temporal-analysis window (Eq. 2.1) included the model response beginning at stimulus onset and ending 25 ms after stimulus offset, to allow for the response delay and the transient onset and offset responses associated with AN fibers [see Figs. 2-1(d) and 2-3] over the range of CF's and stimulus parameters used in this study.

[5]Two factors occasionally led to a slight mismatch between the stimulus parameters used in the present study and those that were used in the psychophysical studies. (1) Stimulus conditions were chosen to be the same for frequency- and level-discrimination predictions in order to allow for direct comparisons of model results between frequency and level discrimination. (2) Tone frequencies were always chosen to be equal to one of the 60 model CF's in order to avoid artifacts related to the spatial undersampling of the human frequency range.

[6]A sampling frequency of 500 kHz was used for all predictions, except for three high-frequency conditions for which higher sampling rates were chosen to avoid sub-harmonic distortion generated by the saturating nonlinearity for specific relations between stimulus and sampling frequency. In general, this sampling artifact was most prominent for high levels and high frequencies, and thus for high frequencies the present predictions were limited to low- to medium-levels.

2.6 Computational Results

2.6.1 Frequency discrimination in quiet

Predicted optimal frequency JND's were calculated as a function of frequency f, level L, and duration T for the all-information and rate-place encoding schemes using Eq. 2.1. Figure 2-4 shows the frequency discrimination predictions along with predictions from Siebert's (1970) study and human performance. Rate-place and all-information predictions from the computational AN model match predictions from Siebert (1970) well, both qualitatively and quantitatively, over the range of frequencies for which Siebert's analytical model is applicable ($200 < f < 2000$ Hz, $T \geq 100$ ms, and $20 \leq L \leq 50$ dB SPL).

The effect of frequency on pure-tone frequency discrimination is shown in Fig. 2-4(a). Rate-place predictions are essentially flat as a function of frequency, and match human performance very closely below 2 kHz. The predictions from the computational AN model show a slight improvement in performance as frequency increases due to the increase in the sharpness of tuning (quality factor $Q = CF/ERB$) as frequency increases (Table 2.1 and Glasberg and Moore, 1990).

All-information performance is roughly two orders of magnitude better than rate-place and human performance, consistent with Siebert (1970). At low frequencies, predicted performance based on all-information improves as frequency increases, due to the increased number of stimulus cycles as frequency increases [i.e., the two tones become more out of phase by the end of the stimulus duration for higher frequencies (see bottom panel of Fig. 2-3)]. All-information predictions, when extended to high frequencies, demonstrate a sharp worsening in performance above 2-3 kHz due to the rolloff of phase-locking at high frequencies [Fig. 2-1(c)]. This worsening matches the general trend observed in human data (Moore, 1973; Wier *et al.*, 1977, Moore and Glasberg, 1989). The all-information predictions converge towards the rate-place predictions as frequency increases, as expected due to the loss of temporal information. However, all-information performance is still one order-of-magnitude better than rate-place performance at 10 kHz [Fig. 2-4(a)], despite low synchrony coefficients at high frequencies [Fig. 2-1(c)]. The upper-CF limit in the model slightly influenced both rate-place and all-information by underestimating the amount of information available (by

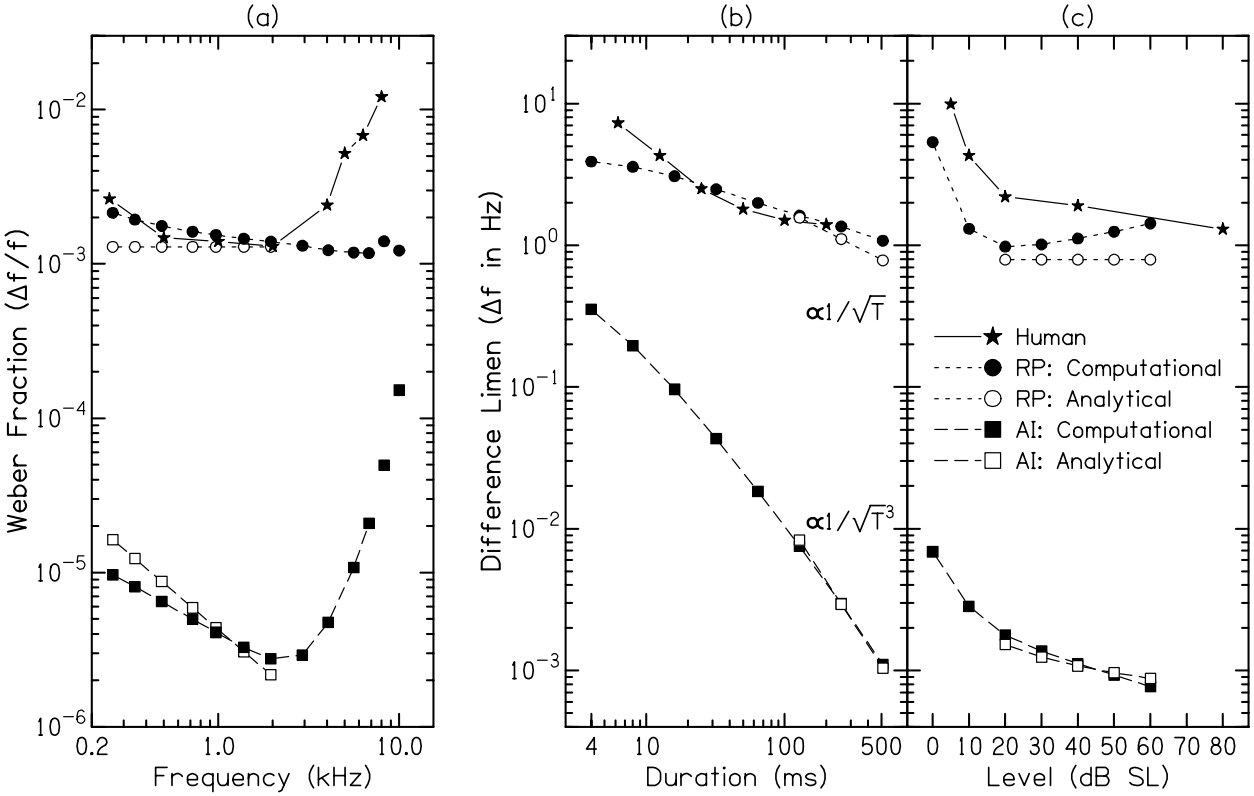

Figure 2-4: Optimal performance for pure-tone frequency discrimination. Predictions from the computational AN model are shown by filled circles for the rate-place (RP) scheme, and by filled squares for the all-information (AI) scheme. Predictions from Siebert's (1970) analytical model are shown for the two encoding schemes by corresponding open symbols. Typical human performance is illustrated by stars. (a) The Weber fraction $\Delta f/f$ is plotted as a function of frequency. All model predictions are for $L = 40$ dB SPL and $T = 200$ ms (20-ms rise/fall). Human data are from Moore (1973) for $T = 200$ ms and equal loudness across frequency ($L = 60$ dB SPL at 1 kHz). The two high-frequency points not connected by lines for each encoding scheme are conditions for which the upper limit of the computational model CF range influenced the results. (b) The difference limen Δf is plotted as a function of duration. Predictions are for $L = 40$ dB SPL, $f = 970$ Hz, and 4-ms rise/fall ramps. Human data are from Moore (1973) for $L = 60$ dB SPL, $f = 1000$ Hz and 2-ms rise/fall ramps. (c) The difference limen Δf is plotted as a function of level. Predictions were made for $f = 970$ Hz and $T = 500$ ms (20-ms rise/fall). Human data are from Wier *et al.* (1977) for $f = 1000$ Hz and $T = 500$ ms.

up to a factor of 2), and thus overestimating the JND (by up to a factor of $\sqrt{2}$).

Performance as a function of duration is shown in Fig. 2-4(b). Predictions for long durations can generally be described by $\Delta f \propto \sqrt{T^{-1}}$ for rate-place, and $\Delta f \propto \sqrt{T^{-3}}$ for all-information. Predicted performance from the all-information scheme improves much faster with duration than the rate-place predictions or the human data. The rapid improvement with increased duration is due to allowing the decision process to make all possible temporal comparisons throughout the entire time course of the stimulus (see the distinct growth of information with time in the bottom panel of Fig. 2-3). Both the rate-place and all-information predictions have a smaller slope at shorter durations than at longer durations, with the all-information slope matching the slope of human performance for $T \leq 16$ ms. The computational AN model included realistic onsets and offsets as well as neural adaptation, which would be expected to be significant for shorter durations.

Frequency discrimination performance as a function of level is shown in Fig. 2-4(c). At low levels, both rate-place and all-information predictions from the computational AN model demonstrate the same sharp improvement in performance with increasing level as seen in the human data. Above 20 dB SPL, rate-place performance is roughly flat, with the predicted JND's from the computational AN model demonstrating a slight increase as level increases. Both all-information and human performance continue to improve slowly as level increases. The improvement with level in the model is due to the increase in the number of excited fibers as level increases.

In order to understand the basis for predicted rate-place and all-information frequency-discrimination performance, it is useful to examine the distribution of information across the population of AN fibers. Information profiles $(\delta'_f[CF])^2$ are plotted in Fig. 2-5 for all 60 AN model CF's for low-, medium-, and high-frequency tones. The information profiles $(\delta'_f[CF])^2$ are plotted on a logarithmic scale to allow comparison across the three tone frequencies shown. The information profiles for the all-information and rate-place schemes are different in shape. A contiguous range of CF's centered at the frequency of the tone provides information in the all-information scheme. Above 10 dB SPL, fibers near the frequency of the tone have similar high levels of synchrony, and thus contribute nearly equally to frequency discrimination. In contrast, model AN fibers firing at the maximum average rate do not

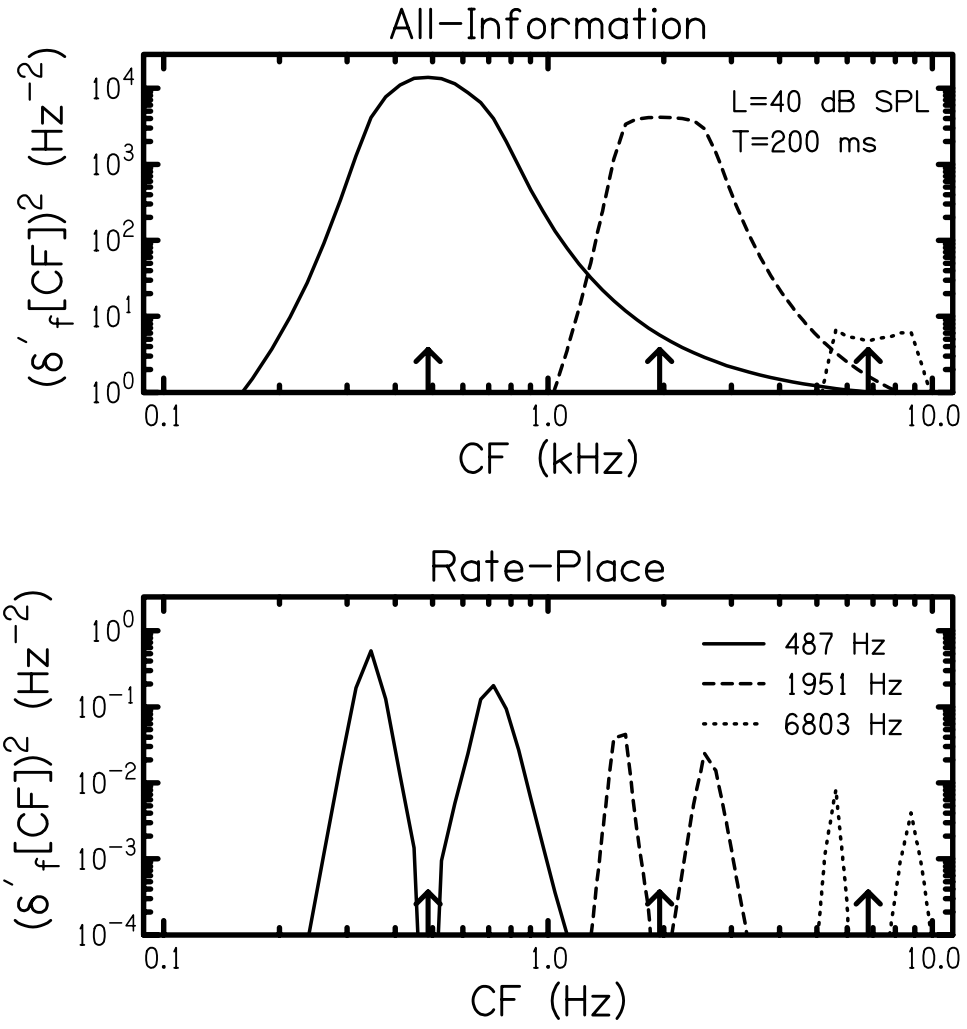

Figure 2-5: Information responsible for optimal frequency discrimination for low- (487 Hz), medium- (1951 Hz), and high-frequency (6803 Hz) conditions [5] (indicated by arrows) with $L = 40$ dB SPL and $T = 200$ ms (20-ms rise/fall). The top and bottom panels show the information $(\delta'_f[CF])^2$ for each AN model CF in the all-information and rate-place encoding schemes, respectively. The linear decline in the peaks of the rate-place information profiles as frequency increases contrasts the sharp drop in the all-information peaks at the highest frequency. These trends correspond directly to the flat rate-place $\Delta f/f$ and to the increasing all-information $\Delta f/f$ seen in Fig. 2-4(a) as frequency increases.

provide any information for estimating frequency in the rate-place encoding scheme. Only AN fibers at the edge of the rate-response area provide useful information for estimating frequency based on rate-place (Siebert, 1968).

2.6.2 Level discrimination in quiet

Predictions of optimal performance for pure-tone level discrimination are compared with human performance in terms of ΔL as a function of level, duration, and frequency (Fig. 2-6).[7] Rate-place predictions from an analytical AN model are also shown for comparison (Siebert, 1968). Optimal performance is roughly an order of magnitude better than human performance, while all-information JND's are roughly a factor of two lower than rate-place predictions [due to the level information from AN phase-locking (Colburn, 1981)].

Both the human and computational model predictions demonstrate a decrease in ΔL as level is increased in Fig. 2-6(a), i.e., the near-miss to Weber's Law. The slope of ΔL versus level in the computational model predictions is similar to that in the human data for levels above 30 dB SPL. Conversely, Siebert's (1968) model demonstrates Weber's Law above 40 dB SPL.

The computational model predictions for both all-information and rate-place illustrate the same rate of improvement in performance as a function of duration as is observed in the human data [Fig. 2-6(b)]. The rate of improvement is shallower in slope than the predictions from the analytical AN model (open circles), and than predictions for an optimal detector applied directly to the stimulus ($\Delta L \propto \sqrt{T^{-1}}$, solid line). The decay in discharge rate due to neural adaptation is likely to be responsible for the shallow slope in optimal performance demonstrated by the computational AN model.

The rate-place predictions are essentially flat as a function of frequency [Figure 2-6(c)], except for the highest-frequency conditions for which the upper-limit of the model CF range limited performance. The all-information predictions converge towards the rate-place pre-

[7]Level discrimination will be discussed in terms of the level difference between the two tones, defined as $\Delta L = 20 \log_{10} [(p + \Delta p)/p]$ dB, where p is the sound pressure of the standard tone. While there is much debate about the most appropriate metric for level discrimination (see Grantham and Yost, 1982; Green, 1988), ΔL has been reported to be proportional to sensitivity d' by several groups (e.g., Rabinowitz *et al.*, 1976; Buus and Florentine, 1991), and thus this metric will be used in the present study.

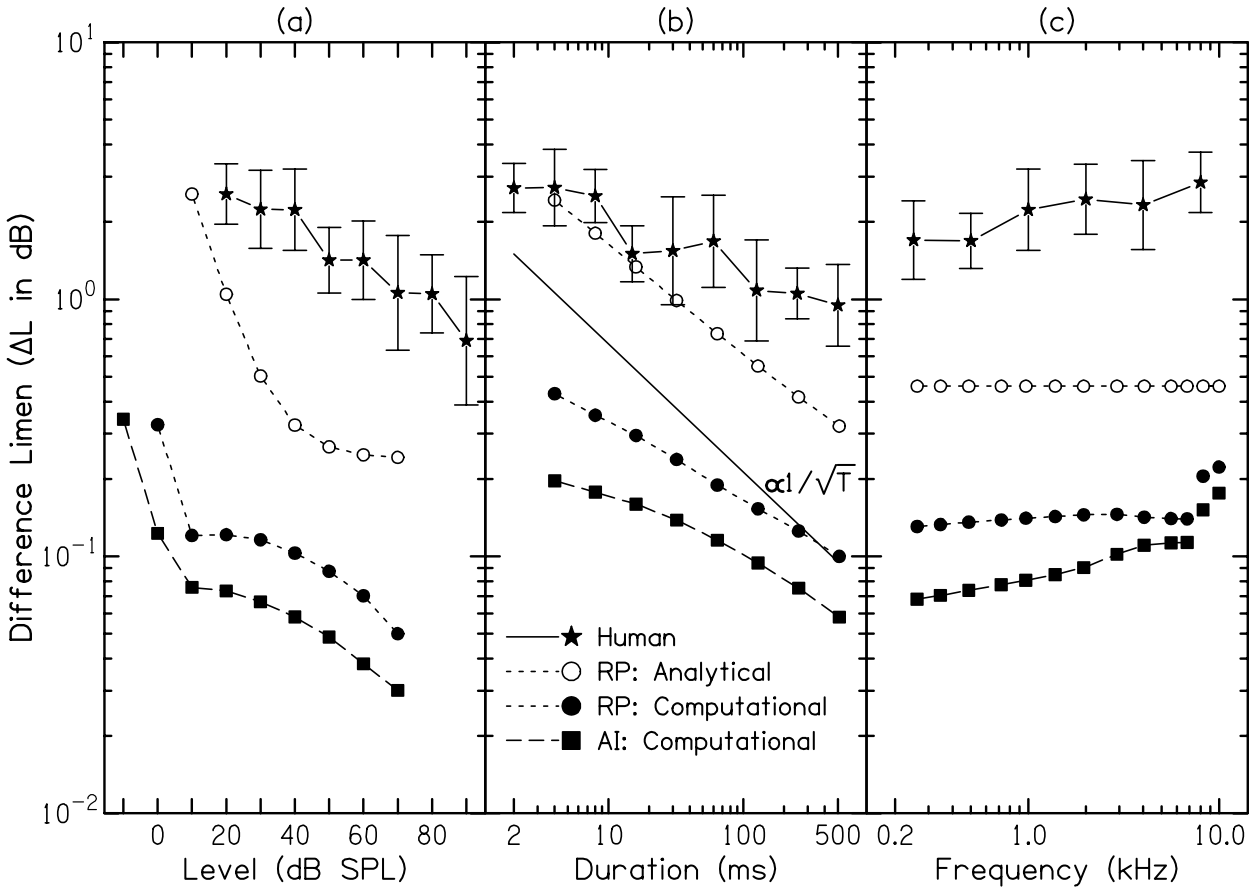

Figure 2-6: Optimal performance in terms of ΔL for pure-tone level discrimination. Predictions from the computational AN model are shown by filled circles for the rate-place scheme, and by filled squares for the all-information scheme. Rate-place predictions from an analytical AN model (Siebert, 1968) are shown as open circles. Typical human performance is illustrated by stars. (a) Effect of level. All model predictions are for $f = 970$ Hz and $T = 500$ ms (20-ms rise/fall). Human data are from Florentine *et al.* (1987) for $f = 1000$ Hz and $T = 500$ ms (20-ms rise/fall). (b) Effect of duration. Predictions are for $L = 40$ dB SPL, $f = 970$ Hz, and 4-ms rise/fall ramps. Human data are from Florentine (1986) for $L = 40$ dB SPL and $f = 1000$ Hz, and 1-ms rise/fall ramps. The solid line represents the predictions for an optimal detector applied directly to the stimulus, i.e., $\Delta L \propto T^{-1/2}$. (c) Effect of frequency. Predictions are for $L = 40$ dB SPL and $T = 200$ ms (20-ms rise/fall). Human data are from Florentine *et al.* (1987) for $L = 40$ dB SPL and $T = 500$ ms (20-ms rise/fall). The two high-frequency points not connected by lines for each encoding scheme are conditions for which the upper limit of the computational model CF range resulted in overestimates of the JND's.

dictions as frequency increases due to the loss of temporal information at high frequencies. The shallow slope versus frequency in the all-information predictions is similar to that in the human JND's for this stimulus level. A much sharper increase with frequency is observed in human data at high levels (Florentine *et al.*, 1987), but this effect is not likely to be present in model predictions at higher levels.

Level-discrimination information profiles (Fig. 2-7) for rate-place and all-information have the same general shape, unlike the information profiles for frequency discrimination (Fig. 2-5). For level discrimination, the edges of the response area provide the only information in the model for both the all-information and rate-place encoding schemes due to the saturation of both rate and synchrony as level increases. The primary source of the near-miss in the model predictions is an increase in the number of fibers that contribute significantly as level increases, especially above the frequency of the tone.

2.7 Discussion

2.7.1 Analytical versus computational approach

The limits on performance for level and frequency discrimination imposed by the random nature of AN responses were first examined roughly thirty years ago by Siebert (1965, 1968, 1970). The primary goal of the present study was to extend the optimal decision theory approach used by Siebert to allow the use of computational, rather than analytical, AN models. In order to verify the new method, a simplified AN model was used in the present study to permit direct comparisons to Siebert's predictions based on his analytical AN model.

The predictions from the present study matched Siebert's predictions well in almost all aspects over the parameter range for which the analytical AN model was applicable: (1) there is significantly more temporal information than rate-place information in the AN for encoding changes in frequency, (2) rate-place performance in terms of $\Delta f/f$ is roughly flat as a function of frequency, (3) all-information performance in terms of $\Delta f/f$ improves as frequency increases for low frequencies, (4) all-information performance improves as $\Delta f \propto \sqrt{T^{-3}}$, while rate-place improves as $\Delta f \propto \sqrt{T^{-1}}$, (5) all-information Δf continues to decrease as level in-

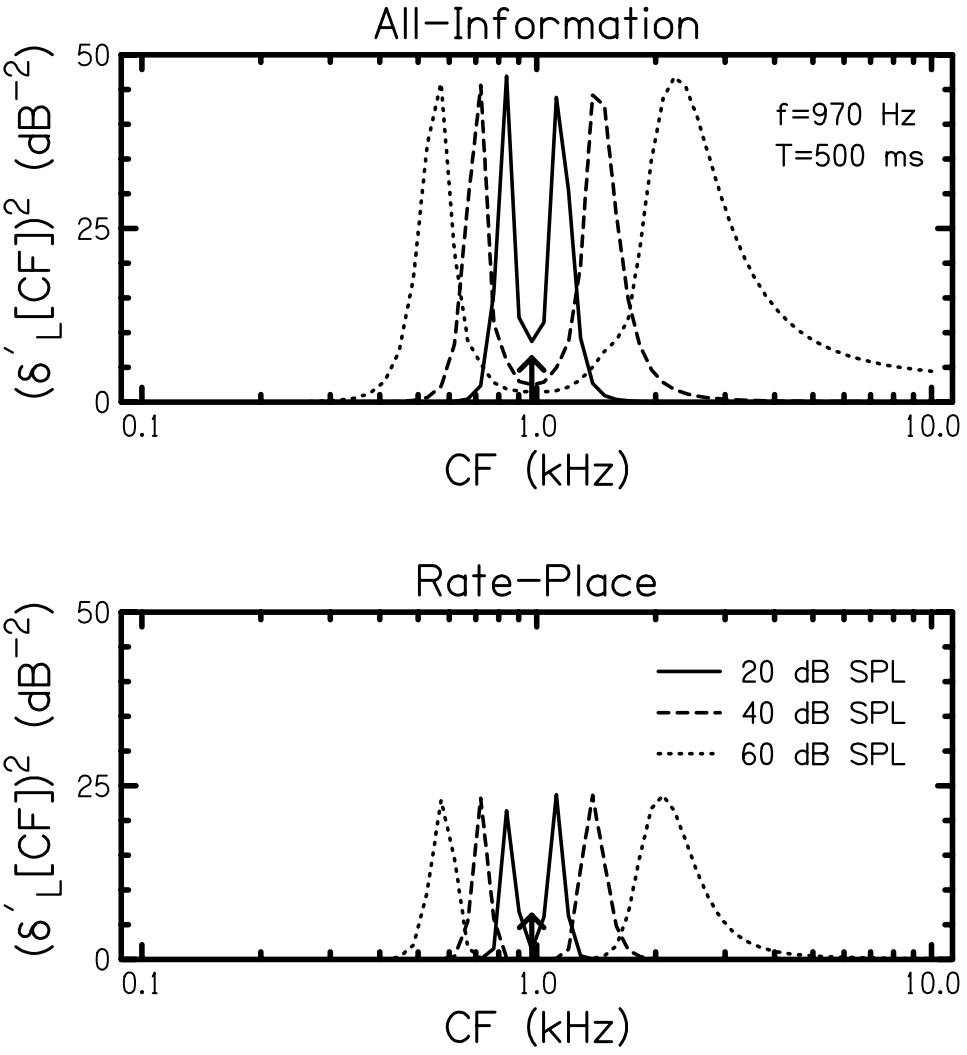

Figure 2-7: Information responsible for optimal level discrimination for three level conditions (20, 40, and 60 dB SPL) with $f = 970$ Hz (indicated by arrow)and $T = 500$ ms (20-ms rise/fall). The top and bottom panels show the information $(\delta'_L[CF])^2$ for each AN model CF in the all-information and rate-place encoding schemes, respectively.

creases, while rate-place Δf is roughly asymptotic above 20 dB SPL, (6) spread of excitation can produce constant level-discrimination performance over a wide range of levels based on only low-threshold fibers with a limited dynamic range, and (7) level discrimination is essentially independent of frequency. The predictions from the computational AN model differed from Siebert's predictions in two cases: (1) level discrimination in the computational model improved at a slower rate versus duration and similarly to human data, due to the effect of neural adaptation that was not included in Siebert's model, and (2) level-discrimination performance in the present model demonstrated the near-miss to Weber's Law that is observed in human performance, due to the nontriangular filters used in the computational model, whereas Siebert's model predicted Weber's Law. Both of these discrepancies represent cases in which the computational model provides a better match to human performance. The general agreement between the predictions from the computational and analytical models establishes the validity of the computational approach.

The predictions from the computational AN model extended Siebert's predictions in two cases: (1) all-information predictions were made at high frequencies due to the inclusion of the rolloff in phase-locking, and (2) predictions could be made at short durations due to the inclusion of realistic onset and offset responses and neural adaptation. Both of these extensions lead to important issues for auditory encoding of frequency and level, as discussed below.

2.7.2 Encoding of frequency

Rate-place

The absolute JND's for optimal use of rate-place information are much closer to human JND's than are all-information predictions (Fig. 2-4). Further, rate-place predictions match the trends in human performance versus frequency below 2 kHz and versus duration above 20 ms. However, there are several inconsistencies between predicted rate-place and human performance that must be explained if a rate-place model is to account for human performance.

(1) Predicted frequency-discrimination performance based on rate-place information is

inconsistent with the distinct worsening in human performance above 2 kHz [Fig. 2-4(a)]. The present prediction of invariant performance with frequency is consistent with all other rate-place models for frequency discrimination that incorporate realistic filter slopes as a function of frequency (e.g., Siebert, 1968, 1970; Javel and Mott, 1988; Delgutte, 1995). Thus, in order for a rate-place model to account for human frequency discrimination, some alteration must be included to account for the order-of-magnitude increase in $\Delta f/f$ between 2 and 8 kHz. Several possible explanations are considered below:

(a) A decrease in the efficiency of processing rate-place information above 2 kHz could explain human performance; however, it is not clear why a mechanism for counting AN discharges would be optimal at low frequencies, but an order-of-magnitude less efficient at high frequencies. (b) A reduction in the innervation density of independent AN fibers at high frequencies by a factor of 100 could explain the order-of-magnitude worsening in human performance. However, this is inconsistent with the increase in AN-fiber innervation density from apex to base of the cat cochlea (Keithley and Schreiber, 1987; Liberman *et al.*, 1990). (c) The maximum potential increase in $\Delta f/f$ that would result from the disappearance of the entire upper side of the information profile (Fig. 2-5) off the basal end of the cochlea as tone-frequency increases is $\sqrt{2}$, which is too small to account for human performance. (d) A decrease in the sharpness of tuning at high frequencies is inconsistent with reports of human [Glasberg and Moore (1990), used in the present study] and cat tuning (Liberman, 1978; Miller *et al.*, 1997), which become sharper as frequency increases.

Thus, there appears to be no physiologically realistic explanation for the worsening of rate-place performance as frequency increases above 2 kHz, as observed in the human data.

(2) The asymptotic rate-place performance as level increases (actually slight increase in computational-model JND's), is inconsistent with the shallow improvement in human performance as level increases [Fig. 2-4(c)]. The inclusion of high-threshold, low-spontaneous rate AN fibers could be expected to lead to improved performance at high levels. However, Erell (1988) demonstrated that inclusion of medium- and low-spontaneous-rate fibers [with higher thresholds and larger dynamic ranges (Liberman, 1978)] did not significantly extend the limited range of levels over which rate-place performance improves.

All-information

In contrast to rate-place, all-information performance matches the trends in human performance across all frequencies, especially at high frequencies, and across level [Figs. 2-4(a), (c)]. The current predictions quantify that there is significant temporal information in the AN up to at least 10 kHz, despite very low synchrony coefficients for AN fibers at high frequencies [Fig. 2-1(c)]. However, there are several inconsistencies between predicted all-information performance and human performance.

(1) There is a two-orders-of-magnitude discrepancy between human performance and optimal all-information performance at both low and high frequencies. (2) All-information performance for frequency discrimination improves much too rapidly as duration increases [$\Delta f \propto \sqrt{T^{-3}}$ rather than $\Delta f \propto \sqrt{T^{-1}}$; Fig. 2-4(b)]. This rapid improvement is due to allowing all possible comparisons between discharges to be used by the optimal processor (see Fig. 2-3).

Potential signal-processing mechanisms

It is of interest to hypothesize signal processing mechanisms that would incorporate the useful aspects of both rate-place and all-information encoding schemes for explaining human performance. A processor that makes use of the temporal discharge patterns appears to be required to produce the U-shaped dependence of $\Delta f/f$ on frequency [Fig. 2-4(a)]; however, the temporal information must be used inefficiently to account for the duration effect [Fig. 2-4(b)]. In a pilot modeling study, the computational AN model population was used to make predictions based on Goldstein and Srulovicz's (1977) analysis of first-order intervals.[8] While the first-order interval behavior demonstrated a $\Delta f \propto \sqrt{T^{-1}}$ relation at long durations ($T \geq 50$ ms), a $\Delta f \propto \sqrt{T^{-3}}$ relation was observed at short durations. The steep improvement at short durations is similar to Goldstein and Srulovicz's (1977) predictions, but is steeper

[8]There are several simplifying assumptions in Goldstein and Srulovicz's (1977) analysis of frequency-discrimination performance based on first-order intervals. Several of the simplifications that were made by assuming that RT was large (where R is average rate), were removed in the pilot study with the computational AN model and had little effect. However, the assumption that all intervals within the duration of the stimulus are independent, which is not true for a limited-duration stimulus or for a nonstationary Poisson process, was not able to be removed. The significance of this assumption for predicting optimal performance based on first-order intervals is not clear.

than the improvement in human performance at short durations [Fig. 2-4(b); Moore, 1973]. The U-shaped dependence of performance versus frequency was maintained by the first-order-interval scheme, similar to Goldstein and Srulovicz (1977), with a parallel upward shift of the entire curve by roughly one order of magnitude (i.e., roughly half-way closer to human performance).

A temporal processor that considered all discharge times (i.e., all-order intervals) within a limited temporal window would be expected to demonstrate the desired $\Delta f \propto \sqrt{T^{-1}}$ dependence for stimulus durations much longer than the length of the temporal window. For durations shorter than the temporal window, such a scheme would be expected to behave similarly to the all-information processor, which matches the rate of improvement in human performance with duration for $T \leq 16$ ms in the present study [Fig. 2-4(b)]. A restricted temporal processor of this type would demonstrate the same shallow improvement with level as the all-information scheme, consistent with human performance, and would be considerably closer to overall human performance levels due to the large amount of temporal information discarded [see Figs. 2-3 and 2-4(b)]. It is reasonable to assume that neurons in the CNS perform signal processing over restricted temporal windows, given that cell membranes have finite time constants; however, specific restricted temporal processors must be proposed and tested quantitatively to determine if the above expectations hold.

2.7.3 Encoding of level

The present predictions for pure-tone level discrimination did not demonstrate a large difference between performance based on the all-information and rate-place encoding schemes. Performance based on optimal use of the information available in the AN was roughly one order-of-magnitude better than human performance. Optimal performance demonstrated the same general trends seen in human performance for the conditions studied [Fig. 2-6], despite the absence of many physiological properties that have been hypothesized to be important for level discrimination, as discussed below.

The computational AN model used in the present study did not include several important aspects of the basilar-membrane (BM) compressive nonlinearity: changes in gain, bandwidth, and phase with level (Rhode, 1971; Anderson *et al.*, 1971; Geisler and Rhode, 1982; Ruggero

et al., 1997; Recio *et al.*, 1998), and suppression (Sachs and Kiang, 1968; Delgutte, 1990b; Ruggero *et al.*, 1992). The perceptual significance of the compressive gain observed on the BM has been primarily examined psychophysically (see Moore, 1995b, and Moore and Oxenham, 1998 for reviews). The compressive gain can extend the dynamic range over which changes in level can be detected in individual AN fibers (Heinz *et al.*, 1999). The level-dependent phase response of AN fibers has a much larger dynamic range than rate-information (Anderson *et al.*, 1971), and thus can provide additional temporal information for level discrimination at high levels for which the majority of AN fibers are saturated in terms of average rate (Carney, 1994; Carney *et al.*, 1999; Heinz *et al.*, 1999). The absence of suppression in the present model prohibits the accurate simulation of AN responses to complex stimuli. Suppression could potentially produce significant interactions between a target and an off-frequency noise masker that are not intuitively clear, nor consistent with the assumption that all fiber information is eliminated within the noise masker (e.g., Viemeister, 1983; Moore and Glasberg, 1989).

The near-miss to Weber's Law in the computational-model predictions matches human performance [Fig. 2-6(a)], but contrasts with Siebert's (1965, 1968) prediction of Weber's Law. Both the computational AN model and Siebert's model have only high-spontaneous-rate (HSR), low-threshold fibers, and rely strongly on spread of excitation across the AN population to encode changes in level for medium and high levels. In the computational model, an increase in the number of CF's contributing useful information as level increases, especially above the frequency of the tone, produces the near-miss (Fig. 2-7). Potentially significant differences between the computational AN model and Siebert's analytical model include: (1) the use of gamma-tone filters rather than triangular filters, (2) the form of the variation in filter bandwidth with CF, and (3) the distribution of CF across the population of AN fibers. The present results, and those of Teich and Lachs (1979), demonstrate that the near-miss does not require Weber's Law to hold in narrow frequency regions [an assumption that is often made in psychophysical models (e.g., Florentine and Buus, 1981)]. However, the simplified AN model used in the present study has several significant limitations that preclude any definitive conclusions regarding level encoding.

The present model would not likely predict Weber's Law for broadband noise (Miller,

1947) or for narrowband noise in notched-noise (Viemeister, 1983), or the nonmonotonic dependence of ΔL on level at high frequencies (Carlyon and Moore, 1984; Florentine *et al.*, 1987). A significant limitation of the AN model used in the present study is the exclusion of low-spontaneous-rate (LSR), high-threshold AN fibers (16% of the population; Liberman, 1978), which have been implicated in accounting for Weber's Law in limited frequency regions (Colburn, 1981; Delgutte, 1987; Viemeister, 1988a,b; Winslow and Sachs, 1988; Winter and Palmer, 1991). However, more efficient processing of the LSR fibers than the HSR fibers is required to achieve Weber's Law in narrow frequency regions, and thus the near-miss when information is combined across the population of CF's (Delgutte, 1987). A more complex AN model that includes the nonlinear changes in gain and phase as well as the LSR population is needed to quantify the relative contributions of these physiological properties to the encoding of sound level at high levels.

An interesting finding from the present study, despite the limitations mentioned above, is that the shallow improvement in ΔL with duration observed in the computational-model predictions matches human performance [Fig. 2-6(b)]. This shallow improvement is likely due to neural adaptation, as suggested by Buus and Florentine (1992), and indicates that a limitation in the temporal-integration capabilities of a central processor is not needed to explain the shallow improvement in level-discrimination performance as duration increases.

2.7.4 Relation between frequency and level encoding

The bandpass tuning in the auditory system implies that a change in rate on an AN fiber can result from either a change in stimulus level or frequency. Siebert (1968) used a simple AN model with triangular-shaped filters to demonstrate a fundamental relationship between level and frequency discrimination based on rate-place information. A shortcoming of rate-place models has been their inability to account for both frequency and level discrimination in terms of the ratio of Weber fractions W_A/W_F (Siebert, 1968; Erell, 1988; Delgutte, 1995). Siebert predicted that the ratio of the Weber fraction for amplitude, $W_A = \Delta A/A$, to the Weber fraction for frequency, $W_F = \Delta f/f$, should be roughly 15 for rate-place models based on the shapes of AN tuning curves. Values of 10-15 for this ratio are typical of most rate-place models; however, human psychophysical data typically yield a ratio closer to 50 (Erell, 1988).

Erell (1988) suggested that this discrepancy could be resolved by using filter bandwidths that were four to five times narrower than those based on tuning in cat, and that human tuning is sharper than in cat. Typical values of the ratio W_A/W_F from the present study (based on human tuning) were 11 for rate-place and 710 for all-information (for $f = 970$ Hz, $L = 40$ dB SPL, $T = 64$ ms, and 4-ms rise/fall). Human behavior lies between the predicted behavior based on rate-place and all-information encoding schemes. This result supports the notion that a restricted temporal encoding scheme should be pursued (i.e., an encoding scheme that is intermediate to rate-place and all-information).

2.7.5 Limitations of the present study

Computational auditory-nerve model

The computational AN model used in this study has a basic filter shape that is the same for low and high CF's, with the only difference being filter bandwidth. This assumption ignores the tails of AN tuning curves observed for high-CF AN fibers (Kiang and Moxon, 1974). Inclusion of tails would affect the details of the predicted performance as a function of level; however, the significance of this assumption is minimized for the predictions versus frequency and duration, which were limited to low- and mid-level stimuli. The omission of tuning-curve tails also limits the applicability of the present AN model for complex sounds.

The Poisson assumption used in the present model ignores refractory effects, and could affect some of the details of the performance limits predicted in this study. Siebert (1965, 1968, 1970) and Colburn (1969, 1973) have argued that this assumption is not significant because the AN never responds at a sustained rate for which the typical inter-spike interval is less than four to five times larger than the absolute refractory period; however, this argument does not rule out a significant influence of the relative refractory period. The standard deviation of AN discharge counts has been reported to be less than that predicted by the Poisson model (Teich and Khanna, 1985; Young and Barta, 1986; Delgutte, 1987; 1996; Winter and Palmer, 1991). The maximum deviation from the Poisson model occurs at high discharge rates, and the Poisson model becomes more accurate as discharge rate decreases. Young and Barta (1986) reported that the standard deviation of discharge counts ranged

from 55% to 80% of the Poisson value at high discharge rates, values that are consistent with other reports. Thus, the maximum discrepancy from using the Poisson model could be an overestimate of some rate-place JND's by a factor of two; however, the discrepancy will typically be much less since fibers with high discharge rates do not contribute significant rate-place information (see Figs 2-5 and 2-7; Colburn, 1981; Winter and Palmer, 1991). Miller *et al.* (1987) provided evidence that single-unit phase-locked responses are consistent with a nonstationary Poisson model. They showed that the mean and variance of period histograms had roughly the same pattern and absolute value. Thus, the errors from using a Poisson model without refractory effects are small compared with the important trends in the present predictions, which often range over more than an order of magnitude.

Optimal decision theory

The Cramér-Rao-bound analysis in the present study is restricted in that it only provides limits on the performance of all possible decision processes for a Poisson process, and thus should be viewed primarily as a method for quantifying the total information available in the auditory nerve. This method does not describe the type of processing required to achieve optimal performance, which is often significantly better than human performance (e.g., Fig. 2-4). Derived optimal processors typically include many physiologically unrealistic operations on the observed population of AN discharges, such as the ability to compare all discharge times across the entire duration of the stimulus, as well as the ability to compare the discharges across all AN fibers (Colburn, 1969, 1973; Siebert, 1970). It is often the case that physiologically realistic, but non-optimal, processors can be described that perform at levels more consistent with human performance (e.g., Colburn, 1969, 1973, 1977a,b; Goldstein and Srulovicz, 1977; see Delgutte, 1996). This approach provides intuition regarding the actual processing performed in the auditory system, and is thus a natural second step after the fundamental performance limits have been described based on the total AN information.

The ODT used in the present study is applicable to any single-parameter discrimination task (e.g., frequency or level discrimination). A useful extension will be to incorporate the potential for adding random noise to the stimulus; however, this will require a significant extension of the ODT used in the present study in order to quantify the influence of stimulus

(external) noise. However, this extension is now warranted because computational AN models that produce responses to arbitrary stimuli provide the potential to predict the effects of noise maskers on psychophysical performance. As mentioned above, there is particular interest in the potential to evaluate quantitatively the effect of noise maskers in experiments that have been used to test rate-place and temporal encoding schemes (e.g., Viemeister, 1983; Moore and Glasberg, 1989). In addition to the effects of suppression, the information contained in the temporal patterns (either fine-time or envelope) of AN fiber responses may be significant in these studies.

2.8 Conclusions

Optimal decision theory can be combined with computational models to predict psychophysical performance limits based on an entire neural population. Auditory discrimination performance limits based on the computational model matched predictions from previous studies over the parameter range for which analytical models were applicable. The parameter range over which psychophysical performance can be predicted was extended using the computational AN model.

The following three conclusions related to the encoding of frequency are not likely to depend on the details of the AN model used in the present study. Optimal use of rate-place information is consistent with human frequency discrimination performance at low frequencies, but is inconsistent with the trends in performance versus frequency at high frequencies. Frequency discrimination based on optimal use of all-information is roughly two orders-of-magnitude better than human performance at all frequencies, but demonstrates the correct trends versus frequency, based on the frequency-dependence of phase-locking in the cat. There is significant temporal information in the AN for frequency discrimination up to at least 10 kHz, and thus temporal schemes cannot be rejected at high frequencies based on the decrease in phase-locking in the AN.

Future studies using the computational approach with more complex AN models are now justified, and will be particularly interesting for the task of level discrimination. In addition, this approach is applicable to any neural population in a sensory system for which a model

of the discharge-pattern statistics is available.

2.9 Acknowledgments

The authors would like to thank Bertrand Delgutte, Christopher Long, Christine Mason, Martin McKinney, Susan Moscynski, Bill Peake, Tom Person, Timothy Wilson, Xuedong Zhang, and Ling Zheng for providing valuable comments on an earlier version of this manuscript. We thank Mary Florentine for providing the data from Florentine (1986) shown in Fig. 2-4(b), and Don Johnson for his synchrony data from cat shown in Fig. 2-1(c). Portions of this work were presented at the 21st and 22nd Midwinter Meeting of the Association for Research in Otolaryngology. The simulations in this study were performed on computers provided by the Scientific Computing and Visualization group at Boston University. This work was supported in part by the National Institute of Health, Grants T32DC00038 and R01DC00100, as well as by the Office of Naval Research, Grant Agmt #Z883402.

2.10 Appendix: Performance Limits for One-Parameter Discrimination

This appendix describes a complete analysis of a general one-parameter discrimination experiment using ODT, and extends several previous results (e.g., Siebert, 1970; Colburn, 1981; Rieke *et al.*, 1997). This analysis is applicable to any physiological model that assumes Poisson discharge statistics. Prediction of performance limits based on the Cramér-Rao bound is described first, because this was the method first described by Siebert (1970); however, this approach lacks an explicit description of the processing that is required to achieve optimal performance. The use of a likelihood ratio test, which is described second, provides insight into the meaning of the Cramér-Rao bound result, describes the form of an optimal processor, and demonstrates that the Cramér-Rao bound is efficient [i.e., can be met with equality (van Trees, 1968)].

The assumption that a nonstationary Poisson process is a good model for the AN discharge patterns allows the random nature of the observations to be described (see Parzen,

1962; Snyder and Miller, 1991; Rieke *et al.*, 1997). The joint probability density of the unordered discharge times $\{t_1^i, \cdots, t_{K_i}^i\}$ on the i^{th} AN fiber is given by

$$p(t_1^i, \cdots, t_{K_i}^i|\alpha) = \frac{\prod_{n=1}^{K_i} r_i(t_n^i, \alpha)}{K_i!} \exp\left[-\int_0^T r_i(t, \alpha)dt\right], \tag{2.3}$$

where $r_i(t, \alpha)$ is the time-varying discharge rate, and α is the stimulus parameter to be discriminated. [Note that this formula applies to unordered discharge times (i.e., the discharge times are *not* constrained such that $t_1^i < t_2^i < \cdots < t_{K_i}^i$).]

Cramér-Rao Bound: The Poisson description of the random nature of the observations can be used in the Cramér-Rao bound (Cramér, 1951; van Trees, 1968) to describe optimal performance for estimating the value of α based on the response of a single AN fiber in terms of its time-varying discharge rate $r_i(t, \alpha)$. The variance $\sigma_{\hat{\alpha}}^2[i]$ of any unbiased estimate of the parameter α from the observed discharge times on the i^{th} fiber is bounded by the Cramér-Rao inequality

$$\frac{1}{\sigma_{\hat{\alpha}}^2[i]} \leq E\left\{\left[\frac{\partial}{\partial\alpha}\ln p(t_1^i, \cdots, t_{K_i}^i|\alpha)\right]^2\right\}, \tag{2.4}$$

$$= \sum_{K_i=0}^{\infty}\int_0^T \cdots \int_0^T \left[\frac{\partial}{\partial\alpha}\ln p(t_1^i, \cdots, t_{K_i}^i|\alpha)\right]^2 p(t_1^i, \cdots, t_{K_i}^i|\alpha)dt_1^i \cdots dt_{K_i}^i, \tag{2.5}$$

where the expectation in Eq. 2.4 is over the random observations $\{t_1^i, \cdots, t_{K_i}^i\}$. This inequality provides a lower bound on the variance of any unbiased estimator by relating the accuracy of the estimate to the rate of change with α of the likelihood of the observations. The variance $\sigma_{\hat{\alpha}}^2[i]$ can be related to the time-varying discharge rate $r_i(t, \alpha)$ by substituting the joint probability density from Eq. 2.3 into Eq. 2.5 and performing some extensive simplifications to obtain

$$\frac{1}{\sigma_{\hat{\alpha}}^2[i]} \leq \int_0^T \frac{1}{r_i(t, \alpha)}\left[\frac{\partial r_i(t, \alpha)}{\partial\alpha}\right]^2 dt. \tag{2.6}$$

Equation 2.6 provides a lower bound on the variance of any unbiased estimate of α based on

the observations from a single AN fiber in terms of the time-varying discharge rate $r_i(t, \alpha)$.

The just-noticeable difference $\Delta\alpha_{JND}$ of the ideal observer is equal to the minimum standard deviation of any estimator based on the observations, when threshold is defined as 75% correct in a two-interval, two-alternative forced-choice task. This definition of threshold corresponds to $d' = 1$, where $d' = \Delta\alpha_{JND}/\sigma_{\hat{\alpha}}[i]$. The JND based on the population of AN fibers is given by

$$\Delta\alpha_{JND} = \left\{ \sum_i \int_0^T \frac{1}{r_i(t,\alpha)} \left[\frac{\partial r_i(t,\alpha)}{\partial \alpha} \right]^2 dt \right\}^{-\frac{1}{2}}, \tag{2.7}$$

because the information from each AN fiber adds under the assumption that the discharge patterns of all fibers are statistically independent.

Likelihood Ratio Test: This analysis extends previous results from Colburn (1981) and Rieke *et al.* (1997) to show that evaluation of the performance of a sufficient statistic for a nonstationary Poisson process derived from a likelihood-ratio test leads to the same equation obtained from the Cramér-Rao bound (i.e., that an efficient estimator exists). This alternative approach provides more intuition, and is therefore more accessible than the Cramér-Rao bound analysis. Most importantly, it is shown that the quantity

$$(\delta'_\alpha[CF_i])^2 \triangleq \int_0^T \frac{1}{r_i(t,\alpha)} \left[\frac{\partial r_i(t,\alpha)}{\partial \alpha} \right]^2 dt \tag{2.8}$$

represents the information available on the i^{th} AN fiber for discriminating α, where $\delta'_\alpha[CF_i]$ represents the *normalized sensitivity* of the i^{th} fiber, and is defined as the sensitivity d' per unit α (see Durlach and Braida, 1969; Braida and Durlach, 1988).

The form of the optimal processor for discriminating between α and $\alpha + \Delta\alpha$ can be

derived from a log-likelihood test,

$$\ln \frac{p(t_1^i, \cdots, t_{K_i}^i | \alpha + \Delta\alpha)}{p(t_1^i, \cdots, t_{K_i}^i | \alpha)} \begin{array}{c} \alpha + \Delta\alpha \\ > \\ < \\ \alpha \end{array} 0, \tag{2.9}$$

where the threshold of 0 corresponds to the minimum-probability-of-error criterion with equal *a priori* probabilities (van Trees, 1968, pp. 23-30). Substituting the joint probability density of the discharge times (Eq. 2.3) into Eq. 2.9 and simplifying, one obtains the following form of an optimal test

$$Y(t_1^i, \cdots, t_{K_i}^i) \triangleq \sum_{n=1}^{K_i} \ln \frac{r_i(t_n^i, \alpha + \Delta\alpha)}{r_i(t_n^i, \alpha)} \begin{array}{c} \alpha + \Delta\alpha \\ > \\ < \\ \alpha \end{array} \int_0^T [r_i(t, \alpha + \Delta\alpha) - r_i(t, \alpha)]\, dt. \tag{2.10}$$

Since the decision variable Y is a sum of random variables, a Gaussian assumption is reasonable if K_i is large, and $(d')^2$ is an appropriate performance measure. Thus, optimal performance based on $Y(t_1^i, \cdots, t_{K_i}^i)$ can be evaluated as (see Green and Swets, 1966)

$$(d'[CF_i])^2 = \frac{\left\{ E\left[Y(t_1^i, \cdots, t_{K_i}^i) \middle| \alpha + \Delta\alpha \right] - E\left[Y(t_1^i, \cdots, t_{K_i}^i) \middle| \alpha \right] \right\}^2}{Var\left[Y(t_1^i, \cdots, t_{K_i}^i) \middle| \alpha \right]}. \tag{2.11}$$

It can be shown (e.g., see Rieke *et al.*, 1997) that for any decision variable Y of the form

$$Y(t_1, \cdots, t_K) = \sum_{n=1}^{K} g(t_n), \tag{2.12}$$

where $g(t_n)$ is any function of the Poisson discharge time t_n, that the expected value and

variance of Y are given by

$$E\left[Y(t_1,\cdots,t_K)|\,\alpha\right] = \int_0^T g(t) r_i(t,\alpha) dt, \tag{2.13}$$

$$Var\left[Y(t_1,\cdots,t_K)|\alpha\right] = \int_0^T g^2(t) r_i(t,\alpha) dt. \tag{2.14}$$

Given these relations with $g(t) = \ln[r_i(t,\alpha+\Delta\alpha)/r_i(t,\alpha)]$,

$$\begin{aligned}\left(d'[CF_i]\right)^2 &= \left\{\int_0^T \ln\frac{r_i(t,\alpha+\Delta\alpha)}{r_i(t,\alpha)}\left[r_i(t,\alpha+\Delta\alpha)-r_i(t,\alpha)\right]dt\right\}^2 \Bigg/ \\ &\qquad \left\{\int_0^T \left[\ln\frac{r_i(t,\alpha+\Delta\alpha)}{r_i(t,\alpha)}\right]^2 r_i(t,\alpha)dt\right\}. \end{aligned}\tag{2.15}$$

With the additional assumption that $r_i(t,\alpha)$ varies linearly and slowly over the incremental range from α to $\alpha+\Delta\alpha$ (Colburn, 1981), i.e.,

$$r_i(t,\alpha+\Delta\alpha) \simeq r_i(t,\alpha) + \dot{r}_i(t,\alpha)\Delta\alpha, \tag{2.16}$$

where $\dot{r}_i(t,\alpha) = \partial r_i(t,\alpha)/\partial\alpha$, one obtains

$$\begin{aligned}\left(d'[CF_i]\right)^2 &\simeq \left\{\int_0^T \ln\left[1+\frac{\dot{r}_i(t,\alpha)\Delta\alpha}{r_i(t,\alpha)}\right]\left[\dot{r}_i(t,\alpha)\Delta\alpha\right]dt\right\}^2 \Bigg/ \\ &\qquad \left(\int_0^T \left\{\ln\left[1+\frac{\dot{r}_i(t,\alpha)\Delta\alpha}{r_i(t,\alpha)}\right]\right\}^2 r_i(t,\alpha)dt\right). \end{aligned}\tag{2.17}$$

The approximation $\ln(1+x) \simeq x$ for small x then leads to the equations

$$\begin{aligned}\left(d'[CF_i]\right)^2 &\simeq \left[(\Delta\alpha)^2\int_0^T \frac{\dot{r}_i(t,\alpha)^2}{r_i(t,\alpha)}dt\right]^2 \Bigg/ \left[(\Delta\alpha)^2\int_0^T \frac{\dot{r}_i(t,\alpha)^2}{r_i(t,\alpha)}dt\right] \\ &= (\Delta\alpha)^2\int_0^T \frac{1}{r_i(t,\alpha)}\left[\frac{\partial r_i(t,\alpha)}{\partial\alpha}\right]^2 dt \qquad (2.18) \\ &= (\Delta\alpha)^2\left(\delta'_\alpha[CF_i]\right)^2. \qquad (2.19)\end{aligned}$$

Note that the "small-x" approximation may not be appropriate for low discharge rates. To

avoid this potential problem in the valleys of the phase-locked response, a constant was added to all discharge rates $r(t, f, L)$ prior to the ODT analysis in the present study. [A value of $\simeq 7$ spikes/s was used, based on the minimum value of $r(t, f, L)$ in Siebert's (1970) model. The precise value of this constant is not significant, as long as the instantaneous discharge rate does not go to zero.]

Finally, taking $d' = 1$ as the JND and combining over independent fibers, one obtains

$$\Delta\alpha_{JND} \simeq \left\{ \sum_i \left(\delta'_\alpha[CF_i]\right)^2 \right\}^{-\frac{1}{2}}, \tag{2.20}$$

which is the same form derived from the Cramér-Rao bound (Eq. 2.7). [Note that the summation over fibers in Eq. 2.20 results from the independence assumption, and would come directly in the analysis if the probability function in Eq. 2.3 were written as a product of the fiber probabilities.]

Chapter 3

Evaluating auditory performance limits: II. One-parameter discrimination with random level variation

3.1 Abstract

Previous studies have combined analytical models of neural responses with optimal decision theory (ODT) to predict psychophysical performance limits; however, these studies have typically been limited to simple models and simple psychophysical tasks. A companion paper describes an extension of the ODT approach to allow the use of computational models that provide more accurate descriptions of neural responses. The present paper describes an extension to more complex psychophysical tasks. A general method is presented for evaluating psychophysical performance limits for discrimination tasks in which one stimulus parameter is randomly varied. Psychophysical experiments often randomly vary a single parameter in order to restrict the cues that are available to the subject. The new method is demonstrated for the auditory task of random-level frequency discrimination using a computational auditory-nerve (AN) model. Performance based on AN discharge times (*all information*) is compared to performance based only on discharge counts (*rate place*). Both decision models are successful in predicting that random level variation has no effect on performance in quiet, which is the typical result in psychophysical tasks with random level variation. The distribution of information across the AN population provides insight into how different types of AN information can be used to avoid the influence of random level variation. The rate-

[0]A similar version of this chapter has been submitted to *Neural Computation* (Heinz *et al.*, 2000b).

place model relies on comparisons between fibers above and below the tone frequency (i.e., the population response), while the all-information model does not require such across-fiber comparisons. Frequency discrimination with random level variation in the presence of high-frequency noise is also simulated. No effect is predicted for all-information, consistent with the small effect in human performance; however, a large effect is predicted for rate-place in noise with random level variation.

3.2 Introduction

The use of optimal decision theory (ODT) combined with models of neural responses has provided much insight into neural encoding of sensory stimuli (e.g., Fitzhugh, 1958; Siebert, 1965,1968,1970; Colburn, 1969, 1973, 1977a,b, 1981; Goldstein and Srulovicz, 1977; Delgutte, 1987; Erell, 1988; see Delgutte, 1996 and Parker and Newsome, 1998 for reviews). These studies evaluated psychophysical performance limits based on the stochastic behavior of neural responses. However, the application of this approach has been limited to simple psychophysical tasks due to the use of simple analytical models and by the restricted use of ODT to deterministic stimuli. In a companion paper, an extension of the ODT approach to incorporate computational neural models is presented, which allows a much wider range of stimuli to be studied (Chapter 2). The present study extends the ODT analysis to allow more complicated psychophysical tasks to be evaluated, specifically discrimination tasks in which one parameter is randomly varied from trial to trial.

Many psychophysical experiments have used random variation of certain stimulus parameters in order to limit the cues available to the subject. For example, McKee *et al.* (1986) observed that human velocity discrimination was unaffected by random variation of either contrast or temporal frequency, and concluded that performance was mediated by sensing velocity. In the auditory system, this method has been used in profile-analysis experiments to demonstrate that level discrimination of a single component of a tone complex, or detection of a tone in noise, is not affected by the randomization of overall level, and thus these tasks can be performed without relying on an absolute energy cue (Green *et al.*, 1983; Kidd *et al.*, 1989; see Green, 1988 for a review). While psychophysical performance in the presence of overall-level randomization is typically unchanged, performance based on single neurons would likely be severely degraded. Therefore, it is important to quantitatively

evaluate how physiological responses can account for behavior in these types of psychophysical tasks. Durlach *et al.* (1986) described a quantitative model for profile-analysis tasks based on across-frequency level comparisons, which included the effects of both external (stimulus) variations and internal processing noise. However, the internal noise used in their model was not directly related to the known physiological noise that exists in AN fibers, and thus was somewhat arbitrary. The present study extends the approach used by Siebert and others to include the effect of random stimulus variation in a single parameter on psychophysical performance limits. In order to demonstrate this new method, predictions for a random-level frequency discrimination task are evaluated using the same computational auditory-nerve (AN) model used in Chapter 2. Predictions are made for both *rate-place* and *all-information* (including temporal and rate-place information) encoding schemes, as was done in the companion study of pure-tone frequency and level discrimination.

Random level variation has been used in several frequency-discrimination experiments to test rate-place models for frequency encoding (Henning, 1966; Verschuure and van Meeteren, 1975; Emmerich *et al.*, 1989; Moore and Glasberg, 1989). In this task, the listener is asked to discriminate the frequency of two tones whose levels are varied randomly and independently from trial to trial. Frequency-discrimination-in-quiet could hypothetically be performed by observing the average discharge rate of a single frequency channel tuned either above or below the tone frequency (e.g., Zwicker, 1956; 1970; Henning, 1967). Such rate-based models would be expected to be greatly affected by random level variation because changes in level could not be discriminated from changes in frequency. Conversely, temporal models would not be expected to be affected by random level variation. It is often stated that listeners use rate-place information to encode frequency at high frequencies, (e.g., Wever, 1949; Moore, 1973, 1989; Dye and Hafter, 1980; Wakefield and Nelson, 1985; Javel and Mott, 1988; Pickles, 1988; Moore and Glasberg, 1989), because AN phase-locking rapidly degrades above 2-3 kHz [Johnson, 1980; Joris *et al.*, 1994; or see Fig. 2-1(c)].

Several studies have observed an effect of random level variation on frequency discrimination (e.g., Henning 1966, Emmerich *et al.*, 1989); however, the observed effect of random level variation is likely to be largely due to pitch shifts associated with the changes in level over the broad range of level variation used in these studies (Verschuure and van Meeteren, 1975;

Emmerich *et al.*, 1989). Moore and Glasberg (1989) measured frequency discrimination as a function of frequency with a smaller range of random level variation, and observed virtually no effect. Moore and Glasberg also measured frequency discrimination in the presence of high-frequency noise that was used to mask the characteristic frequencies (CF's) above the tone frequency. This experiment was designed to test the idea that rate-place models could avoid the effect of random level variation by comparing information from CF's above and below the tone frequency. A small but significant effect of adding the high-frequency noise was observed, with a slightly larger effect when the noise was added to the random-level condition. Despite a similar effect at all tone frequencies, Moore and Glasberg (1989) concluded that their results were consistent with the "duplex theory" for frequency encoding (Wever, 1949), i.e., that rate-place information is used at high frequencies and temporal information is used at low frequencies. The results in the present study question this conclusion by quantifying that there is insufficient rate-place information in the AN model to account for human performance with random level variation under the assumption that the high-frequency noise masks all CF's above the tone frequency. The quantitative approach described in the present study can be used for any sensory system for which there are statistical descriptions of neural responses as a function of the relevant stimulus parameters.

3.3 General Methods

3.3.1 Computational auditory-nerve model

The computational AN model used in the present study was the same model described in Chapter 2.[1] This AN model was a simplified version of a previous nonlinear AN model (Carney, 1993), and was used in order to simplify the verification of the new ODT methods. The major components of the AN model are summarized below [see Chapter 2 for a detailed description].

The initial model stage was a linear fourth-order gamma-tone filter bank, which was used to represent the frequency selectivity of AN fibers. Model filter bandwidths were based

[1]Code for the AN model used in the present study is available at http://earlab.bu.edu/.

on estimates of human bandwidths from Glasberg and Moore (1990). Each bandpass filter was followed by a memoryless, asymmetric, saturating nonlinearity, which represents the mechano-electric transduction of the inner hair cell (IHC). All AN model fibers had a rate threshold of roughly 0 dB SPL, a spontaneous rate of 50 spikes/s, and a maximum sustained rate of roughly 200 spikes/s. The model dynamic range for sustained rate was roughly 20-30 dB, while the dynamic range for onset rate was much larger. The rolloff in phase-locking was consistent with all species discussed in Weiss and Rose (1988), and the cutoff frequency matched data from cat (Johnson, 1980). Neural adaptation was introduced through a simple three-stage diffusion model for the IHC-AN synapse based on data from Westerman and Smith (1988). The output of the i^{th} AN model fiber represents the instantaneous discharge rate $r_i(t, f, L)$ of an individual high-spontaneous-rate, low-threshold AN fiber in response to an arbitrary stimulus. The AN discharges are assumed to be produced by a population of nonstationary Poisson processes with rate functions described by $r_i(t, f, L)$.

3.3.2 Optimal decision theory

The application of ODT in the present study is extended from the companion study (Chapter 2) to evaluate psychophysical tasks in which a single parameter is randomly varied. Optimal performance is evaluated for both rate-place and all-information models and compared to data from human listeners. *Rate-place* predictions are based on the assumption that the population of AN-fiber discharge counts $\{K_i\}$ over the duration of the stimulus is the only information used by the listener. In contrast, the *all-information* predictions are based on the assumption that the listener uses the population of discharge times and counts $\{t_1^i, \cdots, t_{K_i}^i\}$, where t_j^i represents the j^{th} discharge from the i^{th} AN fiber. The contribution of temporal information in the responses can be inferred by comparing the predictions of the rate-place and all-information models. The all-information model does not assume any specific forms of temporal processing, such as calculating synchrony coefficients or creating interval histograms, and thus provides an absolute limit on achievable performance given the total information available in the AN.

3.4 General Analytical Results: Random-Level Frequency Discrimination

The use of ODT with stochastic models of neural responses has never been used to evaluate the many psychophysical tasks that randomly vary one parameter in order to restrict the cues that are available to the subject. The ODT analysis of a general one-parameter discrimination experiment with one unwanted random parameter is a straightforward extension to the multiple-parameter case (see Appendix in Section 3.10). In the present study, this analysis is described in terms of a random-level frequency-discrimination task. The just-noticeable difference in frequency for this task can be calculated as

$$\Delta f_{JND} = \left\{ E_L \left\{ \sum_i \left(\delta'_f[CF_i] \right)^2 \right\} - \frac{\left\{ E_L \left(\sum_i \delta'_{Lf}[CF_i] \right) \right\}^2}{E_L \left\{ \sum_i \left(\delta'_L[CF_i] \right)^2 \right\} + API_L} \right\}^{-\frac{1}{2}}, \tag{3.1}$$

where E_L denotes the expected value over the random-level range, $(\delta'_f[CF_i])^2$ and $(\delta'_L[CF_i])^2$ represent the normalized sensitivities for estimating frequency and level, respectively, API_L represents the *a priori* information about level (e.g., the random-level range), and $\delta'_{Lf}[CF_i]$ represents the cross-interaction between changes in level and frequency on the i^{th} AN fiber (see Appendix in Section 3.10, and Chapter 2). Each of the normalized-sensitivities and the cross-interaction term in Eq. 3.1 can be evaluated in terms of the time-varying discharge rates $r_i(t, f, L)$ for all AN fibers. Thus, Eq. 3.1 can be used with any AN model (e.g., analytical or computational), and is applicable to any single-parameter discrimination experiment with one randomized parameter. This analysis is also applicable to any sensory system for which the statistics of the neural responses can be described as a function of the stimulus parameters of interest.

This general form of the relation between optimal performance and the relevant information quantities provides insight about the influence of random level variation on the ability to perform frequency discrimination. Equation 3.1 illustrates that the AN information available to estimate frequency in the random-level experiment is equal to the average (over level) of the information available for fixed-level frequency discrimination, $E_L \left\{ \sum_i (\delta'_f[CF_i])^2 \right\}$, minus

the amount of information that is lost due to the random level variation. The numerator of the second term, $\left\{E_L\left(\sum_i \delta'_{Lf}[CF_i]\right)\right\}^2$, represents the square of the average total correlation between the effect of changes in frequency and changes in level on the AN observations. If changes in level influence the observations in the same way as changes in frequency, then it is more difficult to estimate frequency in the presence of random level variation. The denominator of the second term, $E_L\left\{\sum_i(\delta'_L[CF_i])^2\right\} + API_L$, is a normalization factor that represents the total information available to estimate level. The first term in the denominator is the average information for estimating level from the AN observations, while the second term is the *a priori* information available for estimating level from the limited random-level range. If the tone level can be estimated accurately (relative to the square of the total correlation between changes in level and frequency), then the denominator of the second term is large relative to the numerator and there is little loss of information (i.e., this corresponds to the known-level case averaged over level).

3.5 Computational Methods: Use of Auditory-Nerve Models

All predictions in the present study were made using the computational AN model in the identical manner used in the companion paper (Chapter 2). Briefly, predictions are based on 60 model CF's ranging from 100 Hz to 10 kHz, which are uniformly spaced in location according to a human cochlear map (Greenwood, 1990). The predictions are based on a population of 12,200 total AN fibers, where each of the 60 model responses represent roughly 200 independent, high-spontaneous-rate AN fibers. Tone frequencies were always chosen to be equal to one of the 60 model CF's. Stimulus duration was defined as the duration between half-amplitude points on the stimulus envelope. All rise/fall ramps were generated from raised-cosine functions. The temporal window in the all-information analysis included the model response beginning at stimulus onset and ending 25 ms after stimulus offset, in order to allow for the response delay and the transient onset and offset responses associated with AN fibers over the range of CF's and stimulus parameters used in this study.

Predictions for the rate-place encoding scheme were based on the average discharge rate across the entire temporal-analysis window (i.e., including the extra 25 ms after the nominal offset of the stimulus).

3.6 Computational Results

3.6.1 Random-level frequency discrimination in quiet

Predictions of optimal performance for the random-level frequency discrimination task were calculated using Eq. 3.1 for the same values of frequency, level, and duration used in Chapter 2. A random-level range was uniformly distributed and centered around the nominal level for each condition. For all conditions, there was no effect of random level variation for the rate-place or all-information schemes for either a 6-dB or a 20-dB random-level range (neither shown).[2] This result is consistent with human performance measured by Moore and Glasberg (1989), in which no effect of random level variation was observed with a 6-dB range. Conversely, Emmerich *et al.* (1989) observed a factor of three degradation in human performance when measured with a 20-dB random-level range; however, they showed that much of this effect was due to the confounding role of level-dependent shifts in pitch (Verschuure and van Meeteren, 1975), which are not likely to be produced by the simplified AN model used in the present study. The present results for both rate-place and all-information models thus support the idea that there is no effect of random level variation on frequency discrimination when a small enough random-level range is used to avoid the influence of level-dependent-pitch effects (Moore and Glasberg, 1989).

In order to illustrate how each encoding scheme can discriminate frequency accurately in the presence of random level variation, the distribution of information quantities in Eq. 3.1 across the population of AN fibers is shown in Fig. 3-1. The first row shows the average

[2]The effect of random level variation was evaluated by comparing random-level performance (given by Eq. 3.1) to average (across level) fixed-level performance (determined by the first term in Eq. 3.1, $E_L\left\{\sum_i \left(\delta'_f[CF_i]\right)^2\right\}$). This comparison avoids any potential effect of variation in performance across the levels within the random-level range, which could result in a difference between fixed-level performance and average fixed-level performance that was not truly an effect of random level variation.

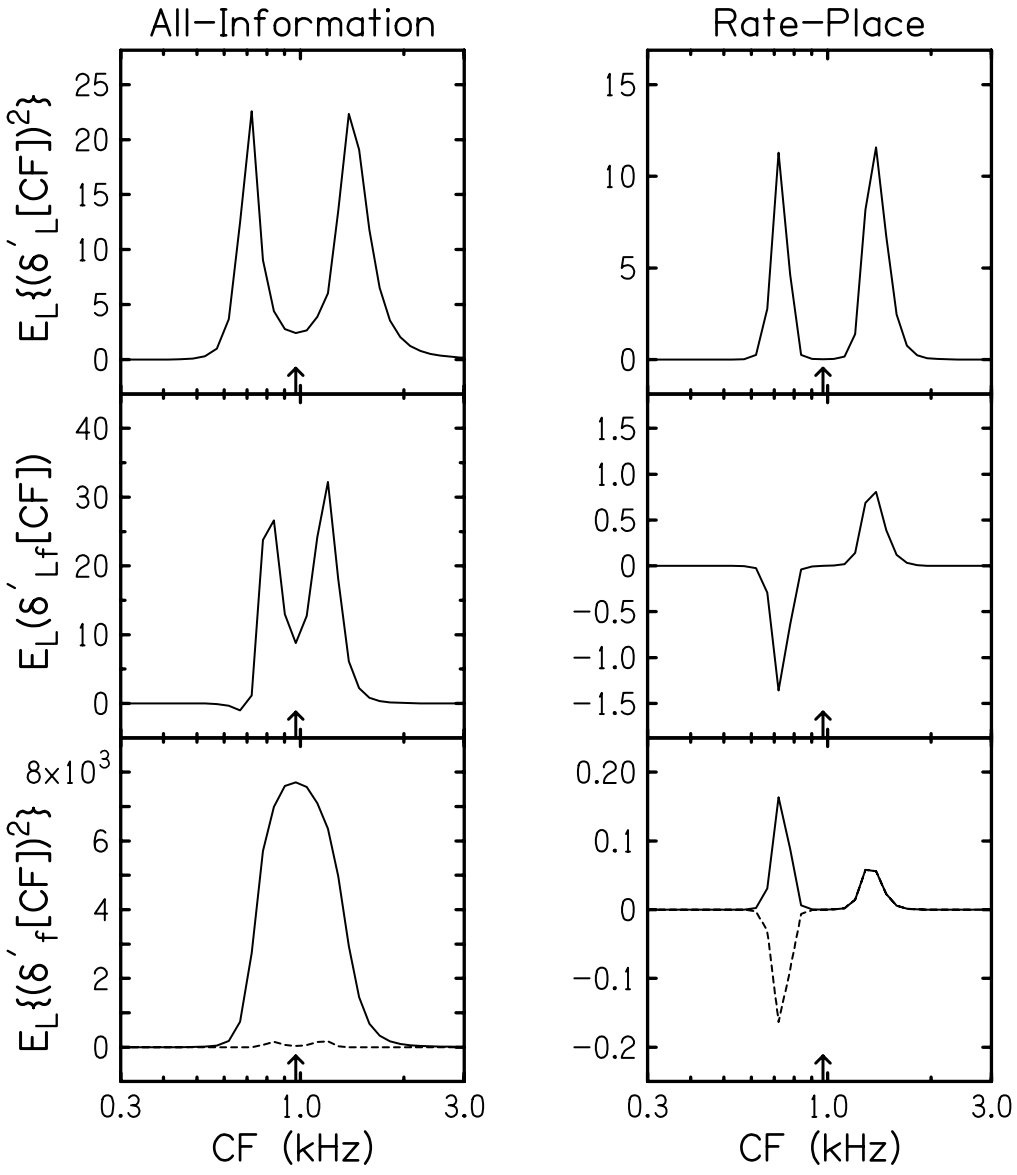

Figure 3-1: Information responsible for optimal performance on a random-level frequency discrimination task for $f = 970$ Hz (indicated by arrows), $L = 40$ dB SPL, $T = 200$ ms (20-ms rise/fall), and a 6-dB random-level range. The left and right columns illustrate information from the all-information and rate-place encoding schemes, respectively. The top panel represents the average level information $E_L\{(\delta'_L[CF])^2\}$ available for each AN model fiber. The middle panel shows the average cross-interaction between level and frequency for each AN fiber, $E_L(\delta'_{Lf}[CF])$. In the bottom panel, the average information available for fixed-level frequency discrimination, $E_L\{(\delta'_f[CF])^2\}$, is shown by the solid line. The dashed line represents the information for estimating frequency based on individual AN fibers that is lost due to the random level variation (second term in Eq. 3.1). The amount of lost information (a positive quantity) is plotted so that those CF's that have negative correlation between changes in level and changes in frequency have a negative value (for illustrative purposes only).

information about level available for each CF, $E_L\{(\delta'_L[CF])^2\}$. The second row shows the average cross-interaction term as a function of CF, $E_L\left(\delta'_{Lf}[CF]\right)$. The average fixed-level information about frequency for each AN fiber, $E_L\left\{(\delta'_f[CF])^2\right\}$, is shown by the solid line in the bottom row. The dashed line in the bottom row represents the average frequency information that is lost due to the random level variation based on estimating frequency from single AN fibers (i.e., the second term in Eq. 3.1 evaluated for single CF's, and with $API_L = 0$). The amount of lost information is plotted so that those CF's that have a negative correlation between changes in level and frequency have a negative value. This is solely for illustrative purposes, as the second term in Eq. 3.1 is always positive because it is the ratio of a squared value and a positive information quantity.

The curves in the bottom panel of Fig. 3-1 illustrate how both encoding schemes overcome the influence of random level variation. In the all-information scheme, each fiber possesses significantly more information for estimating frequency than is lost due to the random level variation (compare solid and dashed lines, Fig. 3-1 bottom left). However, the situation with single-fiber rate-place information is very different. The average amount of rate-place information on a single fiber that is lost due to random level variation is equal to the average amount of information available for estimating frequency with fixed level (compare solid and dashed lines, Fig. 3-1 bottom right). However, when the population response is considered, there is no effect of random level variation for the rate-place scheme due to the opposite polarity of the cross-interaction term above and below the frequency of the tone (middle panel, Fig. 3-1). The lack of an effect for the rate-place population response can be seen in Eq. 3.1 to result from the summation of the cross-interaction term $\delta'_{Lf}[CF]$ over all CF's prior to squaring the total interaction. The rate-place cross-interaction profile is an odd function around the frequency of the tone (Fig. 3-1, middle panel), and thus the positive interaction above the frequency of the tone cancels the negative interaction below the frequency of the tone so that the overall effect of random level variation is negligible.

In summary, single AN fibers in the all-information scheme can perform random-level frequency discrimination equally as well as fixed-level frequency discrimination. In contrast, it is not possible to discriminate frequency in the presence of random level variation based on rate-place information from a single AN fiber. However, a rate-place model that compared

information in CF's above and below the frequency of the tone could make use of the opposite interaction to separate the effect of changes in level from the effect of changes in frequency, and thereby discriminate frequency in the presence of random level variation.

3.6.2 Random-level frequency discrimination in noise

Moore and Glasberg (1989) measured frequency discrimination in four conditions in order to test the duplex theory of frequency coding, i.e., that rate-place information is used at high frequencies while temporal information is used at low frequencies. They reported performance as a function of frequency for four conditions: (1) fixed-level frequency discrimination in quiet, (2) random-level frequency discrimination in quiet, (3) fixed-level frequency discrimination with a high-frequency noise masker that spanned from $1.1f$ to $1.4f$, and (4) random-level frequency discrimination in the presence of the high-frequency noise. Moore and Glasberg suggested that if the high-frequency noise were assumed to mask completely all CF's above the frequency of the tone, then the performance of a rate-place model should be significantly affected by randomizing the level of the tone in the presence of the noise. The random-level frequency discrimination analysis in the present study was used to simulate all four conditions. The first two conditions have been described above, while the conditions that included the high-frequency noise were simulated by discarding the information from all CF's above the frequency of the tone. While this simulated effect of the noise is extreme [e.g., given the effects of suppression (Sachs and Kiang, 1968; Delgutte, 1990b; Ruggero *et al.*, 1992)], this is the assumption that was made by Moore and Glasberg (1989), and has been used often to interpret psychophysical experiments with noise maskers (e.g., Viemeister, 1983). Future studies using more complex AN models and ODT analyses will evaluate the validity of this assumption.

The group mean results from Moore and Glasberg (1989) are shown in Fig. 3-2(a), the all-information predictions in Fig. 3-2(b), and the rate-place predictions in Fig. 3-2(c). Moore and Glasberg reported that there was no statistically significant interaction between condition and frequency, as indicated by the roughly parallel shifts of the curves for the different conditions [Fig. 3-2(a)]. Table 3.1 shows the average factor across frequency by which thresholds for each condition became worse relative to frequency discrimination in quiet [condition

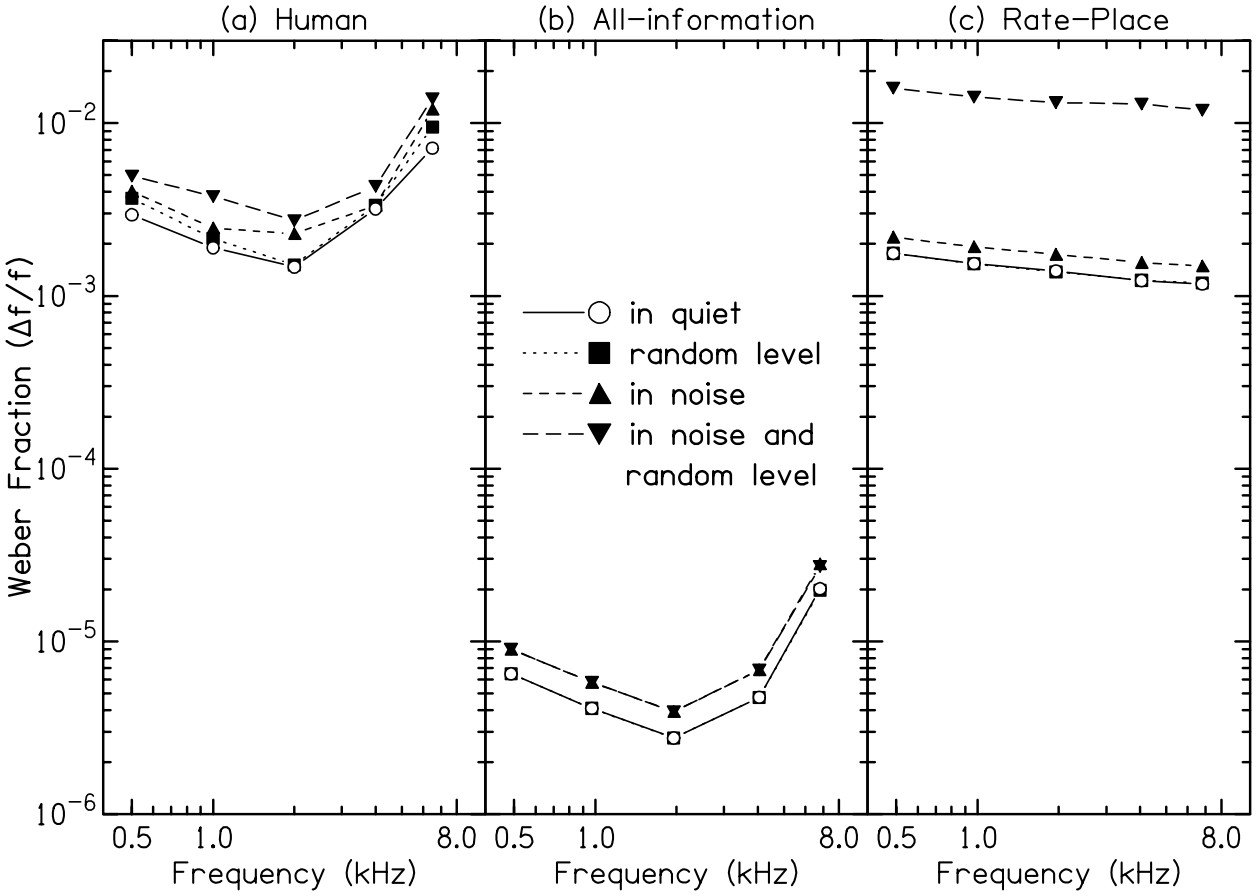

Figure 3-2: Comparison between human and predicted optimal performance for the random-level frequency-discrimination conditions reported by Moore and Glasberg (1989). The Weber fraction $\Delta f/f$ is plotted as a function of frequency in each panel. Human performance is shown in (a); all-information predictions in (b); rate-place predictions in (c). Condition (1) (circles): fixed-level frequency discrimination in quiet; Condition (2) (squares): random-level frequency discrimination in quiet; Condition (3) (upward triangles): fixed-level frequency discrimination in the presence of high-frequency noise; Condition (4) (downward triangles): random-level frequency discrimination in high-frequency noise. Human data are for $L = 70$ dB SPL, T=200 ms (10-ms rise/fall), 6-dB random-level range, and overall noise level of 75 dB SPL (Moore and Glasberg, 1989); Model predictions are for $L = 40$ dB SPL, $T = 200$ ms (20-ms rise/fall), and 6-dB random-level range.

Table 3.1: Average ratio between thresholds in conditions (2)-(4) relative to condition (1) (frequency discrimination in quiet). Conditions are defined in the text and the caption to Fig. 3-2.

	Human	Model	
Condition	Moore and Glasberg (1989)	All-information	Rate-place
random level (2)	1.15[a]	1.00	1.00
in noise (3)	1.37	1.41	1.26
random level in noise (4)	1.65	1.40	9.68

[a] not statistically significant.

(1)] for human and model performance. Moore and Glasberg reported that there was not a statistically significant difference between frequency discrimination in quiet and with a random level [i.e., between conditions (1) and (2)]. Human thresholds became worse by an average factor of 1.37 (statistically significant) for the in-noise condition, while the most significant effect was seen for the random-level in noise condition in which performance became worse by an average factor of 1.65. However, all of these effects in human performance are relatively small, especially compared with the factor-of-five degradation in performance between 2 and 6.5 kHz in all four conditions.

Both all-information [Fig. 3-2(b)] and rate-place predictions [Fig. 3-2(c)] in quiet were unaffected by random-level variation, as described above. The simulated effect of adding the high-frequency noise for the fixed-level task resulted in similar increases in predicted thresholds for both the rate-place and all-information encoding schemes [Figs. 3-2(b),(c)]. The increase is consistent with the removal of roughly half of the available information (see solid curve in bottom panels of Fig. 3-1) and the resultant increase in threshold by a factor of $\sqrt{2}$, and is similar to the size of the observed effect in the human data. The only relative difference between the rate-place and all-information predictions was for the random-level in-noise condition. All-information thresholds were worse than condition (1) by a factor of

1.40; there was no effect of imposing random level variation in the presence of high-frequency noise [i.e., conditions (3) and (4) were identical, Fig. 3-2(b), Table 3.1]. In contrast, the rate-place thresholds increased by a factor of 9.68 compared to frequency discrimination in quiet [Fig. 3-2(c), Table 3.1]. The large predicted increase in rate-place thresholds for the random-level in-noise condition compared to the noise-alone condition is due to the inability of the rate-place model to compare CF's above and below the tone frequency (see Fig. 3-1). This large effect is inconsistent with the small effect observed in human performance at all frequencies.

3.7 Discussion

The companion study (Chapter 2) and the present study each describe new methods to generalize the use of neural models with optimal decision theory (ODT) to relate physiological response properties to psychophysical behavior. The present study describes an extension of the ODT analysis to more complex psychophysical tasks in which one stimulus parameter is randomly varied in order to restrict the cues available to the subject. Optimal frequency-discrimination performance based on either the population of discharge counts (rate-place) or discharge times (all information) was unaffected by random level variation, consistent with human performance. The distributions of frequency and level information across the AN population demonstrated how both rate-place and all-information encoding schemes avoid the effect of random level variation in quiet. Predictions were also made for random-level frequency discrimination in the presence of high-frequency noise, based on the simplified assumption that the noise masker acts to eliminate all information above the frequency of the tone.

When the simple model of the effect of noise masking is used, the predictions for the random-level frequency discrimination in noise experiment (Fig. 3-2) are inconsistent with the use of rate-place information to encode frequency. The predicted effect in rate-place performance of adding random level variation in the presence of high-frequency noise [Fig. 3-2(c)] is much larger than the small effect observed in human performance [Fig. 3-2(a); Moore and Glasberg (1989)]. In fact, there is insufficient rate-place information in the total AN

model population to account for human performance in the random-level-in-noise condition. The cause of the large reduction in information in the random-level-in-noise condition is the inability of the rate-place model to compare CF's above and below the tone frequency. Neither an increase in the total number of AN fibers nor an alteration of the innervation density across CF would reduce the large discrepancy (factor of 9.68) between rate-place JND's for the in-quiet and random-level-in-noise conditions. The implications of other, less easily quantified, alterations to the AN model for the assumed effect of the masking noise are discussed below.

The inconsistency between the effect of random level variation with a noise masker for the rate-place model and human performance occurs for all frequencies. This finding, combined with the finding by Moore and Glasberg (1989) that there was no statistically significant interaction between the effects of condition [(1)-(4), Fig. 3-2] and frequency, suggests that a single encoding scheme is responsible for performance at all frequencies, rather than the duplex theory often invoked to explain frequency encoding (e.g., Wever, 1949; Moore, 1973; Dye and Hafter, 1980; Wakefield and Nelson, 1985; Moore and Glasberg, 1989). As discussed in Chapter 2, rate-place performance in quiet is closer to human performance than all-information performance in terms of absolute performance level. However, the trends in rate-place performance versus frequency are inconsistent with human performance at high frequencies, converse to the duplex theory. This strong discrepancy between the trends in rate-place and human performance is shown in the present study to exist for all four frequency-discrimination conditions described by Moore and Glasberg (1989). In contrast to rate-place, all-information performance matches the trends in human performance across all frequencies and all four conditions. Notably, and contrary to general beliefs (e.g., Moore, 1973, 1989; Dye and Hafter, 1980; Wakefield and Nelson, 1985; Javel and Mott, 1988; Pickles, 1988; Moore and Glasberg, 1989), there is significant temporal information at high frequencies for all four conditions in the present study. Also, the present study shows that all-information performance is unaffected by random level variation in quiet and in the presence of high-frequency noise (Fig. 3-2), consistent with the small effects on human performance.

The computational AN model used in the present study did not include several important aspects of AN responses that could affect the predictions for the masking conditions.

The absence of suppression (Sachs and Kiang, 1968; Delgutte, 1990b; Ruggero *et al.*, 1992) in the present model prohibits the accurate simulation of AN responses to noise stimuli. Suppression could potentially produce effects that contradict the assumption that the high-frequency noise acts to mask completely all AN fibers with CF above the tone frequency. Another potentially significant limitation of the current AN model is the exclusion of high-threshold AN fibers with low and medium spontaneous rates (Liberman, 1978). Low- and medium-spontaneous-rate AN fibers (16% and 23% of the AN population, respectively) have been suggested to contribute to level encoding at higher levels (e.g., Colburn, 1981; Delgutte, 1987; Winter and Palmer, 1991), and therefore should be included in future AN models to quantify the true effect of masking noise on random-level frequency discrimination. In addition, a significant extension of the ODT analysis is necessary in order to accurately evaluate psychophysical tasks with complex random stimuli, such as noise. Despite these limitations, the present study provides interesting new analyses of how the information in a neural population can be used to perform a complex psychophysical task. In addition, the present analysis motivates several important questions about the ability of rate-place information to account for human frequency-discrimination performance, which can be addressed by future extensions of this approach using more complex AN models and ODT analysis.

The primary goal of the present study was to demonstrate a method for relating neural responses to behavior in psychophysical tasks that include random variation of one parameter. This new method was demonstrated for the auditory task of random-level frequency discrimination, but is applicable to any psychophysical discrimination experiment in which one parameter is randomly varied. The analysis presented applies to any sensory system for which there are models that describe the statistical properties of neural responses to the relevant stimuli. Equation 3.1 is valid for any neural model, while some of the analysis in the Appendix in Section 3.10 is specific for neural responses that are well described statistically by a nonstationary Poisson process. Thus, the present study describes a general modeling approach for quantitatively relating physiological responses to behavior in complex psychophysical tasks that are often used to test neural encoding hypotheses.

3.8 Conclusions

1. Optimal decision theory can be used to quantify the effects of both physiological noise and stimulus variation on psychophysical performance in discrimination experiments with random variation in one stimulus parameter.

2. Optimal frequency-discrimination performance based on the population of auditory-nerve (AN) discharge counts (rate-place) or based on discharge times on individual AN fibers is unaffected by random level variation in quiet.

3. There is insufficient rate-place information in the present AN model to account for human performance in a random-level frequency-discrimination task with high-frequency noise.

4. All-information performance with high-frequency noise is unaffected by random level variation, consistent with human performance.

3.9 Acknowledgments

The authors would like to thank Bertrand Delgutte, Christopher Long, Christine Mason, Martin McKinney, Susan Moscynski, Bill Peake, Tom Person, Timothy Wilson, Xuedong Zhang, and Ling Zheng for providing valuable comments on an earlier version of this manuscript. We thank Brian Moore for providing the human data shown in Fig. 3-2(a). Portions of this work were presented at the 21st and 22nd Midwinter Meeting of the Association for Research in Otolaryngology. The simulations in this study were performed on computers provided by the Scientific Computing and Visualization group at Boston University. This work was supported in part by the National Institute of Health, Grants T32DC00038 and R01DC00100, as well as by the Office of Naval Research, Grant Agmt #Z883402.

3.10 Appendix: One-Parameter Discrimination with One Unwanted Random Parameter

This appendix describes an extension of the analysis presented in Chapter 2 to more complex psychophysical tasks. Extension of the ODT analysis to the multiple-parameter case allows for the evaluation of psychophysical tasks in which a single parameter (e.g., overall level) is randomly varied from trial to trial in order to restrict the cues that are available to the subject.

Optimal performance in a random-level frequency discrimination experiment can be evaluated by considering the general case of a one-parameter discrimination experiment in the presence of one unwanted random parameter. In this experiment, the observations (Poisson discharge times on M AN fibers, $\mathcal{T} = \{t_1^i, \cdots, t_{K_i}^i; i = 1, \cdots, M\}$, where t_j^i represents the j^{th} discharge on the i^{th} fiber) are influenced by both the nonrandom, unknown frequency f and the random level L. The Cramér-Rao bound for estimating a vector of random parameters ($N = 2$ in this case) can be used (Cramér, 1951; see van Trees, 1968, pp. 84-85). In general, the information available to an observer to estimate a random parameter is the sum of the *a priori* information based on the known distribution of the parameter, and the information available from the data. In order to treat the frequency parameter as nonrandom, but unknown, the *a priori* information is set to zero. The Cramér-Rao bound provides a lower bound on the variance of any unbiased estimate of f, in the presence of randomized level L, and is given by

$$\begin{aligned} \frac{1}{\sigma_{\hat{f}}^2} \;\leq\; & E_{L,\mathcal{T}}\left\{\left[\frac{\partial}{\partial f}\ln p\left(\mathcal{T}|L;f\right)\right]^2\right\} \\ & - \frac{\left(E_{L,\mathcal{T}}\left\{\left[\frac{\partial}{\partial L}\ln p\left(\mathcal{T}|L;f\right)\right]\left[\frac{\partial}{\partial f}\ln p\left(\mathcal{T}|L;f\right)\right]\right\}\right)^2}{E_{L,\mathcal{T}}\left\{\left[\frac{\partial}{\partial L}\ln p\left(\mathcal{T}|L;f\right)\right]^2\right\} + E_L\left\{\left[\frac{\partial}{\partial L}\ln p\left(L\right)\right]^2\right\}}, \end{aligned} \tag{3.2}$$

where $E_{L,\mathcal{T}}$ indicates the expectation over both the random level L and the random observations $\mathcal{T}$, and $p(L)$ is the probability density used to specify the random-level distribution and determines the *a priori* information for level. Equation 3.2 can be written in a more

useful form using iterated expectations

$$\begin{aligned}\frac{1}{\sigma_{\hat{f}}^2} \;\leq\; & E_L\left(E_{\mathcal{T}}\left\{\left[\frac{\partial}{\partial f}\ln p\left(\mathcal{T}|L;f\right)\right]^2\middle| L\right\}\right) \\ & -\frac{\left[E_L\left(E_{\mathcal{T}}\left\{\left[\frac{\partial}{\partial L}\ln p\left(\mathcal{T}|L;f\right)\right]\left[\frac{\partial}{\partial f}\ln p\left(\mathcal{T}|L;f\right)\right]\middle| L\right\}\right)\right]^2}{E_L\left(E_{\mathcal{T}}\left\{\left[\frac{\partial}{\partial L}\ln p\left(\mathcal{T}|L;f\right)\right]^2\middle| L\right\}\right)+E_L\left\{\left[\frac{\partial}{\partial L}\ln p\left(L\right)\right]^2\right\}}. \end{aligned} \tag{3.3}$$

The probability density of the observed Poisson discharge times on all fibers is the product of the densities for individual fibers, assuming each fiber is conditionally independent given L (see Parzen, 1962; Snyder and Miller, 1991; Rieke *et al.*, 1997; Chapter 2. The conditional expectations with respect to $\mathcal{T}$ are of a form which has been previously evaluated for Poisson observations (Siebert, 1970; Chapter 2).

The just-noticeable difference Δf_{JND} of the ideal observer is equal to the minimum standard deviation of any estimator based on the observations, where threshold is defined as 75% correct in a two-interval, two-alternative forced-choice task. This threshold definition corresponds to $d' = 1$, where $d' = \Delta f_{JND}/\sigma_{\hat{f}}$. The JND based on the population of AN fibers is given by

$$\begin{aligned}\frac{1}{(\Delta f_{JND})^2} \;=\; & E_L\left\{\sum_i \int_0^T \frac{1}{r_i(t,f,L)}\left[\frac{\partial r_i(t,f,L)}{\partial f}\right]^2 dt\right\} \\ & -\left[\left(E_L\left\{\sum_i \int_0^T \frac{1}{r_i(t,f,L)}\left[\frac{\partial r_i(t,f,L)}{\partial L}\frac{\partial r_i(t,f,L)}{\partial f}\right]dt\right\}\right)^2\Bigg/ \right. \\ & \left.\left(E_L\left\{\sum_i \int_0^T \frac{1}{r_i(t,f,L)}\left[\frac{\partial r_i(t,f,L)}{\partial L}\right]^2 dt\right\}+E_L\left\{\left[\frac{\partial \ln p_L(L)}{\partial L}\right]^2\right\}\right)\right]. \end{aligned} \tag{3.4}$$

Equation 3.4 describes optimal performance for discriminating frequency in the presence of random level variation based on the AN-population response in terms of the time-varying discharge rates $r_i(t,f,L)$. Following the notation used in Chapter 2,

$$(\delta_f'[CF_i])^2 \;\triangleq\; \int_0^T \frac{1}{r_i(t,f,L)}\left[\frac{\partial r_i(t,f,L)}{\partial f}\right]^2 dt. \tag{3.5}$$

and

$$(\delta'_L[CF_i])^2 \triangleq \int_0^T \frac{1}{r_i(t,f,L)} \left[\frac{\partial r_i(t,f,L)}{\partial L}\right]^2 dt, \tag{3.6}$$

The quantities $(\delta'_f[CF_i])^2$ and $(\delta'_L[CF_i])^2$ represent the information available on the i^{th} fiber for estimating frequency f and level L, respectively, and are shown in Chapter 2 to represent normalized sensitivities, defined as the sensitivity d' per unit f or L (also see Durlach and Braida, 1969; Braida and Durlach, 1988). Similarly, the cross-interaction term is defined as

$$\delta'_{Lf}[CF_i] \triangleq \int_0^T \frac{1}{r_i(t,f,L)} \left[\frac{\partial r_i(t,f,L)}{\partial L}\frac{\partial r_i(t,f,L)}{\partial f}\right] dt, \tag{3.7}$$

and represents the correlation between changes in level and changes in frequency on the i^{th} fiber. Eq. 3.4 can thus be written as

$$\Delta f_{JND} = \left\{ E_L\left\{\sum_i \left(\delta'_f[CF_i]\right)^2\right\} - \frac{\left\{E_L\left(\sum_i \delta'_{Lf}[CF_i]\right)\right\}^2}{E_L\left\{\sum_i (\delta'_L[CF_i])^2\right\} + API_L}\right\}^{-\frac{1}{2}}, \tag{3.8}$$

where API_L represents the *a priori* information available about level (e.g., the range of levels used in the random variation of level).

A Gaussian distribution was used to calculate API_L due to the analytical difficulty for a uniform distribution that results from the undefined derivative with respect to L for levels at the edges of a uniform distribution. The *a priori* information for level was calculated to be $API_L = 2\pi/R^2$ for a Gaussian distribution with a variance of $R^2/2\pi$, where R represents the random-level range in dB. A Gaussian distribution with variance $R^2/2\pi$ has the same equivalent-rectangular-width as a uniform distribution with random-level range R.

Chapter 4

Monaural, cross-frequency coincidence detection as a mechanism for decoding perceptual cues provided by the cochlear amplifier

4.1 Abstract

The perceptual significance of the cochlear amplifier was evaluated by predicting optimal level-discrimination performance based on auditory-nerve (AN) discharge counts and on discharge times. An analytical AN model was used that included compressive-magnitude and level-dependent phase responses associated with the cochlear amplifier, and high-, medium-, and low-spontaneous-rate (SR) AN-fiber populations. Strongly compressive basilar membrane (BM) responses at high frequencies interact with the high thresholds associated with low-SR AN fibers to produce large dynamic ranges, while nonlinear BM phase responses are encoded in AN fibers at levels above 30 dB SPL at low frequencies. Broadened tuning produces an increase in the correlation between responses of adjacent AN fibers. Thus, performance was also predicted based on a monaural, across-frequency coincidence detection mechanism that can extract nonlinear phase cues. Nonlinear phase changes are required to encode levels up to 120 dB SPL at low frequencies based on AN fibers with a narrow range of characteristic frequencies (CF). In contrast, average-rate information from LSR fibers is required for robust level encoding within a narrow CF range at high frequencies. Coincidence performance matched human performance at all levels and at both low and high frequencies. The "near-miss" to Weber's Law at low frequencies is accounted for by nonlinear phase cues, while the high-frequency "mid-level bump" is produced by compressive magnitude responses. The current predictions suggest that the cochlear amplifier is primarily beneficial to normal-

hearing listeners for encoding sound level within narrow frequency regions, and has only a small influence on performance in simple listening tasks, e.g., in quiet.

4.2 Introduction

The cochlear amplifier is the name used to describe an active mechanism within the cochlea that is thought to provide amplification of low-level sounds in the cochlea. The active mechanism is believed to involve motile outer-hair cells, which transduce mechanical motion on the basilar membrane (BM) into intracellular electrical potentials that are in turn converted into mechanical energy that enhances BM motion (Yates, 1995; Mountain, 1996). While the mechanism of amplification is not well understood, several physiological response properties associated with the cochlear amplifier are clear. The most significant of these is that the active mechanism is vulnerable to cochlear damage, and has been shown to be absent in many common forms of sensorineural hearing loss (Patuzzi *et al.*, 1989). This finding raises the question of how the cochlear amplifier benefits normal hearing listeners, especially in complex listening environments such as the understanding of speech in noise, for which hearing-impaired listeners have much difficulty (Moore, 1995b). The present study evaluates quantitatively some of the benefits of the cochlear amplifier for extending the dynamic range of the auditory system. The absence of the cochlear amplifier in damaged cochleae is likely responsible for the common report of loudness recruitment by listeners with sensorineural hearing loss, and the associated reduction in dynamic range.

It is still not well understood how the auditory system overcomes the dynamic-range problem (for reviews see Evans, 1981; Viemeister, 1988a,b), i.e., the discrepancy between the large dynamic range of human hearing [over 120 dB (Viemeister and Bacon, 1988)], and the limited dynamic range of most auditory-nerve (AN) fibers [less than 30 dB (May and Sachs, 1992)]. The present study focuses on quantifying how the cochlear amplifier helps to extend the dynamic range of the auditory system in terms of the AN information available for encoding changes in level. A psychophysical experiment in which the dynamic-range problem is most clearly evident was examined in the present modeling study: level discrimination of high-level, narrowband signals in conditions for which information is restricted to frequency

regions near the frequency of the signal (e.g., Carlyon and Moore, 1984; Viemeister, 1983). An influential experiment for level-encoding hypotheses was performed by Viemeister (1983), who found that Weber's Law (i.e., constant just-noticeable-difference in level as a function of level), was achieved for high-frequency, narrow-band noise in the presence of band-reject noise. This experiment was designed to prevent the spread of excitation by using a band-reject noise masker, and to prevent the use of temporal information by using a high-frequency signal. Viemeister's finding has been taken as evidence that Weber's Law must hold in narrow frequency regions and must rely on the use of average rate information.

The population of low-spontaneous-rate (LSR), high-threshold AN fibers (Liberman, 1978) have been implicated in the encoding of high sound levels based on average-rate information in narrow frequency regions (Colburn, 1981; Delgutte, 1987; Viemeister, 1988a,b; Winslow and Sachs, 1988; Winter and Palmer, 1991). However, when models have been used to quantify the total information available in a restricted CF-region with physiologically realistic distributions of the spontaneous-rate (SR) groups, performance has been predicted to degrade as level increases above 40 dB SPL (Colburn, 1981; Delgutte, 1987), inconsistent with Weber's law and with trends in human performance (Viemeister, 1983; Carlyon and Moore, 1984). Delgutte (1987) demonstrated that Weber's Law could be achieved in single CF-channels by processing high-threshold, LSR AN fibers more efficiently than low-threshold, high-spontaneous-rate (HSR) fibers. He showed that the "near-miss" to Weber's Law (i.e., a slight improvement in performance as level increases), which is observed in human performance for tones in quiet (e.g., McGill and Goldberg, 1968; Rabinowitz *et al.*, 1976; Jesteadt *et al.*, 1977; Florentine *et al.*, 1987), could be obtained by combining information across CF channels that individually achieved Weber's Law. This idea is similar to the assumption made by Florentine and Buus (1981) in their excitation-pattern model. However, there is no strong evidence for the type of preferential processing of LSR fibers used by Delgutte (1987). In addition, the wide dynamic range of LSR fibers depends on the compressive BM responses (Sachs and Abbas, 1974), which appears to be much less compressive at low frequencies (Cooper and Rhode, 1997; Hicks and Bacon, 1999a). Reduced compression at low frequencies is consistent with the presence of non-saturating ("straight") rate-level curves only at high frequencies in guinea pig (Winter and Palmer, 1991). Thus, it

is desirable to investigate other potential sources of information that could produce Weber's Law in narrow frequency regions.

The cochlear amplifier is potentially relevant for the encoding of sound level in narrow frequency regions because the associated nonlinear properties have typically been reported to influence primarily characteristic frequencies (CF's) near the frequency of a tone. Specifically, the nonlinear near-CF response properties include both compressive magnitude responses (Rhode, 1971; Ruggero *et al.*, 1997), as well as level-dependent phase shifts (BM: Ruggero *et al.*, 1997; IHC: Cheatham and Dallos, 1998; AN: Anderson *et al.*, 1971). In evaluating the potential of the cochlear amplifier to extend the dynamic range of the auditory system, it is important to consider several limiting transformations that occur between the BM and the AN. These include: (1) saturating rate-level curves (Kiang *et al.*, 1965; Sachs and Abbas, 1974), which act to limit the encoding of nonlinear gain, (2) roll-off of phase-locking at high frequencies (Johnson, 1980; Joris *et al.*, 1994), which limits nonlinear phase encoding, and (3) randomness of AN responses (Young and Barta, 1986; Miller *et al.*, 1987; Winter and Palmer, 1991; Delgutte, 1996), which limits overall psychophysical performance. Thus, it is important to consider the encoding of information in the AN, not just based on the compression in BM responses, when evaluating the significance of the cochlear amplifier.

The nonlinear phase changes associated with the cochlear amplifier, which have not been studied in as much detail as the compressive magnitude responses (Sachs and Abbas, 1974; Winter and Palmer, 1991; Moore, 1995b; Moore and Oxenham, 1998), are a focus of the present study. These phase cues continue to encode changes in stimulus level at high levels, despite the saturation of average rate above 40 dB SPL for the majority of AN fibers (Sachs and Abbas, 1974; May and Sachs, 1992), and thus may provide a partial solution to the dynamic-range problem. However, in order for the central nervous system to make use of these changes in phase, a phase reference is required. Therefore, it is important to consider physiologically realistic mechanisms that could make use of the information provided by nonlinear phase shifts. Due to the differences in phase changes in adjacent CF's, a relative phase reference can be obtained by looking across neighboring CF's. Carney (1994) demonstrated that nonlinear phase shifts on single AN fibers result in systematic changes in the temporal discharge patterns across CF (i.e., spatio-temporal patterns that vary with

level over a wide dynamic range), and hypothesized that changes in spatio-temporal patterns may be important for the encoding of sound level. Any two AN fibers with different CF's have a relative phase difference that varies with level, independent of the absolute phase of the stimulus. Thus, a mechanism that compared the relative timing of two AN fibers would be sensitive to changes in level, without requiring an absolute phase reference.

The present study considers monaural, cross-frequency coincidence detection as a mechanism for decoding the nonlinear phase cues provided by the cochlear amplifier. Coincidence detection is a physiologically realistic mechanism, because any neuron with multiple subthreshold inputs acts as a coincidence detector (Carney, 1994; Joris *et al.*, 1998). Carney (1990) has shown that several response types in the antero-ventral cochlear nucleus (AVCN) were sensitive to changes in relative phase across their inputs, consistent with a coincidence detection mechanism. In addition, there is much evidence for coincidence detection in the binaural system (Colburn, 1996). Neurons in the medial superior olive (MSO) and inferior colliculus (IC) have responses that are consistent with coincidence detection between inputs from each ear as a mechanism for decoding interaural time differences that are known to be important for sound localization (Yin and Chan, 1990; Goldberg and Brown, 1969; Rose *et al.*, 1966; Yin *et al.*, 1987; Joris *et al.*, 1998).

In the present study, methods from statistical decision theory (SDT) were combined with an analytical nonlinear AN model and a simple coincidence-counting model. The relative contributions of nonlinear magnitude and nonlinear phase responses to level encoding are compared by using four versions of the analytical AN model, which are created by turning on and off the nonlinear gain and nonlinear phase responses in all possible combinations. The nonlinear gain provides large dynamic ranges for high-threshold, LSR fibers, which support consistent level encoding based on average-rate information in narrow CF regions for high-frequency tones, but not for low-frequency tones. The nonlinear phase cues support Weber's Law for low-frequency, high level tones based on temporal information in low-threshold, HSR fibers near the frequency of the tone, unlike rate and synchrony information, as was shown by Carney *et al.* (1999). The performance of a population of monaural, cross-frequency coincidence detectors is shown to be consistent with human performance for both low- and high-frequency tones in quiet. This physiologically realistic mechanism is extremely

general in that it can decode AN information from all SR groups in the form of average-rate, synchrony, and nonlinear phase cues, and thus is robust across a wide range of sound levels.

4.3 Methods

4.3.1 Auditory-nerve model

The nonlinear AN model used in the present study is an extension of simple linear analytical AN models used by Siebert (1965, 1968, 1970) and by Colburn (1969, 1973, 1977a,b, 1981). The linear AN model was modified to include the main properties of the cochlear nonlinearities associated with the active process, including: (1) nonlinear compressive responses from 30 to 120 dB SPL, (2) compressive nonlinearity restricted to "near-CF" regions, (3) compression that varies with CF, (4) maximum phase shifts of $\pm\pi/2$ both above and below CF, (5) no phase shifts at CF, and (6) dynamic range of each SR group that depends on the compressive magnitude response. The response properties of the model are described in the text below, while the assumptions and equations used to specify the model are described in the Appendix in Section 4.8. This analytical nonlinear AN model was purposefully kept as simple as possible in order to provide greater intuition and to allow the contribution of each nonlinear property to be investigated separately. The analyses presented below are not limited to this AN model, however, and could be pursued in the future with more complex computational nonlinear models.

The statistics of the AN discharges are modeled by a non-stationary Poisson process with rate function $r(t)$. The phase-locked response of the i^{th} AN fiber (with characteristic frequency CF_i) to a tone burst of level L, frequency f_0, duration T, and phase ϕ, is described by a time-varying rate function similar to that used by Colburn (1981), i.e.,

$$r_i(t; L, f_0, T, \phi) = \frac{\bar{r}[L_{eff}(L, f_0, CF_i)]}{I_0\{g[L_{eff}(L, f_0, CF_i), f_0]\}} \times \exp\{g[L_{eff}(L, f_0, CF_i), f_0] \cos[2\pi f_0 t + \theta(L, f_0, CF_i) + \phi]\}, \quad (4.1)$$

where $I_0\{g\}$ is the zeroth-order Bessel function (equal to the time-average of the exponential term). Both the average rate $\bar{r}[L_{eff}]$ and synchrony $g[L_{eff}, f_0]$ are affected by saturating nonlinearities, where the effective level L_{eff} is determined by the nonlinear BM filtering properties and by the level and frequency of the tone. The term $g[L_{eff}, f_0]$ also depends on f_0 such that the strength of phase locking decreases at high frequencies. The nonlinear phase response $\theta(L, f_0, CF_i)$ depends on the level and frequency of the tone as well as on the CF of the AN fiber (see Anderson *et al.*, 1971; Ruggero *et al.*, 1997), and is described similarly to Carney *et al.* (1999). The stimulus is assumed to have random phase ϕ in order to avoid the assumption that the phase of the tone is known.

Many basic response properties of the AN model are illustrated in Fig. 4-1. Panels (a)-(c) show the implementation of the nonlinear magnitude responses, which are consistent with physiological data from Ruggero *et al.* (1997). Normalized BM response versus frequency for a 10-kHz CF is shown in Fig. 4-1(a), for a range of levels. The filters are triangular at low levels, consistent with the linear AN models used by Siebert (1965, 1968, 1970) and Colburn (1969, 1973, 1977a,b, 1981) to fit AN tuning curves in cat. The maximum gain of the cochlear amplifier (i.e., the gain relative to high levels, or equivalently the amount of compression relative to low levels) occurs at CF, and is equal to 60 dB for this CF. The nonlinear gain decreases as tone frequency moves away from CF, and the response is linear well away from CF (roughly more than $\pm 1/2$ octaves). Fig. 4-1(b) shows BM output at CF as a function of level for the 10-kHz place. The solid curve represents the nonlinear BM response, while the dashed line represents the linear version of the model. The compressive region extends from 30 to 120 dB SPL, and the model responses are linear below this range. Fig. 4-1(c) shows the cochlear-amplifier gain at CF as a function of CF. The maximum gain decreases as CF decreases, with 60 dB of gain for frequencies above 8 kHz, 20 dB of gain for frequencies below 500 Hz, and a smooth transition for CF's in between. This pattern of nonlinear gain across CF is consistent with both physiological and psychophysical evidence, although the exact amount of gain at low frequencies is still unclear. The majority of BM data has been obtained at high CF's, and indicate a maximum gain of roughly 50-60 dB (Ruggero *et al.*, 1997; Nuttall and Dolan, 1996). The BM data at low CF's is less abundant, but indicates reduced nonlinearity at low CF's (e.g., Cooper and Rhode, 1997). Hicks and

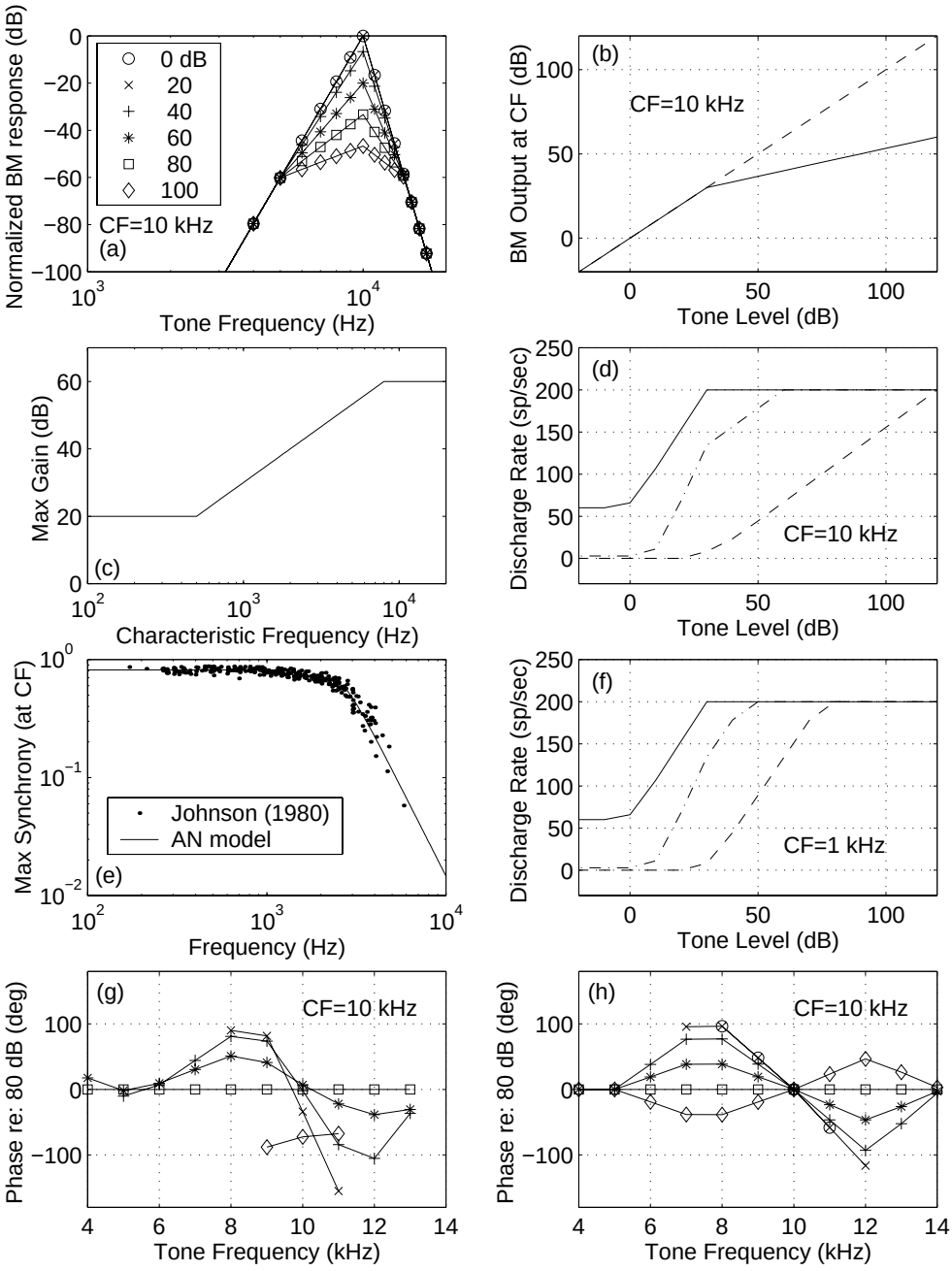

Figure 4-1: Nonlinear AN model response properties. (a) Normalized basilar-membrane (BM) response for a 10-kHz place as a function of frequency for levels ranging from 0-100 dB SPL. (b) BM output at CF as a function of level for a 10-kHz place (solid: nonlinear; dashed: linear). (c) Nonlinear gain at CF as a function of CF. (d) Rate-level curves for a 10-kHz tone at CF for three SR groups (HSR: solid, MSR: dash-dot, LSR: dashed). (e) Maximum synchrony versus frequency. Model responses are compared to data measured in cat (Johnson, 1980). (f) Rate-level curves for a 1-kHz tone. (g) BM phase-response areas (phase relative to 80 dB SPL) from chinchilla for a 10-kHz CF (data from Ruggero *et al.*, 1997). (h) AN-model phase-response areas for a 10-kHz CF [(g,h): same symbols as in (a)].

Bacon (1999a) presented psychophysical evidence that cochlear nonlinearity is reduced at low frequencies, and is characterized by a gradual, rather than steep, transition as CF decreases.

Figure 4-1(d) and (f) illustrates how average rate varies with level for the three spontaneous-rate (SR) groups of AN fibers at high and low frequencies, respectively. The AN model represents all fibers within each SR population with a fixed threshold and SR. Based on data from Liberman (1978), SR values of 60, 3, and 0.1 sp/sec, thresholds of 0, 10, and 30 dB SPL, and population percentages of 61%, 23%, and 16%, were used for the HSR, MSR, and LSR populations, respectively. A saturated rate of 200 sp/sec was used for all three SR groups. Note that the rate-level curves at low frequencies are either "saturating" or "sloping saturating," while at high frequencies there is a third class of "straight" rate-level curves. This pattern is consistent with rate-level curves in guinea pig described by Winter and Palmer (1991), who found no "straight" rate-level curves below 1.5 kHz, and results from the decrease in cochlear-amplifier gain as frequency decreases. Figure 4-1(e) compares the rolloff in phase-locking versus frequency in the model to that observed in cat (Johnson, 1980).

Figure 4-1(g) illustrates nonlinear physiological BM phase responses for a 10-kHz CF (Ruggero *et al.*, 1997), while Fig. 4-1(h) shows the AN-model phase responses. Phase is plotted relative to the phase at a high level (80 dB SPL) in both panels, where each curve represents a different tone level. Thus, any difference from zero represents a phase response that changes with level. The major properties of this nonlinear response, observed for BM responses at high frequencies (Geisler and Rhode, 1982; Ruggero *et al.*, 1997), and IHC (Cheatham and Dallos, 1998) and AN (Anderson *et al.*, 1971) responses at low frequencies, are that (1) phase lags as level increases for $f < CF$, (2) phase leads as level increases for $f > CF$, (3) there are no phase changes at CF, (4) the nonlinear-phase region is the same width in frequency as the nonlinear region for the magnitude response, and (5) the maximum phase shifts observed are roughly $\pm\pi/2$ and occur about half-way into the nonlinear region. The nonlinear phase responses are consistent with broadened tuning as level increases and the associated changes in the phase-versus-frequency slope (i.e., the slope becomes more shallow as filters broaden).

All predictions in the present study were made with 120 distinct model CF's spaced

logarithmically from 300 to 20,000 Hz. It was assumed that the total AN population consists of 30,000 total AN fibers with CF's ranging from 20 to 20,000 Hz.

4.3.2 Monaural, across-frequency coincidence counting model

The present study uses a simple coincidence-counting model that was described by Colburn (1969, 1973, 1977a,b) in his studies of binaural phenomena (Fig. 4-2). A coincidence detector receives two AN-fiber inputs, and is assumed to discharge only when the two input fibers discharge within a narrow temporal window. The coincidence counter outputs the number of coincident discharges within the duration of the stimulus. The present use of the coincidence-counting model differs from that of Colburn only in the source of the two AN inputs. In the binaural model, each AN fiber was from a different ear and had the same CF. In the present study, the two AN inputs are from the same ear, but can have different CF's. In both studies, performance was assumed to depend only on the number of coincidences between two AN fibers (i.e., the timing of the coincidences was ignored). The number of coincidences between two AN fibers with discharge times $\mathcal{T}^i = \{t_1^i, \cdots, t_{K_i}^i\}$ and $\mathcal{T}^j = \{t_1^j, \cdots, t_{K_j}^j\}$ is given by

$$\mathcal{C}_{ij}\{\mathcal{T}^i, \mathcal{T}^j\} = \sum_{l=1}^{K_i} \sum_{m=1}^{K_j} f(t_l^i - t_m^j), \tag{4.2}$$

where t_l^i is the l^{th} discharge out of K_i on the i^{th} AN fiber, and $f(\cdot)$ is a rectangular coincidence window with 10-μs width and unity height (see Colburn, 1969).

4.3.3 Evaluation of psychophysical performance limits

Psychophysical performance is limited by the random nature of AN responses (i.e., that different responses are observed for two identical stimulus presentations). Psychophysical performance limits for discrimination experiments have been evaluated with methods from statistical decision theory (SDT) by using a nonstationary Poisson process with a time-varying discharge rate $r(t)$ to describe the stochastic nature of AN discharges (e.g., Siebert, 1968; 1970; Colburn, 1969; 1973; Chapters 2 and 3). Chapter 2 described a general com-

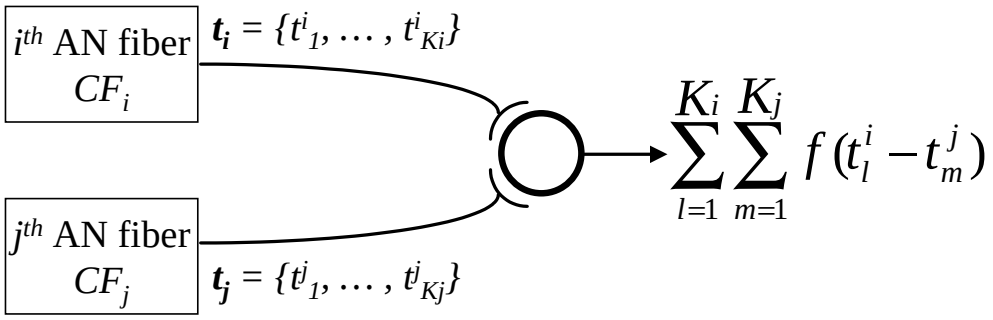

Figure 4-2: Simple model of a monaural, across-frequency coincidence counter. The coincidence detector receives two AN inputs with characteristic frequencies CF_i and CF_j, and discharge times $\mathcal{T}^i = t^i_1, \cdots, t^i_{K_i}$ and $\mathcal{T}^j = t^j_1, \cdots, t^j_{K_j}$, where t^i_l is the l^{th} discharge on the i^{th} AN fiber. The coincidence detector discharges if both inputs discharge within the narrow coincidence window $f(x)$. The output of the coincidence counter is the number of coincidences that occur within the duration of the stimulus.

putational method for evaluating psychophysical performance limits using any AN model that describes $r(t)$ for the stimulus conditions of interest. The analytical AN model used in the present study was implemented computationally, and psychophysical performance limits were evaluated based on two hypotheses for the type of information used to perform the task, *all-information* and *rate-place*. The all-information model assumes that the observations used by the optimal processor consist of the complete set of discharge times across the entire AN population, $\mathcal{T} = \{t^i_l\}_{i=1,\cdots,M;l=1,\cdots,K_i}$, where M =30,000 total AN fibers, and K_i is the number of discharges on the i^{th} AN fiber. Thus, the all-information model assumes that the processor makes optimal use of all available information from the AN (e.g., average-rate, synchrony, and phase information). The rate-place model assumes that the optimal processor only uses the number of discharges observed on each AN fiber, $\{K_i\}$ (i.e., the processor is assumed to not make use of any temporal information).

The contribution of each AN fiber to a level discrimination task can be quantified by calculating the normalized sensitivity δ' to changes in stimulus level [δ' is defined as the sensitivity d' per dB (see Durlach and Braida, 1969; Braida and Durlach, 1988); Chapter 2]. The square of the normalized sensitivity of the i^{th} AN fiber in the all-information model is

given by

$$(\delta'[CF_i])^2 = \int_0^T \frac{1}{r_i(t)} \left[\frac{\partial r_i(t)}{\partial L}\right]^2 dt, \tag{4.3}$$

where T is the duration of the stimulus (Siebert, 1970; Chapter 2). The total normalized sensitivity based on the population of AN fibers is the sum of the individual normalized sensitivities, $(\delta')^2 = \sum_i (\delta'[CF_i])^2$, based on the assumptions of independent AN fibers for deterministic stimuli (Johnson and Kiang, 1976), and an optimal combination across AN fibers. The just-noticeable-difference (JND) in level is given by

$$\Delta L = \frac{1}{\sqrt{(\delta')^2}}. \tag{4.4}$$

Equation 4.3 describes the normalized sensitivity based on the all-information model; the normalized sensitivity based on rate-place information can be calculated with Eq. 4.3 by assuming that $r(t)$ is constant across the duration of the stimulus and equal to the average-discharge rate $\bar{r}$. The contribution of temporal information in AN discharges can be discerned from a comparison between performance based on the all-information and rate-place models. Comparisons between predicted and human performance in terms of trends and absolute performance levels provide insight into the potential significance of average-rate and temporal information for the encoding of sound level in the AN.

Performance based on the outputs of a set of coincidence counters is calculated and compared to rate-place, all-information, and human performance. While the all-information predictions represent the optimal performance of any decision device based on the AN discharge times, the coincidence mechanism represents a specific physiologically realistic processor that uses the discharge times sub-optimally. The number of discharges from a single coincidence detector, $\mathcal{C}_{ij}\{\mathcal{T}^i, \mathcal{T}^j\}$, is a simple function (Eq. 4.2) of the two sets of Poisson AN discharge times, $\mathcal{T}^i$ and $\mathcal{T}^j$, and thus the statistics of the coincidence counts can be described (Appendix in Section 4.9). The performance of a single coincidence counter $\mathcal{C}_{ij}$ for

level discrimination can be evaluated by calculating the sensitivity index

$$Q_{ij} = \frac{\left(E[\mathcal{C}_{ij}|L + \Delta L] - E[\mathcal{C}_{ij}|L]\right)^2}{Var[\mathcal{C}_{ij}|L]}, \tag{4.5}$$

where the just-noticeable difference for this coincidence counter, $\Delta L_{ij,JND}$ corresponds to $Q_{ij} = 1$. The sensitivity index Q_{ij} corresponds to the commonly used sensitivity index $(d')^2$ if $\mathcal{C}_{ij}$ has a Gaussian distribution. It can be assumed that $E[\mathcal{C}_{ij}|L]$ varies linearly over the incremental level range from L to $L + \Delta L_{JND}$. Thus, the normalized sensitivity squared for a single coincidence counter, defined as $(\delta'_{ij})^2 \triangleq Q_{ij}/(\Delta L)^2$, can be approximated as

$$(\delta'_{ij})^2 \simeq \frac{\left(\frac{\partial}{\partial L} E[\mathcal{C}_{ij}|L]\right)^2}{Var[\mathcal{C}_{ij}|L]}. \tag{4.6}$$

The expectation and variance in Eq. 4.6 can be evaluated in terms of the stimulus parameters and the two AN CF's (Appendix in Section 4.9). The partial derivative with respect to level in Eq. 4.6 can be approximated computationally by the difference between the expected value at two slightly different levels divided by the incremental level difference (Chapter 2). The total normalized-sensitivity-squared for a population of coincidence counters is given by the sum of the individual normalized-sensitivities-squared, if the individual coincidence counts are assumed to be independent and combined optimally. In order to satisfy the independence assumption, it is assumed throughout the present study that no AN fiber innervates more than one coincidence counter. The JND based on coincidence counts is calculated from Eq. 4.4.

4.4 Results

4.4.1 Distribution of rate, synchrony, and phase information across CF

In order to illustrate the potential benefit of nonlinear phase cues to the encoding of sound level, we first focus on level discrimination at high levels, where the dynamic-range problem

is most prominent. Figure 4-3 illustrates the distribution and relative contributions of rate, synchrony, and phase cues across the AN population of high-spontaneous-rate (HSR) fibers for level discrimination of a 1-kHz, 100-dB SPL tone. The rate responses (i.e., average discharge rate as a function of CF) are shown for two tones of slightly different level in panel (b). The louder tone produces a wider activation pattern; however, the discharge rate for a wide range of CF's near the tone frequency is the same for both tones due to saturation. The distribution of rate information [i.e., normalized sensitivity squared, $(\delta'[CF])^2$] shown in panel (c) illustrates which AN fibers across the population contribute information for level discrimination. Near the frequency of the tone, there is no rate information due to the saturation of rate-level curves for the HSR fibers. The only rate information available for HSR fibers is at frequencies well away from the tone frequency (Siebert, 1965, 1968).

The situation of primary interest in this study is when information is restricted to AN fibers with CF's near the frequency of the tone. This situation is thought to occur in experiments that use a notched noise masker to limit the spread of excitation (e.g., Viemeister, 1983; Carlyon and Moore, 1984). A narrow frequency region that will be considered in the current study is indicated by the vertical dotted lines. This region represents seven model CF's, three above and three below the CF equal to the tone frequency. The narrow frequency region for the 1-kHz tone is 896-1107 Hz, and is similar to the notch width used by Carlyon and Moore (1984), which extended $\pm 10\%$ from the tone frequency. Humans are typically able to perform level discrimination well in the presence of a notched noise (e.g., Viemeister, 1983; Carlyon and Moore, 1984; Schneider and Parker, 1987); however, Fig. 4-3(c) shows that there is no average-rate information in HSR AN fibers within the narrow-CF region.

AN phase-locking has a different level dependence than average rate, and thus it is important to examine the distribution of synchrony information in addition to rate information. Fig. 4-3(d) shows the 1-kHz synchrony coefficient (or vector strength, which ranges from 0 to 1, see Johnson, 1980) for each tone plotted as a function of CF. Synchrony-level curves have thresholds that are roughly 20 dB below rate thresholds, and typically saturate just above rate threshold (Johnson, 1980). The synchrony-response regions are thus slightly wider than the rate response regions, but are also saturated over a wide range of CF's near the tone frequency. The distribution of information available from both rate and synchrony informa-

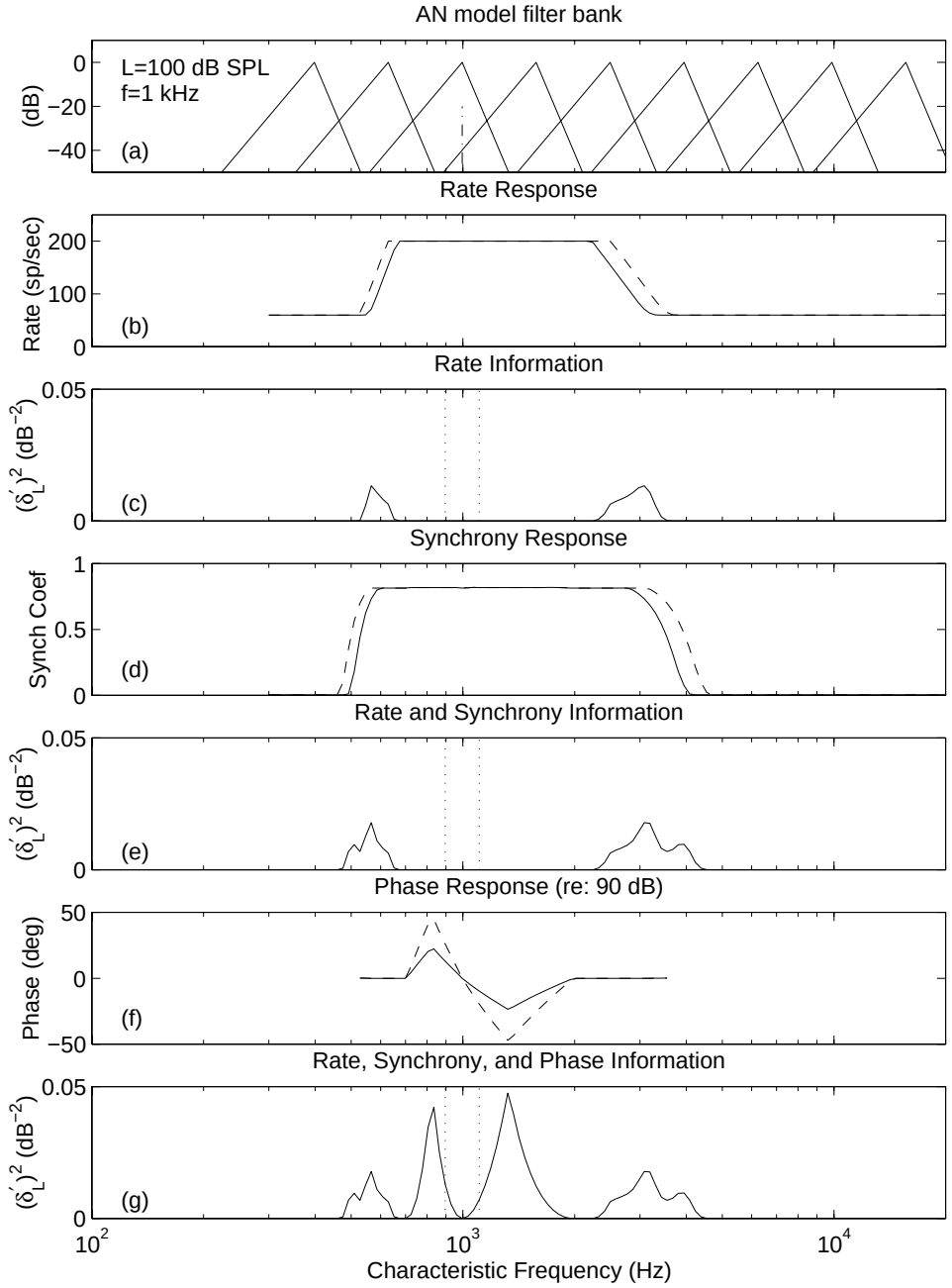

Figure 4-3: Distribution of rate, synchrony, and phase information across the AN population of high-spontaneous-rate fibers for level discrimination of a low-frequency, high-level tone. (a) AN filter bank with a 1-kHz, 100-dB SPL tone. (b) Average discharge rate as a function of CF for two tones of slightly different level. (c) Average-rate information (normalized sensitivity squared) as a function of CF. The vertical dotted lines indicate the restricted-CF region used in the present study to emphasize the dynamic-range problem. (d) Synchrony coefficient (or vector strength, which ranges from 0 to 1, see Johnson, 1980) as a function of CF for both tones. (e) Information available from both rate and synchrony cues. (f) Normalized phase response (relative to 90 dB) for both tones. (g) Total information from rate, synchrony, and phase cues.

tion is shown in panel (e). Similar to rate information, there is no synchrony information near the frequency of the tone for high levels due to saturation of the synchrony coefficient. The useful information from synchrony cues is spread further away from the tone than rate information due to lower synchrony thresholds (Colburn, 1981).

The distribution and relative contribution of nonlinear phase cues is illustrated in panels (f) and (g). The phase responses (relative to the phase at 90 dB SPL) of both tones are shown as a function of CF in panel (f). Auditory-nerve fibers with CF's above and below the tone frequency have phase responses that change with level, and thus contribute information. There are no changes in phase at CF, or well away from CF where the BM response is linear. The distribution of the total information provided from rate, synchrony, and phase cues is shown in panel (g). By comparing panels (e) and (g), the significant contribution of nonlinear-phase cues to the encoding of level can be seen. While there is no information for the CF equal to the tone frequency, there is significant phase information just below and just above the tone frequency since the nonlinear phase responses are restricted to near-CF frequencies. The amount of phase-information is roughly twice as large as the rate and synchrony information. Most importantly, the only information available in the restricted-CF region is information from nonlinear phase cues. In considering the benefits of the cochlear amplifier to level encoding, it is important to note that while compression is present over a wide dynamic range on the BM, saturation of average discharge rate below 40 dB SPL limits the encoding of this nonlinear property in the majority of AN fibers.

There are a small number of low-spontaneous-rate (LSR) fibers that have a larger dynamic range than the majority of AN fibers [Fig. 4-1(d) and (f)]. Therefore, it has been suggested that the LSR fibers are responsible for the encoding of level based on average-rate information in narrow frequency regions at high sound levels (e.g., Delgutte, 1987; Viemeister, 1988a,b; Winslow and Sachs, 1988; Winter and Palmer, 1991). Figure 4-4 compares the distribution of information across CF for HSR and LSR fibers at low and high frequencies. The contribution of nonlinear phase information in HSR fibers is compared to the contribution of average-rate information in LSR fibers. The limited-CF region is indicated by the vertical dotted lines in the top panels, and this region is magnified in the bottom panels. At low frequencies (left column), the HSR fibers (solid curve) contribute significant information from the nonlinear

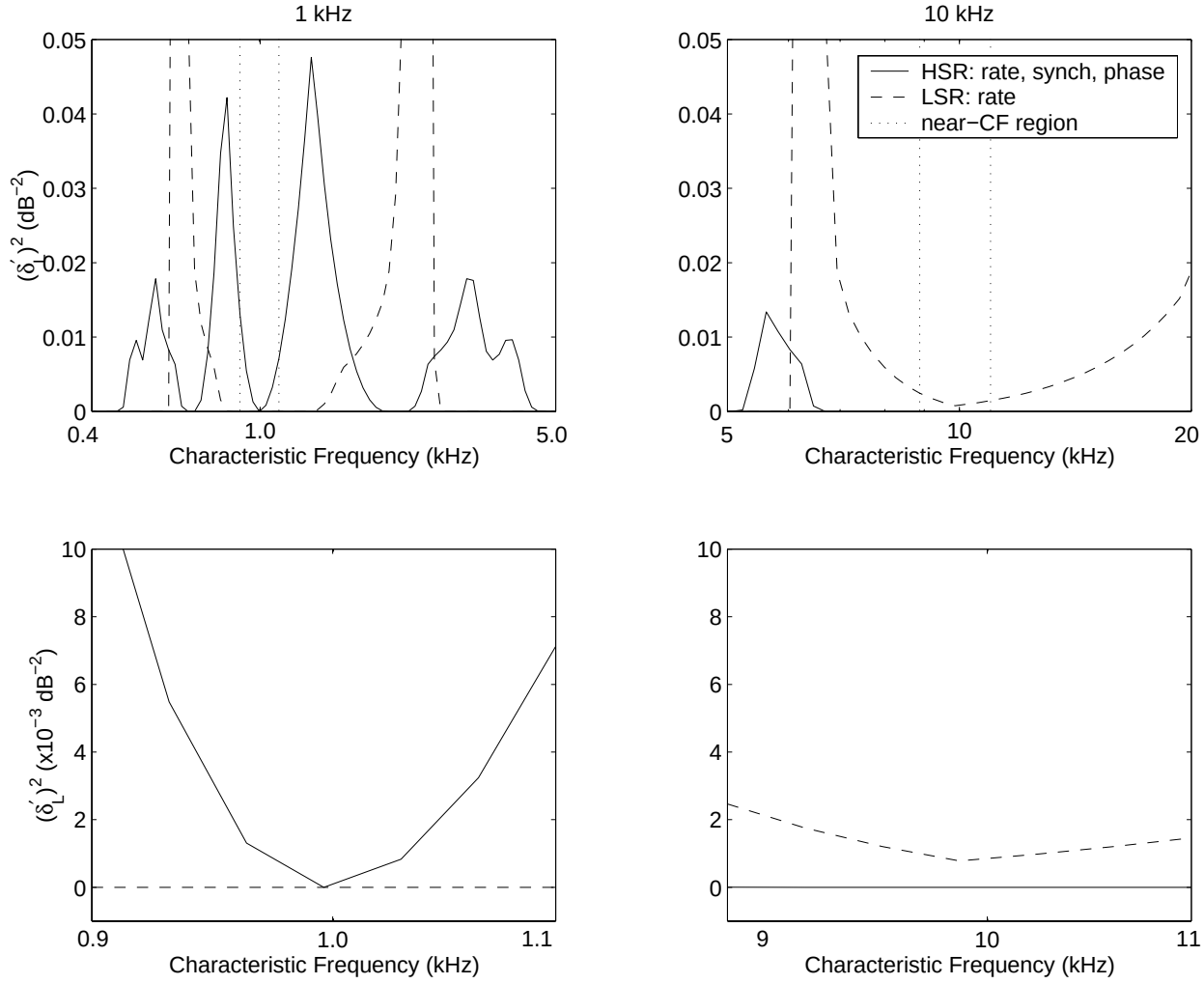

Figure 4-4: Comparison of nonlinear phase information in high-spontaneous-rate (HSR) fibers and average-rate information in low-SR fibers for the encoding of high sound levels in CF's near the tone frequency. The distribution of information across the AN population is shown for level discrimination of 100-dB SPL low- (1 kHz, left column) and high-frequency (10 kHz, right column) tones. The HSR curves (solid) represent rate, synchrony, and phase information, while the LSR curves (dashed) represent only average-rate information. Physiologically realistic properties of HSR and LSR fibers were used to scale the predictions (see text). The vertical dotted lines in the top row represent the narrow-CF region discussed in the present study, which is magnified in the bottom row. Same stimulus conditions as Fig. 4-3.

phase cues within the limited CF region. The LSR fibers (dashed curve) do not contribute any information within the limited CF region for the 1-kHz tone because the LSR fibers saturate at 80 dB [see Fig. 4-1(f)] due to the small amount of gain from the cochlear amplifier at low frequencies [Fig. 4-1(c)]. In contrast, at high frequencies (right column) the HSR fibers contribute no phase information due to the rolloff in phase locking [Fig. 4-1(e)], while the LSR fibers contribute significant average-rate information within the narrow CF region. The large gain from the cochlear amplifier at high frequencies [Fig. 4-1(c)] results in very shallow ("straight") rate-level curves for LSR fibers at high frequencies [Fig. 4-1(d)].

4.4.2 Predicted performance based on a narrow CF region

The dynamic range problem for the encoding of sound level is most severe in the case where the spread of excitation is limited (e.g., due to the presence of a notched-noise masker). This condition is simulated in the present study by considering information contained in a small set of model CF's (seven) surrounding the frequency of the tone. Level-discrimination performance (in terms of ΔL, Eq. 4.4) is predicted for a low- (996 Hz) and a high-frequency (9874 Hz) tone, where the tone frequency was chosen to be equal to one of the 120 model CF's. For the low and high tone frequencies, the near-CF regions used were 896-1107 Hz and 8882-10977 Hz, respectively. Performance was predicted for the HSR, MSR, LSR, and total populations of AN fibers, based on the physiological proportions described by Liberman (1978). Predicted level discrimination performance is compared for the rate-place and all-information models, where information in the seven model CF's was assumed to be combined optimally, and was scaled to account for the number of AN fibers represented by each model CF. Performance for the monaural-coincidence scheme was calculated based on the same number of total AN fibers as the rate-place and all-information predictions. The seven model CF's in the narrow-frequency region were assumed to innervate a set of four coincidence counters, one of which had both CF-inputs equal to the tone frequency. The other three coincidence counters received one CF-input above and one below the tone frequency, which were both separated from the tone frequency by an equal number (one, two, or three) model CF's. Based on the assumption that each AN fiber innervates only one coincidence counter, one-half as many same-CF-input coincidence counters were included in the total coincidence

population as were coincidence counters with different CF inputs. Thus, rate-place, all-information, and coincidence predictions are all based on the same set of AN fibers with CF's near the tone frequency.

Figure 4-5 (left column) shows level-discrimination performance based on the HSR fibers within the narrow frequency region in terms of ΔL as a function of stimulus level L for a low-frequency tone. Rate-place, all-information, and coincidence performance is shown in the top, middle, and bottom panels, respectively. Note that the scale for the ordinate of the coincidence panel is different than that for the rate-place and all-information panels. In order to illustrate the relative contributions from nonlinear gain and nonlinear phase properties, performance based on four versions of the AN model are shown in each panel. The circles represent performance based on the AN model with both linear gain and phase, the triangles represent nonlinear gain and linear phase, the x's represent linear gain and nonlinear phase, and the squares correspond to the AN model with both nonlinear gain and nonlinear phase. Because all three models are identical below 30 dB SPL, the four curves lie on top of one another at low levels in each panel. Levels for which symbols are not shown for a given model represent conditions in which no information existed for that model (i.e., infinite JND).

Average-rate information encodes changes in level only over a limited dynamic range (top row). Because of the low thresholds and corresponding limited dynamic range of HSR fibers, there is no rate-place information above 40 dB SPL for the AN model versions with linear gain. When the nonlinear gain is included in the AN model, the dynamic range over which changes in level are encoded is extended by 10 dB at low frequencies, and by 20 dB at high frequencies. The degradation in performance at 30 dB SPL for the nonlinear-gain models results from the compressive BM response that begins at 30 dB SPL [Fig. 4-1(b)]. The larger influence of the nonlinear gain at high frequencies compared with low frequencies is due to the CF-dependence of the cochlear-amplifier gain.

The contributions of synchrony and nonlinear-phase information are demonstrated by comparing rate-place and all-information predictions. The role of synchrony information is most clearly illustrated with the linear-phase versions of the AN model. Synchrony information improves performance at low levels for the low-frequency tone; however, synchrony does not extend the dynamic range to higher levels due to the fact that synchrony co-

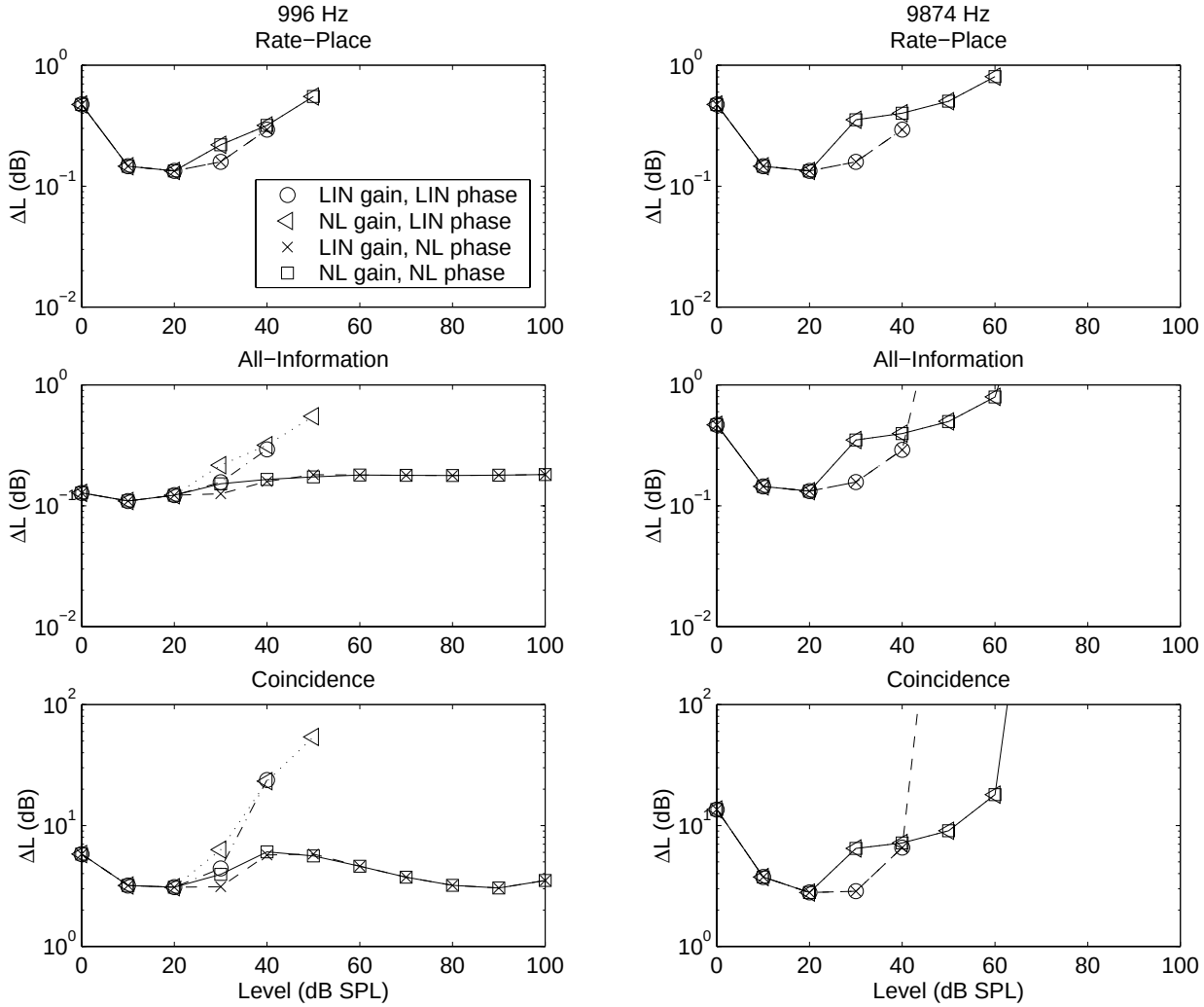

Figure 4-5: Level-discrimination performance based on the population of high-spontaneous-rate (HSR) fibers in a narrow range of CF's near the tone frequency (see text). The just-noticeable difference ΔL is plotted as a function of stimulus level for a 996 Hz (left column) and a 9874 Hz (right column) tone (500-ms duration). Optimal performance based on average-rate and all information is shown in the top and middle rows, respectively. Performance based on a set of monaural coincidence counters is shown in the bottom row. [Note the scale difference between rows.] Four versions of the AN model are shown in each panel to illustrate the effect of nonlinear gain and phase responses. Predictions from the four model versions are identical below 30 dB SPL. Levels for which symbols are not shown represent conditions in which there is no information available for a particular model (i.e., infinite JND).

efficients saturate at lower levels than average rate (Colburn, 1981). At high frequencies, rate-place and all-information predictions are essentially the same because of the sharp rolloff of phase-locking at high frequencies [Fig. 4-1(e)]. The extended dynamic range from the nonlinear gain is also seen in all-information predictions because the all-information encoding scheme includes both temporal and average-rate information. The significant contribution of nonlinear-phase responses at low frequencies is illustrated by comparing the rate-place and all-information predictions for the nonlinear-phase versions of the AN model. The dynamic range for level discrimination is extended up to 100 dB SPL based on the nonlinear phase responses. This is in sharp contrast to the rate-place information, which does not encode changes in level in HSR fibers above 50 dB SPL.

The predictions from the simple coincidence-counter model follow the general trends in the all-information predictions for both low and high frequencies, but are more than an order of magnitude worse than optimal all-information performance. The coincidence-model decodes the average-rate (and some of the synchrony) information that dominates performance below 30 dB SPL. The coincidence mechanism successfully decodes nonlinear phase cues provided by the cochlear amplifier at low frequencies, as illustrated by the relatively flat performance in the coincidence predictions for the AN model with nonlinear phase responses.

Predictions based on the set of LSR AN fibers within the narrow frequency region are shown in Fig. 4-6. Comparison of the rate-place predictions from the AN models with linear and nonlinear gain illustrates the ability of LSR fibers to encode high sound levels in a small set of CF's. At both low and high frequencies, the nonlinear gain extends the dynamic range over which changes in level are encoded; however, performance degrades significantly above 40 dB SPL at low frequencies. Changes in level of a 1-kHz tone are not encoded above 90 dB SPL in the average rate of the set of LSR fibers with CF's near the frequency of the tone. In contrast, LSR rate-place performance for a high-frequency tone is roughly constant across a wide dynamic range, up to 100 dB SPL. The all-information predictions demonstrate that the nonlinear phase responses extend the dynamic range of LSR fibers at low frequencies up to 100 dB SPL.

The coincidence predictions at low frequencies also show a small benefit from the nonlinear gain responses; however, performance based on nonlinear gain alone significantly degrades

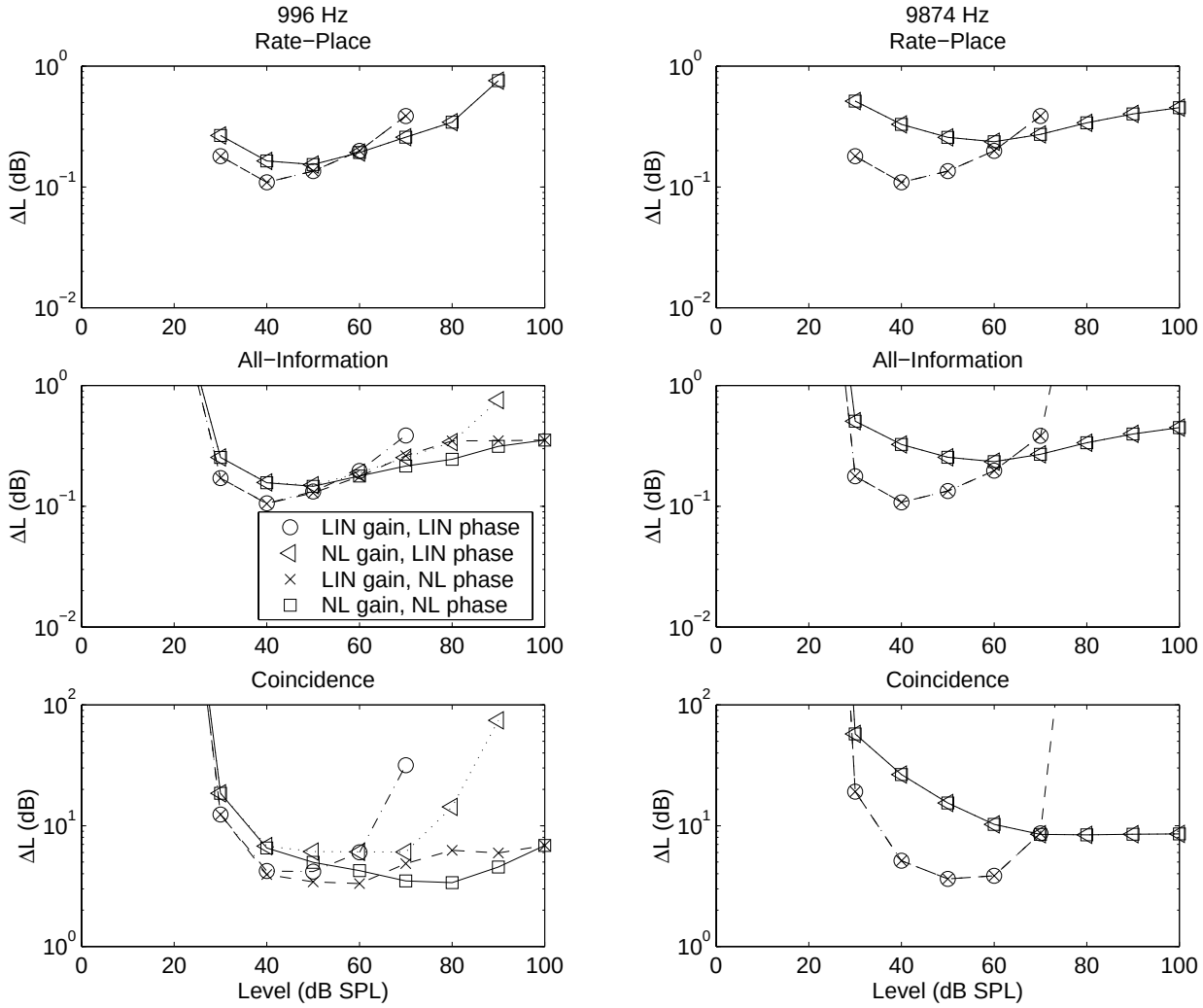

Figure 4-6: Level-discrimination performance based on the population of low-spontaneous-rate (LSR) fibers in a narrow range of CF's near the tone frequency (see text). Same symbols as Fig. 4-5.

above 70 dB SPL. The benefit from the nonlinear phase cues is also decoded by the coincidence counters, extending the dynamic range beyond that based on nonlinear gain alone. However, performance based on the coincidence counters with the nonlinear AN model degrades above 80 dB SPL at low frequencies. At high frequencies, coincidence performance is constant above 70 dB SPL when the nonlinear gain is included in the AN model, while changes in level are not encoded above 70 dB SPL when the nonlinear gain is removed.

The contribution of each of the three SR groups to level-discrimination performance based on the narrow range of CF's is shown in Fig. 4-7 for both low- and high-frequency tones. Performance based on each SR group individually is shown with performance based on the optimal combination of all three SR groups (squares) for the AN model with nonlinear gain and nonlinear phase. The rate-place predictions illustrate that the HSR and MSR fibers are primarily responsible for performance at low levels, while the LSR fibers are responsible for high levels (roughly above 50 dB SPL at both low and high frequencies). Performance based on the combination of average-rate information in the three SR groups degrades by an order of magnitude between 20 and 90 dB SPL at low frequency, and no changes in level are encoded above 90 dB SPL. This finding is consistent with the predictions of Colburn (1981) and Delgutte (1987), and indicates that changes in level are not encoded robustly in average-rate information within a narrow CF region across a wide dynamic range. Changes in level are encoded much more consistently across level at high frequencies, as a result of the large cochlear amplifier gain at high frequencies.

The all-information predictions based on all three SR groups are roughly constant across a dynamic range of 100 dB at low frequencies, unlike the rate-place predictions. Thus, Weber's Law is achieved based on information within a narrow range of CF's at low frequencies only when nonlinear phase information is included. Furthermore, performance based on the HSR fibers is as good and often better than performance based on the MSR and LSR fibers for low frequencies, especially at high levels. At high frequencies, the all-information predictions are similar to the rate-place predictions due to the small amount of temporal information available at high frequencies.

Performance at low frequencies based on the coincidence counters demonstrates roughly the same pattern as the all-information predictions. Performance is roughly constant across

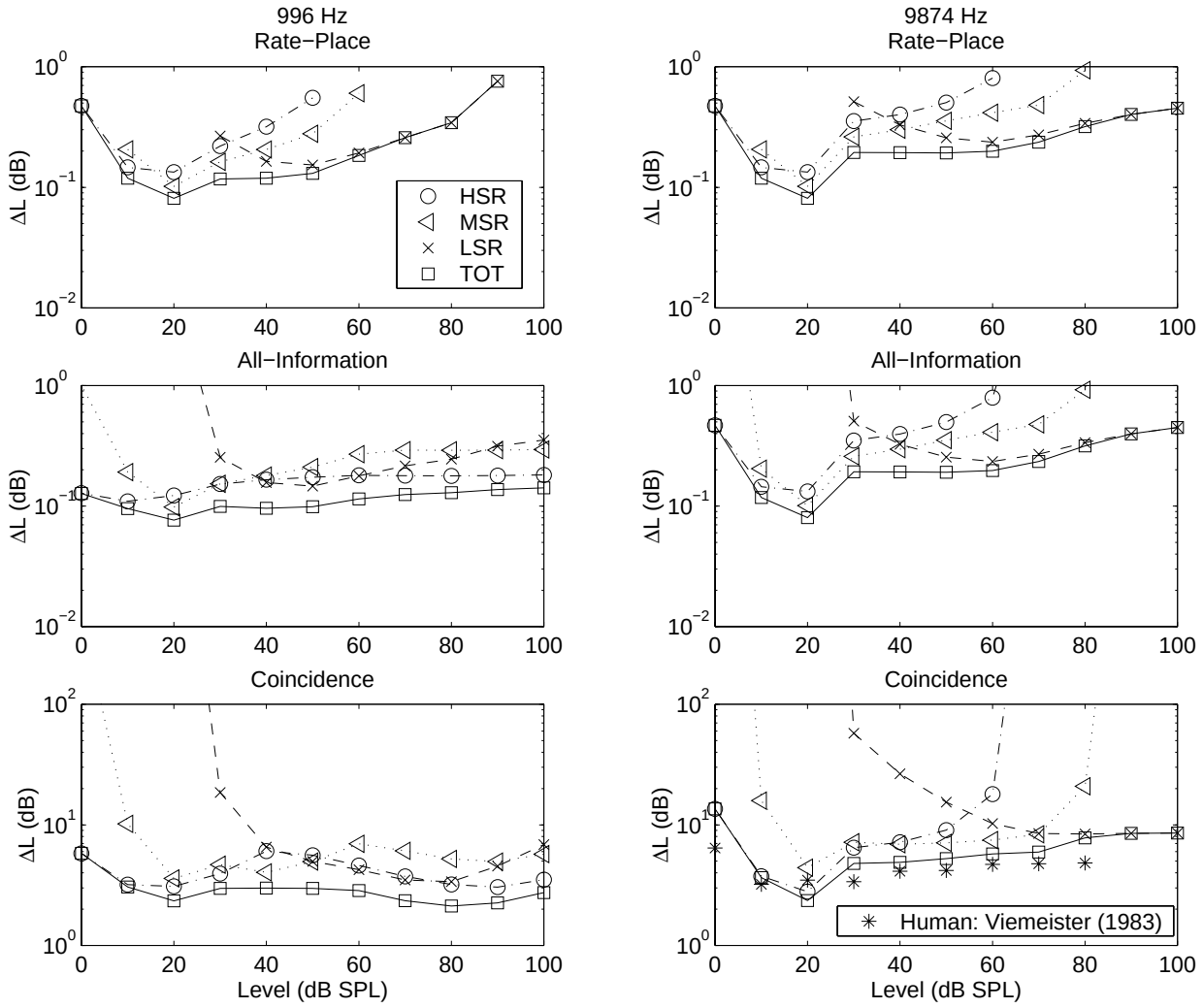

Figure 4-7: Level-discrimination performance based on individual and combined spontaneous-rate groups in a narrow range of CF's near the tone frequency (see text). The just-noticeable difference ΔL for the nonlinear-gain, nonlinear-phase AN model is plotted as a function of stimulus level for a 996 Hz (left column) and a 9874 Hz (right column) tone (500-ms duration). Optimal performance based on average-rate and all information is shown in the top and middle rows, respectively. Performance based on a set of monaural coincidence counters is shown in the bottom row. [Note the scale difference between rows.] HSR: high-spontaneous-rate; MSR: medium-spontaneous-rate; LSR: low-spontaneous-rate; TOT: optimal combination of all three SR groups. Levels for which symbols are not shown represent conditions in which there is no information available (i.e., infinite JND). Human data for level-discrimination of a 200-ms high-frequency noise band (6-14 kHz) in the presence of a notched-noise is shown by the stars in the bottom right panel (Viemeister, 1983).

a wide dynamic range, and is primarily determined by the HSR fibers. At high frequencies, performance based on the coincidence counters is also roughly constant from 10 to 100 dB SPL, and is determined by HSR at low levels, by MSR at medium levels, and by LSR fibers at high levels. For comparison, human performance for level discrimination of a high-frequency, bandpass noise (6-14 kHz) in a noise masker with a 6-14 kHz notch is shown by the stars (Viemeister, 1983). Predicted performance based on the coincidence counters within the narrow-frequency region matches the human performance very closely across the entire range of levels, while optimal rate-place and all-information performance is roughly an order of magnitude better than human performance.

4.4.3 Predicted performance based on all CF's

The results presented above have focused on the ability of human listeners to perform level discrimination in a notched-noise masker across a wide range of levels. Human performance in this task was compared to model predictions from a narrow range of CF's based on the assumption that the notched-noise masker eliminates all information within the noise. However, the effects of suppression and temporal information within the noise may affect this simple assumption and are not included in the present model. The analytical AN model describes responses to pure-tone stimuli, and thus it is most appropriate to compare predicted performance based on the entire population of AN fibers to human performance in a pure-tone level discrimination task in quiet. Level-discrimination performance (Eq. 4.4) was predicted for a low- (996 Hz) and a high-frequency (9874 Hz) tone in quiet based on the HSR, MSR, LSR, and total populations of AN fibers. Rate-place and all-information predictions are based on the optimal combination of information from the 120 model CF's. Performance based on the monaural-coincidence scheme was calculated based on the same number of total AN fibers as the rate-place and all-information predictions. The same four coincidence counters were used as in the narrow-CF predictions described above (i.e., based on seven model CF's surrounding each coincidence CF. The total normalized sensitivity squared was the sum of the individual normalized-sensitivities-squared from each of the four

coincidence counters at each of the 120 model CF's. [1] In this implementation, each AN fiber innervates only one coincidence counter, and thus the normalized sensitivities squared can be summed based on the assumption of independent AN fibers.

Predicted rate-place, all-information, and coincidence performance based on the entire HSR population is shown in Fig. 4-8 for a low- and high-frequency, 500-ms tone. Performance based on the four versions of the AN model are shown in each panel. Rate-place performance based on the linear AN model is flat above 20 dB SPL for the low-frequency tone, consistent with the predictions of Weber's Law based on the spread of excitation (Siebert, 1968). For the high-frequency tone, there is a small rise in ΔL above 60 dB SPL for the linear AN model. This rise is due to the upper side of the excitation spreading beyond the highest CF, and is consistent with the expected $\sqrt{2}$ reduction in ΔL due to the loss of one-half of the information. There is only a small effect of nonlinear gain at 30 and 40 dB SPL for the low-frequency tone; otherwise the cochlear amplifier does not influence rate-place performance based on the population of CF's. In contrast, there is a significant effect of nonlinear gain for the high-frequency tone. Performance is nonmonotonic, with a higher ΔL for 30-50 dB SPL than for 20 or 60 dB SPL. This degradation in performance at mid-levels results from the large amount of compression at high frequencies, and thus is much less prevalent for the low-frequency tone. The presence of a mid-level bump at high-, but not low-, frequencies is consistent with human performance (Florentine *et al.*, 1987).

The all-information predictions in Fig. 4-8 illustrate the contribution of synchrony and nonlinear-phase information to level discrimination of tones in quiet. At low frequencies, synchrony improves performance for the linear AN model by a factor of five at 0 dB SPL, and by a factor slightly less than two for higher levels. Weber's Law is again predicted above 20 dB SPL based on the contributions of rate and synchrony information. The near-miss to Weber's Law is present in the low-frequency predictions when the nonlinear phase responses are included in the AN model. This is consistent with the nonlinear phase responses producing Weber's Law within narrow-CF regions (Fig. 4-5), and the combination of information across CF's producing the near-miss to Weber's Law. High-frequency performance predictions

[1] Coincidence counters for which one input fell outside the model CF range (300-20,000 Hz) were evaluated by using the edge model CF as the input instead. This end effect is not significant for this study's results.

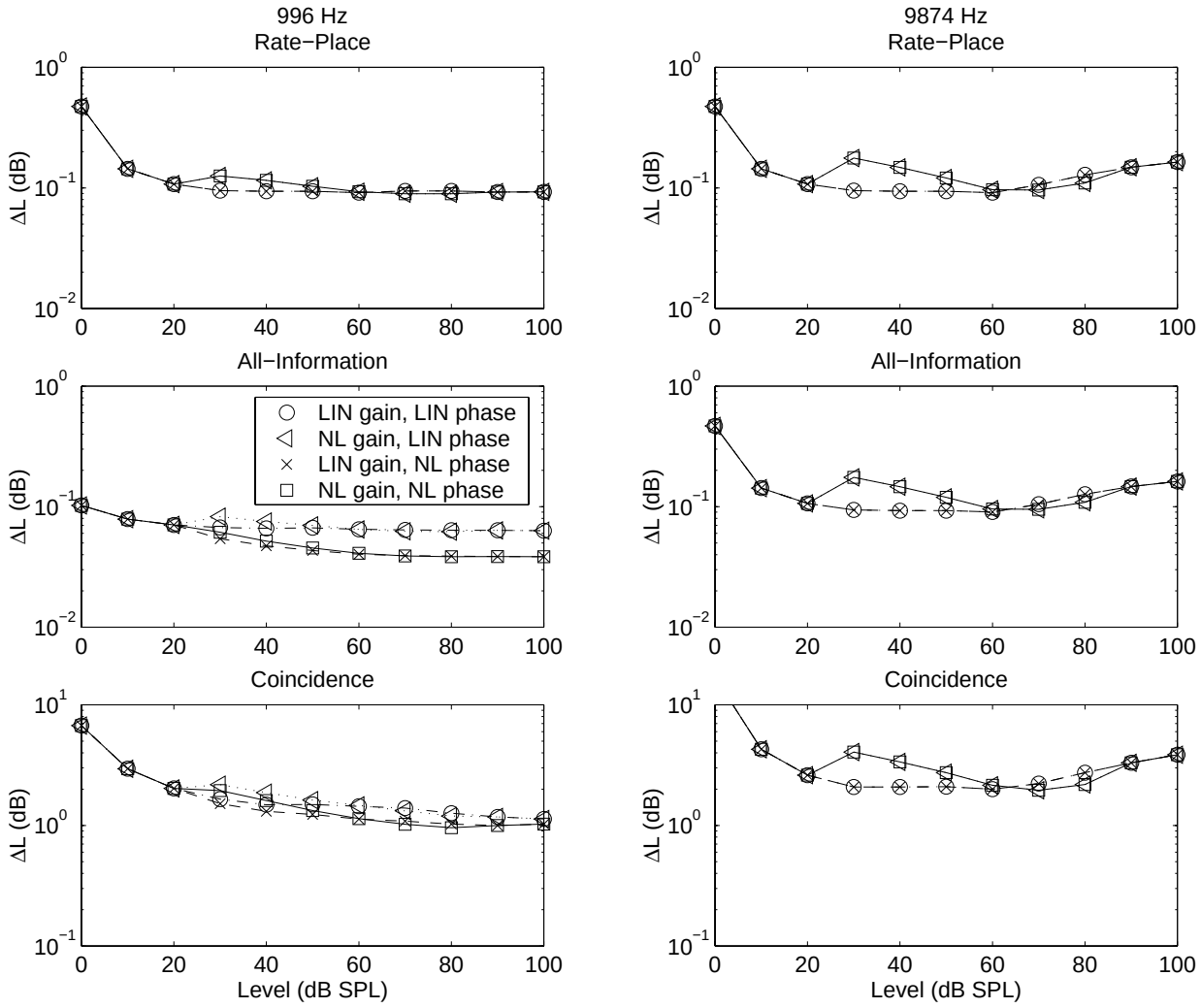

Figure 4-8: Level-discrimination performance based on the total population of high-spontaneous-rate (HSR) fibers. Same symbols as Fig. 4-5.

for rate-place and all-information are essentially identical due to the very low synchrony coefficient at 10 kHz [Fig. 4-1(e)]. Predicted trends based on the population of coincidence counters (lower row) generally resemble the all-information predictions at both low and high frequencies, except that the benefit from synchrony information at very low levels is not observed. Overall performance based on the population of coincidence counters is more than an order of magnitude worse than optimal performance, but is generally between 1 and 5 dB, which is roughly consistent with human performance levels (Florentine, *et al.*, 1987).

Predictions based on the population of LSR fiber are shown in Fig. 4-9. The rate-place predictions demonstrate a small effect of the nonlinear gain at low frequencies, and a larger effect at high frequencies. The shallower slopes of the rate-level curves for LSR fibers in the nonlinear model result in poorer performance at low- and mid-levels, where fibers for which the tone is within the near-CF nonlinear region dominate performance. No mid-level bump is observed in the LSR performance because changes in level below 30 dB SPL are not encoded in LSR fibers. The general trends and absolute performance levels are similar between the rate-place and all-information predictions at both low and high frequencies. The coincidence predictions also match the general trends in the all-information predictions, but are again more than an order of magnitude worse than optimal performance.

Predicted performance based on the three SR groups and on the total AN population of the nonlinear AN model is shown in Fig. 4-10. The rate-place predictions for the low-frequency tone demonstrate that each of the three SR groups contribute essentially equally above 50 dB SPL. Performance based on the total population of AN fibers (squares) decreases only slightly between 30 and 80 dB SPL. Rate-place predictions at high frequencies show that both the HSR and MSR population have a mid-level bump, while the LSR population does not. Performance below 70 dB SPL is determined primarily by the HSR and MSR fibers, while the LSR fibers determine performance above 90 dB SPL.

The all-information predictions for the low-frequency tone demonstrate that performance based on the HSR population is always better than performance based on the MSR or LSR populations when synchrony and nonlinear-phase information are included. All-information performance based on the total AN population shows a slightly larger decrease in ΔL across level than the rate-place predictions based on the contributions of nonlinear phase informa-

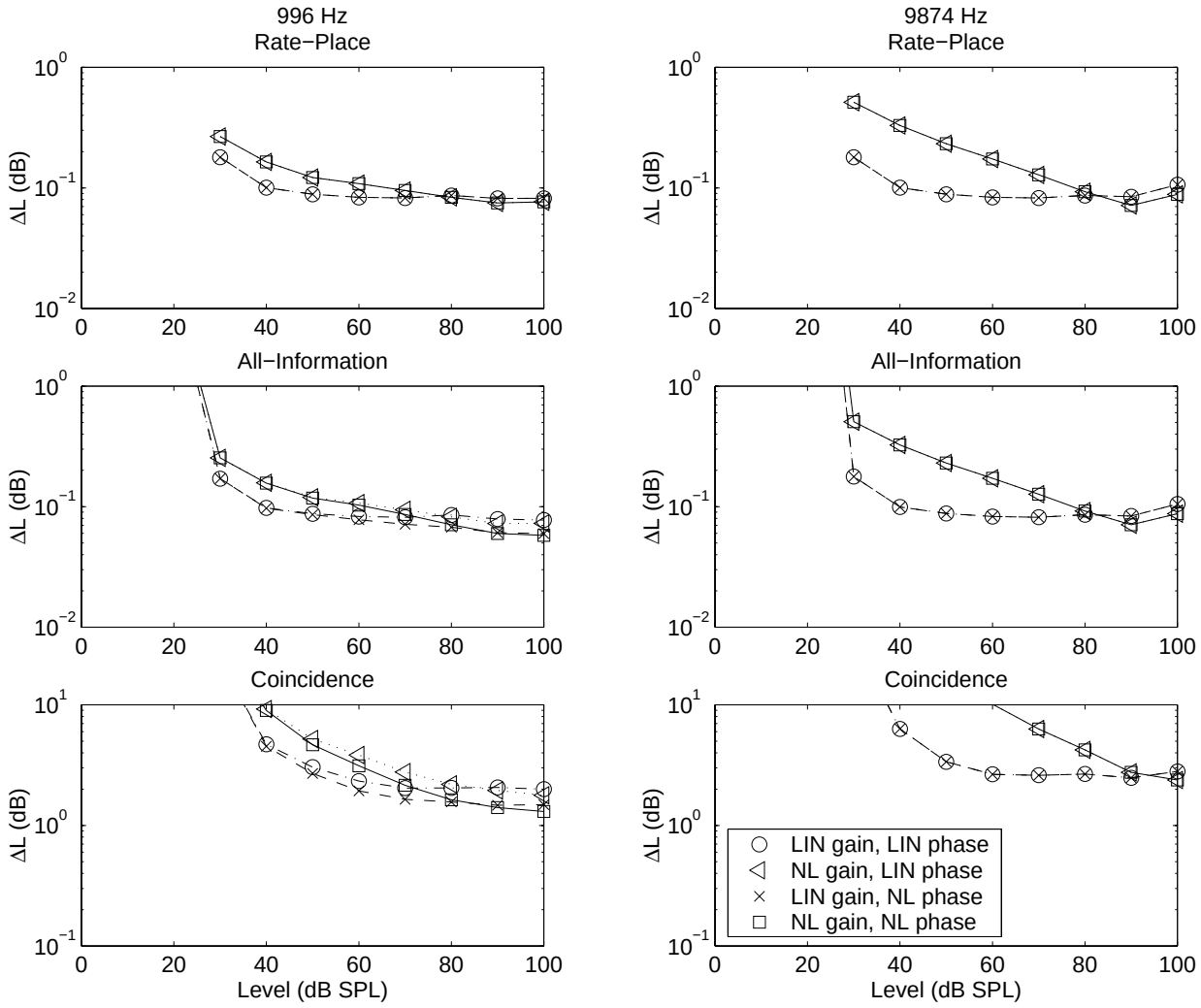

Figure 4-9: Level-discrimination performance based on the total population of low-spontaneous-rate (LSR) fibers. Same symbols as Fig. 4-5.

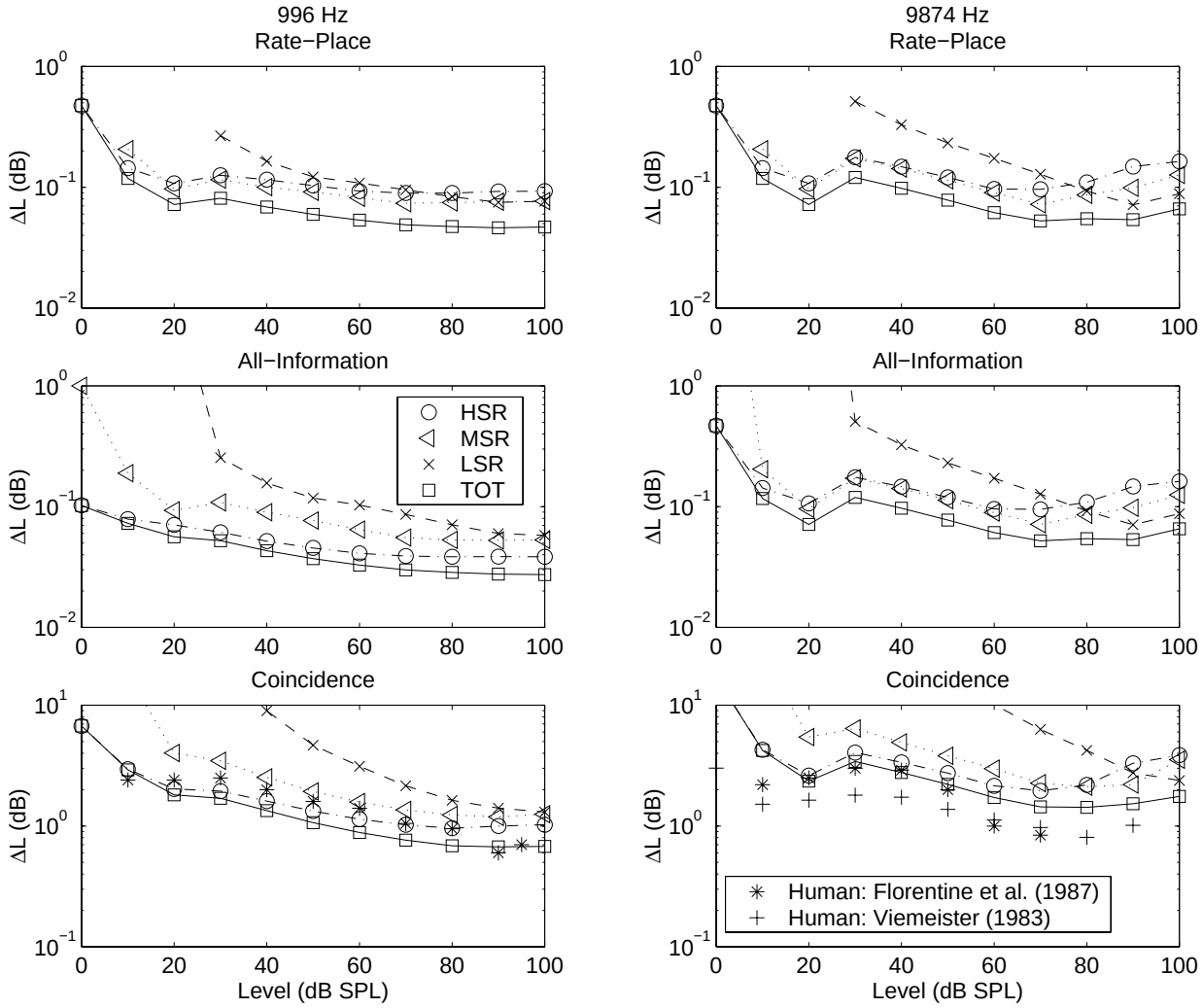

Figure 4-10: Level-discrimination performance based on the total population of individual and combined spontaneous-rate groups (same symbols as Fig. 4-7). Human data for level discrimination of 500-ms tones measured as a function of sensation level is shown by the stars in the bottom left and right panels (Florentine *et al.*, 1987). Human level-discrimination data for a 200-ms high-frequency noise band (6-14 kHz) in quiet is shown by the pluses in the bottom right panel (Viemeister, 1983).

tion. The all-information predictions at high frequencies essentially match the rate-place predictions.

Predictions based on the coincidence counter populations are shown in the bottom row, and are compared to human performance measured by Florentine *et al.* (1987) as a function of sensation level for the same low- and high-frequency conditions.[2] Human performance measured by Viemeister (1983) for level discrimination of a high-frequency, narrowband noise is also shown for comparison in the bottom right panel. At low frequency, performance based on the HSR population is always better than performance based on the MSR and LSR, similar to the all-information predictions. Above 90 dB SPL, all three SR groups contribute essentially equally. Coincidence performance based on the total AN population matches human performance very closely. Predicted performance is slightly better (within a factor of two) than human performance at most levels. The slope of the near-miss to Weber's Law observed in the human performance is matched by the coincidence predictions for the low-frequency tone, as well as the near-miss beginning at 30 dB SPL. The near-miss in the coincidence predictions results primarily from the nonlinear phase cues (which begin at 30 dB SPL) on the HSR fibers, and is not influenced by the population of LSR fibers.

At high frequencies, coincidence performance based on the HSR fibers is best among the three SR groups below 80 dB SPL, MSR performance is best at 90 dB SPL, and performance based on the LSR population is best for levels of 100 dB SPL and above. Coincidence performance based on the total AN population matches the human performance very closely, and is within a factor of two of both data sets at all levels. The nonmonotonic dependence on level (the "mid-level bump") observed in the both sets of human data is also demonstrated in the coincidence predictions. The level at which the bump occurs is well predicted by the coincidence performance. The size of the bump in the coincidence performance matches the data from Viemeister (1983), and is slightly smaller than the data from Florentine *et al.* (1987). A slight rise in ΔL as level increases at high levels is present in the coincidence performance, and is predicted based on the spread of the high-frequency information beyond

[2]The data plotted from Florentine *et al.* (1987) is their data as a function of sensation level (SL), which is the most appropriate comparison to the predictions from the analytical AN model that has a fixed threshold at 0 dB SPL. Some of the effects in the human data that occur at a particular sensation level are reduced when the data is averaged across listeners as a function of SPL, rather than SL.

the highest CF in the model. A slight rise is also often observed in human data when plotted as a function of SPL (Florentine *et al.*, 1987).

4.5 General Discussion

4.5.1 The benefit of the cochlear amplifier for extending the dynamic range of the auditory system

It has often been suggested that the cochlear amplifier is responsible for the extremely wide dynamic range of the auditory system (e.g., Yates, 1995). However, this suggestion has typically been based solely on the compressive magnitude response observed on the basilar membrane, and has never been tested quantitatively by relating the physiological properties associated with the cochlear amplifier to psychophysical performance. The present study quantifies the information available for level discrimination in the auditory nerve (AN) with and without the nonlinear gain and nonlinear phase responses that are associated with the cochlear amplifier. The results from the present study compare the relative contributions of nonlinear gain and nonlinear phase responses, as well as the relative contributions of high-, medium-, and low-spontaneous-rate AN fibers to encoding sound level.

Nonlinear gain

The ability of humans to discriminate changes in level consistently across a wide range of levels in Viemeister's (1983) experiment has been interpreted as demonstrating that Weber's Law is achieved based on average-rate within a narrow range of CF's (e.g., Delgutte, 1987). The notched-noise masker is presumed to mask the CF's away from the signal, and the high-frequency signal is presumed to rule out temporal information. The present predictions support this idea at high frequencies, where the cochlear amplifier gain is large enough that the LSR fibers can encode changes in sound level at high levels (Fig. 4-7). Fig. 4-6 demonstrates that the ability of LSR fibers to encode changes in high sound levels is due to the nonlinear gain associated with the cochlear amplifier. Changes in level of the high-frequency tone were only encoded up to 70 dB SPL in the LSR fibers when the nonlinear

gain response was removed from the AN model, but were encoded up to 100 dB SPL with the nonlinear gain.

However, the idea that average-rate information accounts for Weber's Law in a narrow-CF region fails at low frequencies. There is not enough cochlear amplifier gain at low frequencies to encode changes in level across the entire human dynamic range in the average rate of any of the three SR groups (Fig. 4-7). The LSR fibers only encode changes in level of the low-frequency tone up to 90 dB SPL based on average rate, and performance degrades significantly above 50 dB SPL. Thus, while LSR fibers have been implicated in the encoding of sound level at high levels, their usefulness at low frequencies is restricted. This result is consistent with the lack of "straight" rate-level curves for LSR fibers at low frequencies reported by Winter and Palmer (1991), and in cats at all frequencies (e.g., Sachs and Abbas, 1974; Winslow and Sachs, 1988).

Nonlinear phase

The nonlinear phase responses associated with the cochlear amplifier have rarely been considered for their ability to extend the dynamic range of the auditory system (Carney, 1994); however, they are significant because the wide dynamic range of their information about changes in level is present in all AN fibers, including the HSR fibers which comprise the majority of the AN population. Such a representation of level is preferred to the combination schemes across SR groups that are required for average-rate information to account for level discrimination across a wide range of levels. For example, an enhanced weighting of the average-rate information provided by LSR fibers must be assumed in order to produce Weber's Law at high levels in single-CF regions (Delgutte, 1987). Winslow *et al.* (1987) have suggested that level could be encoded based on average rate with a processor that relies on HSR fibers at low sound levels and LSR fibers at high levels. Figure 4-7 demonstrates that the nonlinear phase responses support Weber's Law based on an optimal combination of the three SR groups within a narrow range of CF's at low frequencies. In fact, performance based on the HSR fibers alone is relatively flat across the entire range of human hearing.

Carney (1994) has illustrated schematically how nonlinear phase shifts on single AN fibers turn into systematic temporal patterns across CF (i.e., spatio temporal patterns). She

showed responses for a bank of model AN fibers with different CF's as a function of time for several levels (her Fig. 5). The main feature is that as level increases, the trajectory across CF of the peaks in the discharge probability as a function of time become steeper (i.e., the responses across CF become more coincident). This motivated her to propose that sound level may be encoded in the spatio-temporal discharge patterns of AN fibers, and that an across-frequency coincidence mechanism could decode these level cues at medium to high levels. Figure 4-7 demonstrates that a set of monaural, across-frequency coincidence counters can encode sound level robustly across the entire range of human hearing based on AN fibers within a narrow range of CF's for both low- and high-frequency tones.

Because the cochlear amplifier acts most strongly in near-CF regions, the benefits from both gain and phase cues in extending the dynamic range within narrow-CF regions are particularly useful for complex stimuli such as speech, where spread of excitation is limited. This suggestion is consistent with the general finding that hearing-impaired listeners have the most difficulty with complex stimuli in difficult listening conditions where spread of excitation may not be possible. Due to the rolloff in AN phase-locking above 2-3 kHz, the greatest benefit from nonlinear phase cues is at low frequencies; however, the majority of important speech information is at low frequencies. The ability of nonlinear phase information to account for Weber's Law in narrow-CF regions at low frequencies (and the inability of average-rate information to do so), suggests that nonlinear phase cues should be considered in the encoding of complex stimuli at high levels. Loss of the cochlear amplifier would be expected to degrade the representation of complex stimuli in impaired ears due to loss of the nonlinear phase cues; however, this impairment would not be observed in physiological studies that quantify the reduction of AN information in impaired animals based only on average rate and synchronized rate responses (e.g., Miller *et al.*, 1997, 1999). In contrast, analyses of rate representations in the cochlear nucleus would be expected to demonstrate an impairment in the encoding of complex stimuli at high levels in impaired animals. The ability of monaural coincidence counters to encode changes in sound level at low frequencies across a much wider dynamic range than average-rate information in the AN may provide a basis for reports of enhanced rate representations in the cochlear nucleus. Blackburn and Sachs (1990) and May *et al.* (1998) have reported that rate representations of speech sounds

are enhanced in the ventral cochlear nucleus (chopper neurons) compared with the AN in normal-hearing animals.

4.5.2 Pure-tone level discrimination in quiet

Near-miss to Weber's Law at low frequencies

It has often been suggested that the near-miss to Weber's Law results from the combination of information across multiple CF's that individually achieve Weber's Law (e.g., Florentine and Buus, 1981; Delgutte, 1987; Viemeister, 1988a,b). The predictions from the present study are consistent with this idea; however, the contribution of nonlinear phase responses is responsible for the robust encoding of level in narrow-CF regions at low frequencies, rather than average-rate information in LSR fibers, which has been most often hypothesized (Delgutte, 1987; Viemeister, 1988a,b; Winter and Palmer, 1991). Only a small improvement in performance across level is predicted for a low-frequency tone based on average rate information combined optimally across the three SR groups (Fig. 4-10). This small improvement is consistent with predictions from Delgutte (1987), which were based on the means and variances of physiologically measured discharge counts as a function of level. Although performance improves slightly as level increases based on average discharge rate, the improvement is too small to account for human performance at low frequencies where the near-miss is observed (Florentine *et al.*, 1987).

Thus, the present predictions suggest that the only effect of the nonlinear responses associated with the cochlear amplifier for level discrimination of low frequency tones is that the near-miss, rather than Weber's Law, is predicted based on the nonlinear phase responses. The degree of the near-miss in the all-information predictions is larger than in the rate-place predictions, but is largest and matches human performance only when performance based on coincidence is considered. Although there are many intuitive reasons to believe that the cochlear amplifier would strongly influence level discrimination of tones, the predicted effect is quite small. This nonintuitive finding results from the fact that at medium to high sound levels, where cochlear compression has a strong effect on BM responses, the primary information about changes in level is contributed by HSR fibers with CF's away from the

frequency of the tone. Thus, the CF's that are dominating performance at medium and high sound levels are responding linearly because the nonlinear effects associated with the cochlear amplifier are restricted to near-CF frequencies.

Several other physiological models have produced a near-miss with only low-threshold, HSR fibers. Teich and Lachs (1979) demonstrated the near-miss with a rate-place model that had more rounded filter shapes than Siebert's filters and that incorporated the effects of refractoriness on AN discharge-count variance. Delgutte's (1987) model included average tuning-curve shaped filters and realistic AN-count variance. The results in Chapter 2 predicted a significant near-miss based on a computational AN model with gamma-tone filters and Poisson discharge statistics. The ability of many models to predict the near-miss based on different mechanisms supports the idea suggested by Viemeister (1988a) that the near-miss to Weber's Law is not a critical aspect of the dynamic-range problem, and that the robust encoding of sound level in narrow-CF regions is the most important issue.

Mid-level bump at high frequencies

A puzzling detail of human level discrimination of tones in quiet is that performance is nonmonotonic at high frequencies, in contrast to the slight improvement in performance with level at low frequencies (e.g., Carlyon and Moore, 1984; Florentine, *et al.*, 1987). Many of the psychophysical experiments exploring the "mid-level bump" (or the "severe departure from Weber's Law") have focused on short-duration signals in various noise maskers, because the effect (when reported as $\Delta I/I$) is generally larger for short-duration signals (e.g., Carlyon and Moore, 1984), and can be enhanced or reduced by various configurations of notched-noise maskers (e.g., Oxenham and Moore, 1995; Plack, 1998). However, the analytical AN model used in the present study is only appropriate to compare to long duration conditions because onset/offset responses and neural adaptation are not included in the model. In addition, the effects of a notched-noise masker on the different types of AN information must be considered quantitatively using methods that are beyond the present study, as discussed below. Thus, the high-frequency mid-level bump reported by Florentine *et al.* (1987) for level discrimination of 500-ms pure tones in quiet is an appropriate comparison for the present predictions.

Plack (1998) has discussed several explanations for the mid-level bump based on both peripheral and central mechanisms; however, only those that are addressed by the present predictions are discussed here. Carlyon and Moore (1984) have suggested that the mid-level bump at high frequencies may be explained by two populations of AN fibers. Good performance was suggested to be provided by the low-threshold, HSR fibers at low levels and by the high-threshold, LSR fibers at high levels, with a degradation in performance at mid levels resulting from neither population encoding changes in sound level. The absence of the mid-level bump at low frequencies was suggested to result from synchrony information providing good performance at mid levels. The present predictions do not support the explanation by Carlyon and Moore (1984). For the high-frequency tone, the transition between HSR and LSR fibers determining performance occurs near 80-90 dB SPL (Fig. 4-10), and no degradation in performance occurs because of the contribution of the third SR group (MSR) reported by Liberman (1978). The contribution of synchrony information at low frequencies is restricted to levels below where HSR fibers contribute rate information (Fig. 4-10), and thus cannot be responsible for good performance at mid levels. von Klitzing and Kohlrausch (1994) have suggested that the mid-level bump can be explained based on mid-level compression on the BM; however, their explanation requires that BM responses become linear above roughly 50 dB SPL, which has been shown not to be true in healthy cochleae (Ruggero *et al.*, 1997).

The present predictions demonstrate that a mid-level bump that is consistent with human data results from the large amount of cochlear amplifier gain at high frequencies (Figs 4-8 and 4-10). The degradation in model performance at mid levels is due to the BM input-output function becoming strongly compressed at 30 dB SPL (Fig. 4-8). As level increases further, the spread of excitation goes beyond the near-CF nonlinear region, and performance is dominated by HSR AN fibers that are responding linearly. Thus, the nonlinear AN model predictions for the HSR fibers match those from the linear AN model at level above 60 dB SPL. The lack of a mid-level bump at low frequencies in the model predictions is consistent with the small cochlear-amplifier gain at low frequencies [Fig. 4-1(c)]. The presence of a mid-level bump at high frequencies, but not at low frequencies, is consistent with human performance (Florentine *et al.*, 1987).

Thus, the present predictions suggest that the main effect of the cochlear amplifier on level discrimination of high-frequency tones in quiet is a degradation in performance at mid levels. This hypothesis suggests that hearing-impaired listeners without a healthy cochlear amplifier would not show a mid-level bump. However, this would be difficult to measure due to the typically limited dynamic range in hearing-impaired listeners (Florentine *et al.*, 1993).

4.5.3 Coincidence detection: A robust, physiologically realistic neural mechanism

The predictions from the present study suggest that a set of monaural, across-frequency coincidence counters that receive AN inputs from a narrow range of CF's can account for Weber's Law across the dynamic range of human hearing, both at low and high frequencies. This finding is significant because Weber's Law in narrow-CF regions appears to be required to account for human level-discrimination performance (Florentine and Buus, 1981; Viemeister, 1983); however, an optimal combination of average-rate information across the set of three SR groups in the AN does not produce Weber's Law across a wide range of levels, at least at low frequencies (Fig. 4-7; Colburn, 1981; Delgutte, 1987). While the predictions in the present study suggest an important role for monaural, across-frequency coincidence detection in the encoding of sound level, it is important to consider whether this mechanism is likely to be present in the auditory system.

Coincidence detection is a physiologically realistic mechanism, because any neuron with multiple sub-threshold inputs acts as a coincidence detector (i.e., several coincident discharges across the inputs are required to produce an output discharge). There is strong evidence that coincidence detection occurs in the binaural auditory system (Yin and Chan, 1990; Goldberg and Brown, 1969; Rose *et al.*, 1966; Yin *et al.*, 1987; Joris *et al.*, 1998), and coincidence detection forms the basis of most models of binaural processing (Colburn, 1996). Carney (1990) showed that several cell types in the antero-ventral cochlear nucleus (AVCN), primarily globular bushy cells, were sensitive to changes in the relative phase across their inputs. Thus, there is physiological evidence for the existence of cells in the AVCN that have responses that are consistent with a monaural, across-frequency coincidence mechanism.

The simple coincidence-counting mechanism analyzed in the present analysis was shown to decode the level-dependent phase cues associated with the cochlear amplifier, as suggested by Carney (1994). However, the coincidence predictions in the present study demonstrate similar trends to the all-information encoding scheme at both low and high frequencies and for all sound levels. Thus, monaural, across-frequency coincidence detection is a robust mechanism for encoding sound level in that both average rate and nonlinear phase information from AN discharges is encoded in the coincidence counts, as well as some information from synchrony cues.

It is generally accepted that there is far more information in AN responses than is used by humans, and that an inefficient processor is needed to account for human level-discrimination performance (e.g., Colburn, 1981; Delgutte, 1987). The coincidence-counting model processes the AN discharge times inefficiently. Information is lost in the process of coincidence detection because only the times of coincident AN discharges are considered. Additional information is lost by basing performance only on the coincidence counts (i.e., by ignoring the coincidence times). Even though the coincidence mechanism in the present study is far from optimal, the coincidence-performance predictions are typically shifted upward roughly in parallel from the all-information predictions. This uniform degradation in performance is desirable because, while the trends in all-information performance often match the trends in human performance, optimal performance is often at least an order of magnitude better than human performance. Figure 4-7 and 4-10 illustrate that the degradation in performance that occurs due to the coincidence mechanism results in absolute performance levels for level discrimination that are very close to human performance. In addition, the monaural coincidence mechanism eliminates the requirement of an inefficient processor that varies its inefficiency as a function of level, which has been suggested to be required based on average-discharge rate information in the AN (Colburn, 1981; Delgutte, 1987).

The derivations of performance based on a monaural, across-frequency coincidence counter described in the Appendix in Section 4.9 suggest an interesting property that could be useful for physiological studies of neurons that are hypothesized to perform coincidence detection. The ratio of the expected value of coincidence counts (Eq. 4.29) to the variance of counts (Eq. 4.30) is dependent only on the properties of the temporal coincidence window

$f(x)$. This ratio would be expected to be independent of stimulus parameters, and therefore the statistics of the observed discharge counts may be able to be used to make inferences about the shape and size of the coincidence window of a given neuron.

4.5.4 Limitations of present study

Predictions in the present study were compared to human data for level discrimination in a notched noise based on the common assumption that the only effect of the notched noise is to eliminate spread of excitation of the tone. In order to accurately evaluate the validity of this assumption, the effect of the noise masker on different types of information must be quantified. This analysis requires two advances beyond the present study: (1) a more complex AN model, and (2) an extension of the statistical decision theory (SDT) analysis.

In order to accurately evaluate complex stimuli, the AN model must include a description of two-tone suppression properties (e.g., Sachs and Kiang, 1968; Delgutte, 1990b; Ruggero *et al.*, 1992). The AN response to a CF-tone in the presence of a notched-noise masker may be suppressed by the noise, whereas the response of AN fibers with CF's within the noise may be suppressed by the tone. Such complex interactions between the tone and noise maskers could contribute significant information to detection or discrimination of signals in noise, and therefore need to be quantified.

An extension of the SDT analysis beyond the present study is required in order to quantify the relative effects of physiological (internal) and stimulus (external) variation on psychophysical performance. Chapter 3 described an extension of the SDT analysis to discrimination tasks in which a single parameter is randomly varied (e.g., random-level variation). A general theoretical analysis of detection or discrimination of a signal in random noise is described in Chapter 5, and has been applied to the detection of tones in notched noise in Chapter 6.

4.6 Conclusions

1. The cochlear amplifier benefits normal-hearing listeners by extending the dynamic range within narrow frequency regions, e.g., for complex stimuli such as speech in

noise for which hearing-impaired listeners have the most difficulty. Highly compressive basilar-membrane responses at high frequencies allow for the robust encoding of level based on average discharge rate; however, average rate does not robustly encode sound level at low frequencies because the cochlear-amplifier gain is too small. Nonlinear phase responses associated with the cochlear amplifier encode changes in level across the entire dynamic range of hearing; however, the rolloff in phase locking reduces the effectiveness of phase cues at high frequencies for simple stimuli.

2. The cochlear amplifier has only a small effect on level discrimination of pure tones in quiet, consistent with the difficulty in measuring a deficit for hearing-impaired listeners in simple tasks. The nonlinear phase cues produce the "near-miss" to Weber's Law at low frequencies, while the nonlinear gain responses account for the nonmonotonic "mid-level bump" at high frequencies.

3. Coincidence detection is a physiologically realistic mechanism that can decode the nonlinear gain and phase cues provided by the cochlear amplifier. Performance based on a population of coincidence counters matches human performance across the entire dynamic range of hearing at both low and high frequencies.

4.7 Acknowledgments

The authors thank Torsten Dau for providing valuable comments on an earlier version of this manuscript. We thank Don Johnson for his synchrony data from cat shown in Fig. 4-1(e), and Mario Ruggero for providing the data shown in Fig. 4-1(g). Part of this work was presented at the joint 137^{th} meeting of the Acoustical Society of America, 2nd convention of the European Acoustics Association: Forum Acusticum 99, and the 25th German Acoustics DAGA conference. This work was supported in part by the National Institute of Health, Grant T32DC00038, and by the National Science Foundation, Grant 9983567.

4.8 Appendix: Nonlinear analytical auditory-nerve - model

The analytical nonlinear auditory-nerve (AN) model used in the present study incorporates the salient features of the cochlear amplifier. The assumptions and equations that specify the analytical model are described in this appendix.

The discharge statistics of AN fibers are assumed to be well described by a non-stationary Poisson process with a time-varying rate function $r(t)$. Equation 4.1 describes the phase-locked response of an AN fiber in response to a tone burst. The average discharge rate $\bar{r}[L_{eff}]$ and the strength of phase locking $g[L_{eff}, f_0]$ both depend on the effective level L_{eff} that drives each AN fiber.

The effective level for the i^{th} AN fiber is determined by the tone level L and by the nonlinear-filter magnitude response H_{NL} for the characteristic frequency CF_i and the tone frequency f_0, i.e.,

$$L_{eff}(L, f_0, CF_i) = L + 20\log_{10}\left[H_{NL}\left(f_0, CF_i, L\right)\right]. \tag{4.7}$$

The implementation of nonlinear-tuning in the present model [see Fig. 4-1(a)-(d)] represents the idea that the cochlear amplifier produces high sensitivity and sharp tuning at low levels by providing amplification to near-CF frequencies, and that the cochlear-amplifier gain is reduced as level increases (Yates, 1995). At low levels, the magnitude response is described by linear triangular filters that are consistent with those used by Siebert (1968, 1970) to describe tuning curves in cat, i.e.,

$$H_S\left(\frac{f_0}{CF_i}\right) = \begin{cases} \left(\frac{f_0}{CF_i}\right)^{10}, & f_0 \le CF_i, \\ \left(\frac{f_0}{CF_i}\right)^{-20}, & f_0 \ge CF_i. \end{cases} \tag{4.8}$$

Nonlinear compression is incorporated into the magnitude response H_{NL} by multiplying the triangular magnitude response H_S by a level- and frequency-dependent attenuation factor,

i.e.,

$$H_{NL}\left(f_0, CF_i, L\right) = \left(10^{[\gamma_{dB}(f_0, CF_i, L)/20]}\right) H_S\left(\frac{f_0}{CF_i}\right). \tag{4.9}$$

The reduction in gain of the cochlear amplifier is specified based on several simple assumptions, consistent with physiological findings (e.g., Ruggero *et al.*, 1997): (1) For a given CF, the maximum gain G provided by the cochlear amplifier is produced for tones presented at CF, and G increases with CF according to

$$G(CF_i) = \begin{cases} G_{min}, & CF_i \leq 500 \text{ Hz}, \\ G_{min} + (G_{max} - G_{min})\frac{\log_{10}(CF_i/500)}{\log_{10}(8000/500)}, & CF_i \in [500, 8000] \text{ Hz}, \\ G_{max}, & CF_i \geq 8000 \text{ Hz}, \end{cases} \tag{4.10}$$

where $G_{min} = 20$ dB, $G_{max} = 60$ dB [see Fig. 4-1(c)]. (2) The cochlear amplifier provides full gain for levels below $L_{thr}^{NL} = 30$ dB SPL, and the cochlear amplifier gain is systematically reduced as level increases from L_{thr}^{NL} to $L_{sat}^{NL} = 120$ dB SPL [see Fig. 4-1(b)]. (3) The cochlear amplifier only provides amplification for frequencies near-CF (i.e., $f_0 \in [f_{lf}^{NL}, f_{hf}^{NL}]$, where

$$\begin{aligned} f_{lf}^{NL} &= CF_i\left(10^{\frac{-G_{max}}{20 \cdot 10}}\right), \\ f_{hf}^{NL} &= CF_i\left(10^{\frac{G_{max}}{20 \cdot 20}}\right), \end{aligned} \tag{4.11}$$

as shown in Fig. 4-1(a). This simple implementation of the nonlinear frequency region results in a flat magnitude response between f_{lf}^{NL} and f_{hf}^{NL} at high levels ($L \geq L_{sat}^{NL}$) and high frequencies ($f \geq 8000$ Hz). Based on these assumptions, the level- and frequency-dependent reduction in gain is given in dB by

$$\gamma_{dB}(f_0, CF_i, L) = \beta_{mag}(L, f_0)\left\{-20\log_{10}\left[H_S\left(\frac{f_0}{CF_i}\right)\right] - G_{max}\right\}\left[\frac{G(CF_i)}{G_{max}}\right], \tag{4.12}$$

where

$$\beta_{mag}(L, f_0) = \begin{cases} 0, & L \le L_{thr}^{NL} \text{ or } f_0 \notin [f_{lf}^{NL}, f_{hf}^{NL}], \\ \left(L - L_{thr}^{NL}\right) \Big/ \left(L_{sat}^{NL} - L_{thr}^{NL}\right), & L \in [L_{thr}^{NL}, L_{sat}^{NL}] \text{ and } f_0 \in [f_{lf}^{NL}, f_{hf}^{NL}], \\ 1, & L \ge L_{sat}^{NL} \text{ and } f_0 \in [f_{lf}^{NL}, f_{hf}^{NL}]. \end{cases} \quad (4.13)$$

In order to evaluate the effect of the nonlinear magnitude response on predictions in the present study, versions of the AN model with and without the nonlinear magnitude responses can be compared. The nonlinear magnitude responses can be turned off by setting $\beta_{mag}(L, f_0) = 0$ for all levels and frequencies.

The dependence of average discharge rate $\bar{r}$ on the effective level L_{eff} of an AN-fiber is specified in terms of a simple saturating nonlinearity (based on Colburn, 1981)

$$\bar{r}[L_{eff}] = \begin{cases} SR, & L_{eff} \le L_{thr} - 5, \\ SR + \frac{1}{600}(R_{sat} - SR)(L_{eff} - L_{thr} + 5)^2, & L_{eff} \in [L_{thr} - 5, L_{thr} + 5], \\ SR + \frac{1}{30}(R_{sat} - SR)(L_{eff} - L_{thr}), & L_{eff} \in [L_{thr} + 5, L_{thr} + 30], \\ R_{sat} & L_{eff} \ge L_{thr} + 30, \end{cases} \quad (4.14)$$

which depends on the spontaneous rate (SR), the saturated rate (R_{sat}), and the rate threshold (L_{thr}). This saturating nonlinearity is shown by the HSR rate-level curves in Fig. 4-1(d) and (f). The dependence of phase locking on effective level L_{eff} is specified using the same general form of saturating nonlinearity, i.e.,

$$g[L_{eff}, f_0] = g_{max}(f_0) \begin{cases} 0, & L_{eff} \le L_{thr} - 25, \\ \frac{1}{600}(L_{eff} - L_{thr} + 25)^2, & L_{eff} \in [L_{thr} - 25, L_{thr} - 15], \\ \frac{1}{30}(L_{eff} - L_{thr} + 20), & L_{eff} \in [L_{thr} - 15, L_{thr} + 10], \\ 1 & L_{eff} \ge L_{thr} + 10, \end{cases} \quad (4.15)$$

where the dependence of synchrony on frequency is matched to data from cat [see Fig. 4-1(e);

Johnson, 1980], and is described by

$$g_{max}(f_0) = 3.1 \begin{cases} 1, & f_0 \leq 1200, \\ \frac{1200}{f_0}, & f_0 \in [1200, 2800], \\ \frac{1200*2800^2}{(f_0)^3}, & f_0 \geq 2800. \end{cases} \tag{4.16}$$

The threshold for phase-locking is specified to be 20 dB below the average-rate threshold.

The implementation of the nonlinear-phase responses in the present study is based on several simple assumptions: (1) The level-dependent phase responses are limited to the same near-CF frequency region as the magnitude responses, $f_0 \in [f_{lf}^{NL}, f_{hf}^{NL}]$. (2) Phase varies linearly with tone level. (3) The maximum phase changes occur half-way into the near-CF nonlinear frequency region, i.e., at frequencies

$$\begin{aligned} f_{lf}^{PH} &= 0.5(CF_i + f_{lf}^{NL}), \\ f_{hf}^{PH} &= 0.5(CF_i + f_{hf}^{NL}). \end{aligned} \tag{4.17}$$

(4) The maximum phase shifts between low levels and 80 dB SPL are roughly $\pi/2$. (5) The total traveling-wave delay at high levels ($L > L_{sat}^{NL}$) is compensated for in each CF with neural delays prior to innervation of the coincidence-detection population. This simple assumption is based on the fact that onset responses to clicks are very strong for many cell types in the cochlear nucleus that receive convergent inputs from a range of CF's (Young, 1984; Rhode and Greenberg, 1992). Based on these simple assumptions, the nonlinear filter phase response is specified by the equation

$$\theta(f_0, CF_i, L) = \beta_{phase}(L, f_0) \cdot 2\Delta\theta_{max} \begin{cases} 0, & f_0 \notin [f_{lf}^{NL}, f_{hf}^{NL}], \\ (f_0 - f_{lf}^{NL})/(CF_i - f_{lf}^{NL}), & f_0 \in [f_{lf}^{NL}, f_{lf}^{PH}], \\ (CF_i - f_0)/(CF_i - f_{lf}^{NL}), & f_0 \in [f_{lf}^{PH}, CF_i], \\ (CF_i - f_0)/(f_{hf}^{NL} - CF_i), & f_0 \in [CF_i, f_{hf}^{PH}], \\ (f_0 - f_{hf}^{NL})/(f_{hf}^{NL} - CF_i), & f_0 \in [f_{hf}^{PH}, f_{hf}^{NL}], \end{cases} \tag{4.18}$$

where $\Delta\theta_{max} = \frac{6}{5}\pi$ is the maximum phase change between L_{thr}^{NL} and L_{sat}^{NL}, and $\beta_{phase}(L, f_0) =$

$1 - \beta_{mag}(L, f_0)$, where $\beta_{mag}(L, f_0)$ is specified by Eq. 4.13. Versions of the AN model with and without the nonlinear phase changes can be compared in order to evaluate the effect of nonlinear phase responses. The level-dependent phase changes can be turned off by setting $\beta_{phase}(L, f_0) = 1$ for all levels and frequencies.

4.9 Appendix: Performance based on a monaural coincidence counter

This appendix presents derivations of the expected value and variance of the coincidence counts $\mathcal{C}_{ij}\{\mathcal{T}^i, \mathcal{T}^j\}$ (Eq. 4.2) that are needed to calculate the normalized sensitivity δ' of a monaural, across-frequency coincidence counter (Eq. 4.6). Similar equations and related discussions are presented by Colburn (1969; 1977b) for a binaural coincidence counter, but without derivations. The expected value and variance in Eq. 4.6 depend on the Poisson statistics of the two sets of independent AN discharge times,$\mathcal{T}^i$ and $\mathcal{T}^j$, and will be shown to be given by

$$E_{\mathcal{T}^i,\mathcal{T}^j,\phi}[\mathcal{C}_{ij}\{\mathcal{T}^i, \mathcal{T}^j\}] = E_\phi \left[\int_0^T \int_0^T f(x-y) r_i(x;\phi) r_j(y;\phi) dx dy \right], \quad (4.19)$$

$$Var_{\mathcal{T}^i,\mathcal{T}^j,\phi}[\mathcal{C}_{ij}\{\mathcal{T}^i, \mathcal{T}^j\}] \simeq E_\phi \left[\int_0^T \int_0^T f^2(x-y) r_i(x;\phi) r_j(y;\phi) dx dy \right], \quad (4.20)$$

where $r_i(t;\phi)$ and $r_j(t;\phi)$ represent the time-varying rate functions of the two AN inputs to the coincidence counter, ϕ is a random phase imposed on every AN fiber, and $f(x)$ is the narrow temporal coincidence window.

The derivations of $E_{\mathcal{T}^i,\mathcal{T}^j,\phi}[\mathcal{C}_{ij}\{\mathcal{T}^i, \mathcal{T}^j\}]$ and $Var_{\mathcal{T}^i,\mathcal{T}^j,\phi}[\mathcal{C}_{ij}\{\mathcal{T}^i, \mathcal{T}^j\}]$ will rely on several general results for decision variables of the form

$$X_i(\mathcal{T}^i) = \sum_{l=1}^{K_i} s(t_l^i), \quad (4.21)$$

where $s(t_l^i)$ is any function of the l^{th} discharge time t_l^i generated from a Poisson process with rate function $r_i(t)$. It can be shown (e.g., Rieke *et al.*, 1997) that the expected value and

variance of $X_i(\mathcal{T}^i)$ are given by

$$E_{\mathcal{T}^i}\left[X_i(\mathcal{T}^i)\right] = \int_0^T s(t)r_i(t)dt, \tag{4.22}$$

$$Var_{\mathcal{T}^i}\left[X_i(\mathcal{T}^i)\right] = \int_0^T s^2(t)r_i(t)dt, \tag{4.23}$$

based on the probability density function for Poisson discharge times (see Parzen, 1962; Snyder and Miller, 1991). Eqs. 4.22 and 4.23 imply that

$$E_{\mathcal{T}^i}\left\{\left[X_i(\mathcal{T}^i)\right]^2\right\} = \int_0^T s^2(t)r_i(t)dt + \left[\int_0^T s(t)r_i(t)dt\right]^2. \tag{4.24}$$

The expected value in Eq. 4.6 can be derived as follows

$$\begin{aligned} E_{\mathcal{T}^i,\mathcal{T}^j,\phi}[\mathcal{C}_{ij}\{\mathcal{T}^i,\mathcal{T}^j\}] &= E_{\mathcal{T}^i,\mathcal{T}^j,\phi}\left[\sum_{l=1}^{K_i}\sum_{m=1}^{K_j} f(t_l^i - t_m^j)\right], \\ &= E_{\mathcal{T}^i,\phi}\left\{\sum_{l=1}^{K_i} E_{\mathcal{T}^j}\left[\sum_{m=1}^{K_j} f(t_l^i - t_m^j)\middle|\mathcal{T}^i\right]\right\}, \\ &= E_{\mathcal{T}^i,\phi}\left[\sum_{l=1}^{K_i}\int_0^T f(t_l^i - y)r_j(y;\phi)dy\right]. \end{aligned} \tag{4.25}$$

Using Eq. 4.22 again, Eq. 4.19 is obtained. Similarly, the term $E_{\mathcal{T}^i,\mathcal{T}^j,\phi}[\mathcal{C}_{ij}^2\{\mathcal{T}^i,\mathcal{T}^j\}]$ can be derived as follows

$$E_{\mathcal{T}^i,\mathcal{T}^j,\phi}[\mathcal{C}_{ij}^2\{\mathcal{T}^i,\mathcal{T}^j\}] =$$

$$\begin{aligned} &= E_{\mathcal{T}^i,\phi}\left[E_{\mathcal{T}^j}\left(\left\{\sum_{m=1}^{K_j}\left[\sum_{l=1}^{K_i} f(t_l^i - t_m^j)\right]\right\}^2\middle|\mathcal{T}^i\right)\right], \\ &= E_{\mathcal{T}^i,\phi}\left\{\int_0^T\left[\sum_{l=1}^{K_i} f(t_l^i - y)\right]^2 r_j(y;\phi)dy + \left[\sum_{l=1}^{K_i}\int_0^T f(t_l^i - y)r_j(y;\phi)dy\right]^2\right\}, \\ &= E_\phi\left\{\int_0^T\int_0^T f^2(x-y)r_i(x;\phi)r_j(y;\phi)dxdy\right. \end{aligned}$$

$$+\int_0^T\int_0^T\int_0^T f(x-y)f(u-y)r_i(x;\phi)r_i(u;\phi)r_j(y;\phi)dxdudy$$
$$+\int_0^T\int_0^T\int_0^T f(x-y)f(x-v)r_i(x;\phi)r_j(v;\phi)r_j(y;\phi)dxdvdy$$
$$+\left[\int_0^T\int_0^T f(x-y)r_i(x;\phi)r_j(y;\phi)dxdy\right]^2\Bigg\}. \quad (4.26)$$

The variance in Eq. 4.6 is then equal to

$$Var_{\mathcal{T}^i,\mathcal{T}^j,\phi}[\mathcal{C}_{ij}\{\mathcal{T}^i,\mathcal{T}^j\}] =$$

$$\begin{aligned}
&= E_{\mathcal{T}^i,\mathcal{T}^j,\phi}[\mathcal{C}_{ij}^2\{\mathcal{T}^i,\mathcal{T}^j\}] - \left(E_{\mathcal{T}^i,\mathcal{T}^j,\phi}[\mathcal{C}_{ij}\{\mathcal{T}^i,\mathcal{T}^j\}]\right)^2 \\
&= E_\phi\Bigg\{\int_0^T\int_0^T f^2(x-y)r_i(x;\phi)r_j(y;\phi)dxdy \\
&\quad +\int_0^T\int_0^T\int_0^T f(x-y)f(u-y)r_i(x;\phi)r_i(u;\phi)r_j(y;\phi)dxdudy \\
&\quad +\int_0^T\int_0^T\int_0^T f(x-y)f(x-v)r_i(x;\phi)r_j(v;\phi)r_j(y;\phi)dxdvdy\Bigg\}. \quad (4.27)
\end{aligned}$$

The second and third terms in the right-hand side of Eq. 4.27 can be shown to be negligible relative to the first term if (1) the coincidence window $f(x)$ is much narrower than the period of variation in discharge rate $r(t)$, which is never less than about 0.5 ms given the rolloff in phase locking of AN fibers above 2 kHz [Fig. 4-1(e); Johnson (1980)], and (2)

$$R_{sat}\left[\int_{-\infty}^{\infty} f(x)dx\right]^2 \ll \int_{-\infty}^{\infty} f^2(x)dx, \quad (4.28)$$

where R_{sat} is the maximum discharge rate of an AN fiber. Both of these conditions are satisfied for the 10-μs rectangular coincidence window in the present study, and thus Eq. 4.20 holds.

Note that Eqs. 4.19 and 4.20 can be used with either analytical or computational AN models, because they are independent of the AN model used to produce the rate functions. More informative expressions for the expected value and variance can be derived for the present analytical AN model. The expected value (Eq. 4.19) and variance (Eq. 4.20) will be

shown to be given by

$$E[\mathcal{C}_{ij}|L] \simeq \frac{T\bar{r}_i\bar{r}_j}{I_0[g_i]\,I_0[g_j]} I_0\left[\sqrt{g_i^2+g_j^2+2g_ig_j\cos(\theta_i-\theta_j)}\right]\int_\infty^\infty f(x)dx, \quad (4.29)$$

$$Var[\mathcal{C}_{ij}|L] \simeq \frac{T\bar{r}_i\bar{r}_j}{I_0[g_i]\,I_0[g_j]} I_0\left[\sqrt{g_i^2+g_j^2+2g_ig_j\cos(\theta_i-\theta_j)}\right]\int_\infty^\infty f^2(x)dx, \quad (4.30)$$

where $g_i = g\left[L_{eff}(L, f_0, CF_i), f_0\right]$, $\bar{r}_i = \bar{r}\left[L_{eff}(L, f_0, CF_i)\right]$, and $\theta_i = \theta(L, f_0, CF_i)$, and $I_0\{g\}$ is the zeroth-order Bessel function.

The expected value can be derived by substituting the rate function $r(t)$ from the analytical AN model (Eq. 4.1) into (Eq. 4.19), i.e.,

$$E_{\mathcal{T}^i,\mathcal{T}^j,\phi}[\mathcal{C}_{ij}\{\mathcal{T}^i,\mathcal{T}^j\}] =$$

$$\begin{aligned}
&= \frac{\bar{r}_i\bar{r}_j}{I_0[g_i]\,I_0[g_j]}\int_0^T\int_0^T f(x-y) \\
&\quad \times\frac{1}{2\pi}\int_0^{2\pi}\exp\left[g_i\cos(2\pi f_0x+\theta_i+\phi)+g_j\cos(2\pi f_0y+\theta_j+\phi)\right]d\phi dxdy, \\
&= \frac{\bar{r}_i\bar{r}_j}{I_0[g_i]\,I_0[g_j]}\int_0^T\int_0^T f(x-y) \\
&\quad \times\frac{1}{2\pi}\int_0^{2\pi}\exp\Big(\sqrt{g_i^2+g_j^2+2g_ig_j\cos[2\pi f_0(x-y)+\theta_i-\theta_j]} \\
&\qquad \times\cos\left\{\phi+\tan^{-1}\left[\frac{g_i\sin(2\pi f_0x+\theta_i)+g_j\sin(2\pi f_0y+\theta_j)}{g_i\cos(2\pi f_0x+\theta_i)+g_j\cos(2\pi f_0y+\theta_j)}\right]\right\}\Big)d\phi dxdy, \\
&= \frac{\bar{r}_i\bar{r}_j}{I_0[g_i]\,I_0[g_j]}\int_0^T\int_0^T f(x-y) \\
&\quad \times I_0\left\{\sqrt{g_i^2+g_j^2+2g_ig_j\cos[2\pi f_0(x-y)+\theta_i-\theta_j]}\right\}dxdy. \qquad (4.31)
\end{aligned}$$

Equation 4.29, and similarly Eq. 4.30, result from the assumption that the coincidence window $f(x)$ is narrow relative to the period of stimulus variation.

Chapter 5

Quantifying the effects of noise maskers on signal information in auditory-nerve responses: I. General theory and application to detection in bandlimited noise

5.1 Abstract

Previous studies have evaluated psychophysical performance limits by combining auditory-nerve (AN) models with methods from statistical decision theory (e.g., Siebert, 1970; Colburn, 1973; Chapter 2). This approach is extended in the present study to include the case of stimuli with random noise maskers. Performance is evaluated for a general processor that makes decisions by evaluating each AN discharge time in terms of *a priori* information about the two alternatives in a discrimination (or detection) experiment. Different assumptions about the *a priori* information used by the processor are evaluated (e.g., information about only the noise ensemble or about individual noise waveforms). Optimum performance is obtained when the processor uses the exact rate waveforms for each noise waveform. This general approach allows the relative contributions of inherent stimulus fluctuations (external noise) and physiological (internal) noise to be evaluated by comparing performance for individual noises and for the random-noise ensemble. The nonlinear computational AN model used in the present study included gamma-tone filters with compressive magnitude responses, two-tone suppression, saturating rate-level curves, low-threshold high-spontaneous-

rate (HSR) fibers, rolloff in phase-locking, neural adaptation, and realistic onsets and offsets. Conditionally independent Poisson processes are used to model the population of AN fibers spaced according to a human cochlear frequency map. The theoretical analysis developed is illustrated for the detection of tones in narrowband and broadband noise maskers. The effect of noise maskers on predicted detection performance is compared for processors based on AN discharge times and based only on AN discharge counts. The distribution of information across the population of AN fibers and the types of temporal information available for the detection of tones in noise are illustrated. The analysis techniques developed here allow quantification of the influence of the correlation between the responses of AN fibers to random-noise stimuli on the ability to combine information across AN fibers.

5.2 Introduction

Noise stimuli have a long history of use as maskers in tests of the signal processing capabilities of the auditory system. Fletcher (1940) evaluated the frequency selectivity of the auditory system by measuring detection thresholds for tones in the presence of band-limited noise as a function of masker bandwidth. As masker bandwidth was increased, detection thresholds increased until a "critical bandwidth" (CB) was reached, beyond which thresholds were essentially constant. Fletcher interpreted these results as being consistent with the use of a single filter to process the tone-in-noise signal, which led to the concept of the "auditory filter." Many well established theories of auditory processing are based on the idea that the auditory system acts as a bank of overlapping filters, including the excitation-pattern model (Zwicker, 1956, 1970; Florentine and Buus, 1981; Glasberg and Moore, 1990), and the articulation-index model of speech perception developed by Fletcher (1953).

Much effort in psychophysics has gone into methods to measure the auditory filter; however, this measurement is difficult using simple stimuli such as single tones due to the presence of many overlapping filters. Complex stimuli, such as noise, have been used in an attempt to force the listener to rely on a single "auditory filter". These experiments are typically interpreted based on the power spectrum model of masking, which has several assumptions including (1) the auditory system processes sounds with an array of overlapping linear filters, (2) a single filter is used to detect signals in noise, typically the filter with the highest signal-to-noise ratio, (3) only the noise that passes through this single filter affects the masking of a tone, and (4) thresholds correspond to a fixed signal-to-noise ratio through the filter in terms

of the long-term power spectra of the stimuli, i.e., the relative phases of components and the short-term fluctuations in the masker are ignored (Moore, 1995a). Despite the fact that none of these assumptions are entirely correct, this model is able to explain many masking phenomena. Thus, the power spectrum model is the current basis for many researchers' view of the effects of noise stimuli on signal information in the auditory system.

Given that the power spectrum model is often used to interpret psychophysical experiments designed to measure physiological frequency selectivity, it seems appropriate to evaluate its assumptions based on information contained in auditory-nerve discharge patterns. Previous studies have evaluated psychophysical performance limits by combining auditory-nerve (AN) models with methods from statistical decision theory (e.g., Siebert, 1965, 1968, 1970; Colburn, 1969, 1973, 1977a,b 1981; Delgutte, 1987; Chapter 2). However, these studies have typically only incorporated response variability due to the stochastic nature of AN discharges, and thus have been limited to deterministic stimuli. The present study presents a new theoretical analysis that extends this approach to include the case of stimuli with random-noise maskers. Several examples of the application of the theory are presented for the detection of tones in narrowband and broadband noise maskers. The predictions from the present study address several assumptions of the power spectrum model by quantifying the effect of noise stimuli on average discharge rate and temporal information in the AN.

5.3 General Theory

In this section, we discuss the application of statistical decision theory (SDT) to calculate performance limits in a general discrimination task in the presence of random stimulus noise. The following analysis describes a level-discrimination experiment in which two tones with levels L_1 and L_2 are to be discriminated; however, this analysis is applicable to the discrimination of any parameter in the presence of random noise. Detection of a signal in random noise can be analyzed as a discrimination experiment between a signal of level $L_1 = -\infty$ dB and one of level L_2.

5.3.1 Separation of stimulus variance from auditory-nerve variance

In a psychophysical experiment with random noise, variability in observer responses results from both the random stimulus (*external noise*) and from the random auditory-nerve (AN) responses (*internal noise*) (Colburn, 1969, 1977a,b; Siegel and Colburn, 1989). It is important to separate the effects of these two noise sources for the interpretation of psychophysical performance limits.

The observations of AN responses available to any processor $Y(\mathcal{T})$ are the collection of Poisson discharge times from the M AN fibers, $\mathcal{T} = \{t_1^1, \cdots, t_{K_1}^1, \cdots, t_j^i, \cdots, t_{K_M}^M\}$, where t_j^i represents the j^{th} discharge time on the i^{th} AN fiber. Optimum performance based on the processor $Y(\mathcal{T})$ can be evaluated in terms of a sensitivity metric Q given by

$$Q(L_1, L_2) = \frac{\left\{E_{n,\mathcal{T}}\left[Y\left(\mathcal{T}\right)|L_2\right] - E_{n,\mathcal{T}}\left[Y\left(\mathcal{T}\right)|L_1\right]\right\}^2}{Var_{n,\mathcal{T}}\left[Y\left(\mathcal{T}\right)|L_1\right]}, \tag{5.1}$$

where the notation $E_{n,\mathcal{T}}$ indicates that the expected value is taken over both the stimulus noise waveform n and the AN discharge times $\mathcal{T}$. It is assumed that $Var_{n,\mathcal{T}}\left[Y\left(\mathcal{T}\right)|L_2\right]$ and $Var_{n,\mathcal{T}}\left[Y\left(\mathcal{T}\right)|L_1\right]$ are approximately equal when L_1 and L_2 are just noticeably different, which is defined as corresponding to $Q(L_1, L_2) = 1$. Note that the sensitivity index $Q(L_1, L_2)$ corresponds to the commonly used sensitivity index $(d')^2$ if $Y(\mathcal{T})$ has a Gaussian distribution.

For a given set of observations $\mathcal{T}$, the decision variable $Y(\mathcal{T})$ is a random variable for which the randomness comes from both the AN response and from the stimulus. The expectations and variance in Eq. 5.1 can be calculated using iterated expectations as

$$\begin{aligned} E_{n,\mathcal{T}}\left[Y\left(\mathcal{T}\right)\right] &= E_n\left\{E_{\mathcal{T}}\left[Y\left(\mathcal{T}\right)|n\right]\right\}, & (5.2)\\ Var_{n,\mathcal{T}}\left[Y\left(\mathcal{T}\right)\right] &= E_n\left\{Var_{\mathcal{T}}\left[Y\left(\mathcal{T}\right)|n\right]\right\} + Var_n\left\{E_{\mathcal{T}}\left[Y\left(\mathcal{T}\right)|n\right]\right\}, & (5.3)\\ &= Var_{AN} + Var_{STIM}, \end{aligned}$$

where the first term on the right side of Eq. 5.3 represents the variability due to the Poisson properties of AN discharges, and the second term represents the variability due to the stimulus (Colburn, 1969; 1977b). The ability to separate the contributions of AN and stim-

ulus variability to the variance of a decision variable will be useful below for evaluating the relative influence of each type of variability on psychophysical performance limits.

5.3.2 General form of optimum processor

The form of the optimum processor $Y_{opt}(\mathcal{T})$ for discriminating between L_1 and L_2 in the presence of a random noise is specified by a log-likelihood-ratio test,

$$Y_{opt}(\mathcal{T}) = \ln \frac{p(\mathcal{T}|L_2)}{p(\mathcal{T}|L_1)} \begin{array}{c} L_2 \\ > \\ < \\ L_1 \end{array} 0, \tag{5.4}$$

where the threshold of 0 corresponds to the minimum-probability-of-error criterion with equal *a priori* probabilities (van Trees, 1968, pp. 23-30). Equation 5.4 can be expanded to show the influence of the random noise n

$$\begin{aligned} Y_{opt}(\mathcal{T}) &= \ln \frac{E_n\left[p(\mathcal{T}|L_2,n)\right]}{E_n\left[p(\mathcal{T}|L_1,n)\right]} \begin{array}{c} L_2 \\ > \\ < \\ L_1 \end{array} 0, \\ &= \ln \frac{E_n\left[\prod_{i=1}^{M} p(\mathcal{T}^i|L_2,n)\right]}{E_n\left[\prod_{i=1}^{M} p(\mathcal{T}^i|L_1,n)\right]} \begin{array}{c} L_2 \\ > \\ < \\ L_1 \end{array} 0, \end{aligned} \tag{5.5}$$

where $\mathcal{T}^i$ is the set of discharge times on the i^{th} AN fiber, and Eq. 5.5 follows from the assumption of conditionally independent AN fibers given the stimulus. Unfortunately, further simplification of $Y_{opt}(\mathcal{T})$ is prohibited by the expectation over the stimulus noise in the numerator and the denominator of the likelihood ratio.

Further simplification of the optimum-processor form is possible if it is assumed that the processor *knows* the individual noise waveforms n. This assumption is unrealistic for a

human observer; however, it allows the processor to be specified in enough detail so that insight can be gained as to how a *smart* processor uses knowledge about the AN responses to each alternative in a discrimination task. This insight will be used below to derive the performance of *smart* processors that only have knowledge of the ensemble properties of the noise, rather than knowledge of the individual noise waveforms. The form of the optimum processor, given that the individual noise waveforms n are known, is specified as

$$\begin{aligned} Y_{opt|n}(\mathcal{T}) &= \ln \frac{p(\mathcal{T}|L_2,n)}{p(\mathcal{T}|L_1,n)}, \\ &= \ln \frac{\prod_{i=1}^{M} p(\mathcal{T}^i|L_2,n)}{\prod_{i=1}^{M} p(\mathcal{T}^i|L_1,n)}, \\ &= \sum_{i=1}^{M} \ln \frac{p(\mathcal{T}^i|L_2,n)}{p(\mathcal{T}^i|L_1,n)}, \\ &= \sum_{i=1}^{M} Y_{opt|n,i}(\mathcal{T}^i). \end{aligned} \tag{5.6}$$

A nonstationary Poisson process can be used to describe the stochastic properties of AN discharges (e.g., Siebert, 1970; Colburn, 1973). The joint probability density of the *unordered* discharge times from a single AN fiber can therefore be described by

$$p(\mathcal{T}^i|L,n) = \frac{\prod_{j=1}^{K_i} r_i(t_j^i|L,n)}{K_i!} \exp\left[-\int_0^T r_i(t|L,n)dt\right], \tag{5.7}$$

where $r_i(t|L,n)$ is the time-varying discharge rate of the i^{th} fiber given the tone level L and the noise waveform n (Parzen, 1962; Siebert, 1970; Snyder and Miller, 1991; Rieke *et al.*, 1997).

By substituting this expression for $p(\mathcal{T}^i|L,n)$ into Eq. 5.6, the form of the optimum processor $Y_{opt|n,i}(\mathcal{T}^i)$ can be evaluated as

$$\begin{aligned} Y_{opt|n,i}(\mathcal{T}^i) &= \ln \left\{ \frac{\left[\prod_{j=1}^{K_i} r_i(t_j^i|L_2,n) \Big/ K_i!\right] \exp\left[-\int_0^T r_i(t|L_2,n)dt\right]}{\left[\prod_{j=1}^{K_i} r_i(t_j^i|L_1,n) \Big/ K_i!\right] \exp\left[-\int_0^T r_i(t|L_1,n)dt\right]} \right\}, \\ &= \ln \left\{ \left[\prod_{j=1}^{K_i} \frac{r_i(t_j^i|L_2,n)}{r_i(t_j^i|L_1,n)}\right] \exp\left[\int_0^T [r_i(t|L_1,n) - r_i(t|L_2,n)]dt\right] \right\}, \end{aligned}$$

$$= \sum_{j=1}^{K_i} \ln \frac{r_i(t_j^i|L_2,n)}{r_i(t_j^i|L_1,n)} + \int_0^T [r_i(t|L_1,n) - r_i(t|L_2,n)]dt. \quad (5.8)$$

Note that only the first term on the right side of Eq. 5.8 depends on the random observations $\mathcal{T}^i$, while the second term depends only on the random noise waveform n. Examination of Eq. 5.8 provides insight as to how the optimum processor for the i^{th} AN fiber uses the observed discharge times to decide between the two alternatives, L_1 and L_2. The optimum processor $Y_{opt|n,i}(\mathcal{T}^i)$ evaluates each discharge time t_j^i in terms of the likelihood that it arose from L_2 relative to the likelihood that it arose from L_1 [i.e., by calculating $\ln \left[r_i(t_j^i|L_2,n) \Big/ r_i(t_j^i|L_1,n) \right]$ in the first term of Eq. 5.8]. This quantity will be greater than zero if the discharge rate at time t_j^i was higher for L_2 than for L_1. The summation of these terms for all observed discharge times thus will be likely to be greater than zero for the stimulus L_2 and less than zero for the stimulus L_1. The second term in Eq. 5.8 represents the expected difference in the number of discharges resulting from L_1 and L_2 for the noise waveform n. This term serves as a correction term based on knowledge of the individual noise waveforms. Thus, the optimum processor evaluates each discharge time based on knowledge of the discharge rates for both alternatives for *each noise waveform* n [i.e., the processor is *assumed to know* $r_i(t|L_2,n)$ and $r_i(t|L_1,n)$ for each n].

5.3.3 General form of a "smart" processor with limited knowledge

The rate waveforms $r_i(t|L_2,n)$ and $r_i(t|L_1,n)$ can be viewed as the optimum processor's *a priori* information about the AN responses for the two alternatives (i.e., the processor's expected AN responses). The assumption that a human listener *knows* the discharge rate waveforms for *each individual noise* n is unrealistic; however, it is reasonable to assume that the listener develops some *sub-optimum a priori* information about the AN responses for each alternative. The form of the optimum processor in Eq. 5.8 will be referred to as a *smart* way to process the observed AN discharge times $\mathcal{T}$ based on any two expected AN responses for the two alternatives.

Thus, performance is derived below for a general processor of the form

$$Y(\mathcal{T}) \;=\; \sum_{i=1}^{M} Y_i(\mathcal{T}^i), \tag{5.9}$$

where,

$$Y_i(\mathcal{T}^i) \;\triangleq\; \sum_{j=1}^{K_i} \ln \frac{x_i(t_j^i|L_2,n)}{x_i(t_j^i|L_1,n)} + \int_0^T x_i(t|L_1,n) - x_i(t|L_2,n)dt, \tag{5.10}$$

and $x_i(t|L_2,n)$ and $x_i(t|L_1,n)$ represent the processor's expected AN responses for the two alternatives. Note that the sum in Eq. 5.9 is an optimum way to combine information across AN fibers if the fibers are independent (see Eq. 5.6); however, in the present analysis this is generally a sub-optimum combination scheme because AN fibers may be correlated in their responses to noise stimuli.

Figure 5-1 illustrates two different processors that will be evaluated in the analysis below. The left column shows individual-noise responses to a noise alone (top row) and in the presence of three tones of different levels. These AN responses represent the knowledge used by the optimal processor for a tone detection in noise task. For example, the optimal model would rely on information related to the presence of the tone in the narrow dips of the noise response (e.g., near 13 ms and 35 ms). The presence of the tone at L_2=20 dB SPL can be seen in these dips, and thus the optimal processor would be able to detect this low-level tone even though the overall shape of the tone-in-noise and noise-alone responses are very similar.

By varying the form of the expected-response waveforms, $x_i(t|L_2,n)$ and $x_i(t|L_1,n)$, performance can be evaluated based on different assumptions of the knowledge used by the processor about the AN responses for each alternative and for each noise waveform. For example, we can remove the assumption that the processor uses the responses for individual noise waveforms by using expected responses that do not depend on individual noises (e.g., by using the average AN response across all noises, or the AN responses for tones in quiet). A more realistic processor that only uses the average properties of the noise-alone and tone-in-noise responses is illustrated in the right column of Fig. 5-1. The average tone-in-noise condition for L=20 dB SPL is indistinguishable from the noise-alone response. Only when

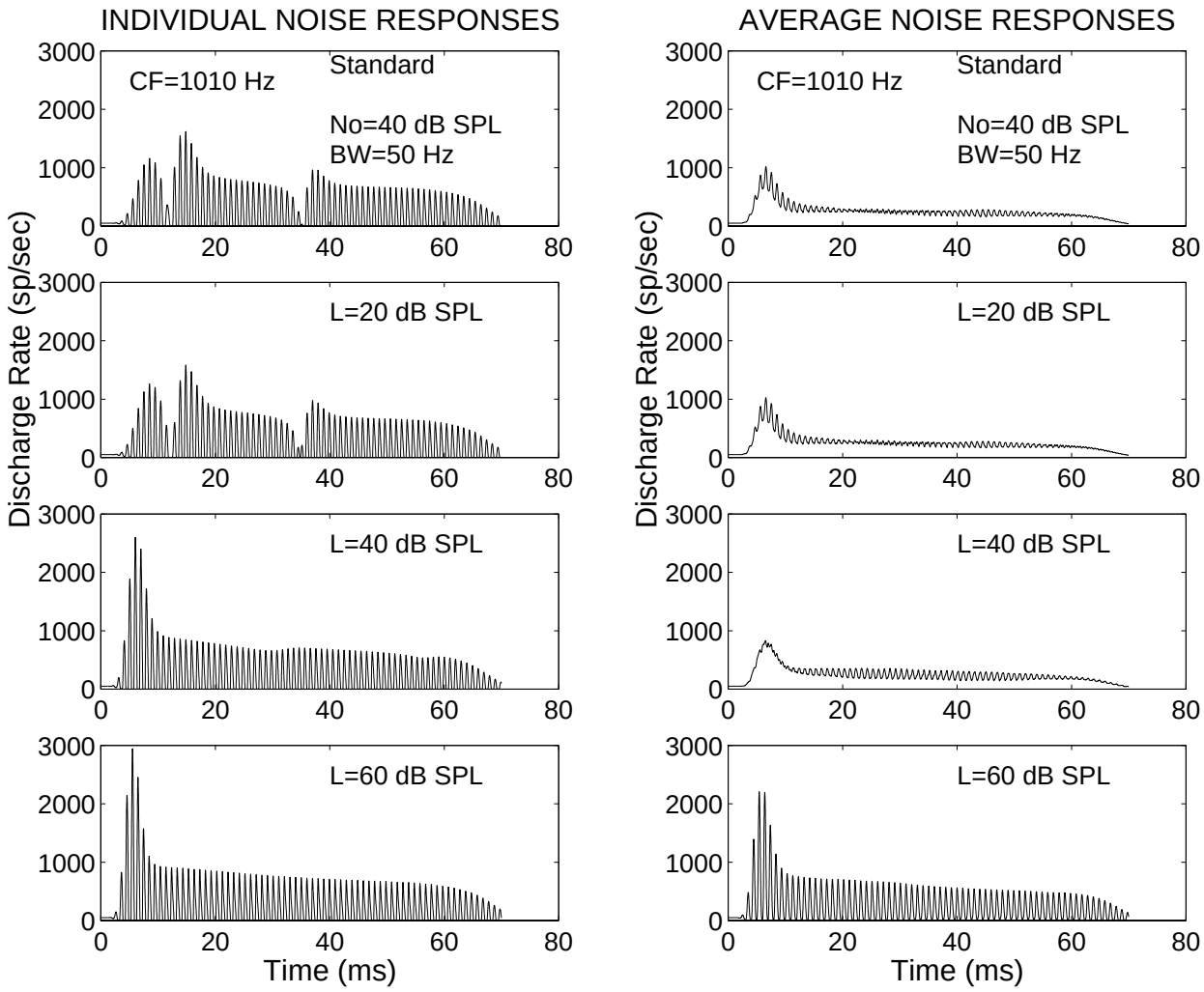

Figure 5-1: Illustration of auditory-nerve responses used in the theoretical analysis. The optimum processor uses knowledge of *individual* noise responses (shown in the left column) to detect a tone in noise. A realistic processor is assumed to use only knowledge of the *average* tone-in-noise response (shown in the right column as the average responses over 100 noises). The level of the 50-ms tone is given by L.

the tone level is raised to 40 dB SPL is a significant difference observed between signal and standard conditions, and thus the realistic processor would be expected to have a much higher threshold than the optimum processor. In summary, the performance of a specific processor that only uses knowledge of the noise ensemble can be evaluated; however, it has not been shown that this processor is optimum given that only the ensemble properties of the noise are known.

5.3.4 Performance of the general processor

In this section, several fundamental properties for decision variables of the form of interest (Eq. 5.10) are used to specify the sensitivity index Q in terms of the discharge rates $r(t)$ of each AN fiber in response to the two stimulus alternatives. The sensitivity index Q (Eq. 5.1) for the general processor (Eq. 5.9) is given by

$$\begin{aligned} Q(L_1, L_2) &= \frac{\left\{E_{n,\mathcal{T}}\left[Y\left(\mathcal{T}\right)|L_2\right] - E_{n,\mathcal{T}}\left[Y\left(\mathcal{T}\right)|L_1\right]\right\}^2}{Var_{n,\mathcal{T}}\left[Y\left(\mathcal{T}\right)|L_1\right]}, \\ &= \frac{\left(E_n\left\{E_{\mathcal{T}}\left[\sum_{i=1}^{M} Y_i\left(\mathcal{T}^i\right)\middle| L_2, n\right]\right\} - E_n\left\{E_{\mathcal{T}}\left[\sum_{i=1}^{M} Y_i\left(\mathcal{T}^i\right)\middle| L_1, n\right]\right\}\right)^2}{E_n\left\{Var_{\mathcal{T}}\left[\sum_{i=1}^{M} Y_i\left(\mathcal{T}^i\right)\middle| L_1, n\right]\right\} + Var_n\left\{E_{\mathcal{T}}\left[\sum_{i=1}^{M} Y_i\left(\mathcal{T}^i\right)\middle| L_1, n\right]\right\}}. \end{aligned} \quad (5.11)$$

Using the linearity of expected values and the assumption of conditionally independent AN fibers given the stimulus, Eq. 5.11 can be simplified further to

$$Q(L_1, L_2) = \frac{\left[E_n\left(\sum_{i=1}^{M}\left\{E_{\mathcal{T}}\left[Y_i\left(\mathcal{T}^i\right)|L_2, n\right] - E_{\mathcal{T}}\left[Y_i\left(\mathcal{T}^i\right)|L_1, n\right]\right\}\right)\right]^2}{E_n\left\{\sum_{i=1}^{M} Var_{\mathcal{T}}\left[Y_i\left(\mathcal{T}^i\right)|L_1, n\right]\right\} + Var_n\left\{\sum_{i=1}^{M} E_{\mathcal{T}}\left[Y_i\left(\mathcal{T}^i\right)|L_1, n\right]\right\}}. \quad (5.12)$$

The terms of the form $E_{\mathcal{T}}\left[Y_i\left(\mathcal{T}^i\right)|L, n\right]$ and $Var_{\mathcal{T}}\left[Y_i\left(\mathcal{T}^i\right)|L, n\right]$ in Eq. 5.12 can be simplified using a general result for decision variables of the form

$$X_i(\mathcal{T}^i) = \sum_{j=1}^{K_i} g(t_j^i), \quad (5.13)$$

where $g(t_j^i)$ is any function of the discharge time t_j^i generated from a nonstationary Poisson

process with time-varying rate function $r_i(t)$. It can be shown (e.g., Rieke *et al.*, 1997) that the expected value and variance of $X_i(\mathcal{T}^i)$ are given by

$$E_{\mathcal{T}}\left[X_i(\mathcal{T}^i)\right] = \int_0^T g(t) r_i(t) dt, \tag{5.14}$$

$$Var_{\mathcal{T}}\left[X_i(\mathcal{T}^i)\right] = \int_0^T g^2(t) r_i(t) dt. \tag{5.15}$$

Using the definition

$$g_i(t) \triangleq \ln \frac{x_i(t|L_2, n)}{x_i(t|L_1, n)}, \tag{5.16}$$

the general processor $Y_i(\mathcal{T}^i)$ for the i^{th} AN fiber given in Eq. 5.10 can be written as

$$Y_i(\mathcal{T}^i) = \sum_{j=1}^{K_i} g_i(t_j^i) + \int_0^T x_i(t|L_1, n) - x_i(t|L_2, n) dt. \tag{5.17}$$

The terms inside the summations in Eq. 5.12 can be simplified using Eqs. 5.14 and 5.15 as

$$E_{\mathcal{T}}\left[Y_i\left(\mathcal{T}^i\right)\middle| L_2, n\right] - E_{\mathcal{T}}\left[Y_i\left(\mathcal{T}^i\right)\middle| L_1, n\right] =$$

$$\begin{aligned}
&= \int_0^T g_i(t) r_i(t|L_2, n) dt + \int_0^T x_i(t|L_1, n) - x_i(t|L_2, n) dt \\
&\qquad - \left[\int_0^T g_i(t) r_i(t|L_1, n) dt + \int_0^T x_i(t|L_1, n) - x_i(t|L_2, n) dt\right], \\
&= \int_0^T g_i(t) \left[r_i(t|L_2, n) - r_i(t|L_1, n)\right] dt. \qquad (5.18)
\end{aligned}$$

$$\begin{aligned}
Var_{\mathcal{T}}\left[Y_i\left(\mathcal{T}^i\right)\middle| L_1, n\right] &= Var_{\mathcal{T}}\left[\sum_{j=1}^{K_i} g_i(t_j^i)\middle| L_1, n\right], \\
&= \int_0^T g_i^2(t) r_i(t|L_1, n) dt. \qquad (5.19)
\end{aligned}$$

$$E_{\mathcal{T}}\left[Y_i\left(\mathcal{T}^i\right)\middle|\, L_1, n\right] = \int_0^T g_i(t) r_i(t|L_1, n)dt + \int_0^T x_i(t|L_1, n) - x_i(t|L_2, n)dt. \tag{5.20}$$

Thus, the sensitivity index Q for the general processor is given by

$$\begin{aligned} Q(L_1, L_2) = & \left(E_n\left\{\sum_{i=1}^{M}\int_0^T \ln\frac{x_i(t|L_2, n)}{x_i(t|L_1, n)}\left[r_i(t|L_2, n) - r_i(t|L_1, n)\right]dt\right\}\right)^2 \Bigg/ \\ & \left(E_n\left\{\sum_{i=1}^{M}\int_0^T \left[\ln\frac{x_i(t|L_2, n)}{x_i(t|L_1, n)}\right]^2 r_i(t|L_1, n)dt\right\}\right. \\ & \left. + Var_n\left[\sum_{i=1}^{M}\int_0^T \ln\frac{x_i(t|L_2, n)}{x_i(t|L_1, n)} r_i(t|L_1, n) + x_i(t|L_1, n) - x_i(t|L_2, n)dt\right]\right). \end{aligned} \tag{5.21}$$

The numerator of Eq. 5.21 represents the square of the expected value of the difference between the decision variable for the two alternatives L_1 and L_2. The denominator of Eq. 5.21 represents the variance of the decision variable as the sum of the AN variance (first term) and the stimulus variance (second term).

Equation 5.21 represents the fundamental result of the present analysis, and can be used to analyze several types of discrimination- or detection-in-noise experiments that are commonly used in psychophysics. A *random noise* experiment uses independent noise waveforms in all observation intervals (e.g., in both intervals of a two-interval, two-alternative forced-choice experiment). A *fixed noise* experiment uses the identical noise waveform in all observation intervals, and thus the influence of stimulus variability on the observations is removed.

For a fixed-noise experiment, Eq. 5.21 can be used to calculate the sensitivity metric for each noise waveform n as

$$\begin{aligned} Q(L_1, L_2, n) = & \left\{\sum_{i=1}^{M}\int_0^T \ln\frac{x_i(t|L_2, n)}{x_i(t|L_1, n)}\left[r_i(t|L_2, n) - r_i(t|L_1, n)\right]dt\right\}^2 \Bigg/ \\ & \left\{\sum_{i=1}^{M}\int_0^T \left[\ln\frac{x_i(t|L_2, n)}{x_i(t|L_1, n)}\right]^2 r_i(t|L_1, n)dt\right\}. \end{aligned} \tag{5.22}$$

The average sensitivity across noise waveforms is then given by

$$Q(L_1, L_2) = E_n \left\{ Q(L_1, L_2, n) \right\}. \tag{5.23}$$

Psychophysical performance has been shown to vary from noise waveform to noise waveform in some conditions (e.g., Gilkey *et al.*, 1985; Isabelle and Colburn, 1991; Siegel and Colburn, 1989), and thus Eq. 5.22 could be used to evaluate the peripheral contribution to the variation in performance across noise waveforms.

5.3.5 Simplifications for suprathreshold conditions

The sensitivity metric Q derived in the previous section (Eq. 5.21) can be simplified for suprathreshold discrimination conditions (i.e., most discrimination experiments excluding detection). In suprathreshold conditions, it is reasonable to assume that $r_i(t|L,n)$ and $x_i(t|L,n)$ vary linearly and slowly over the incremental range from L to $L+\Delta L$, where ΔL is small (Colburn, 1981). Thus,

$$r_i(t|L+\Delta L, n) \simeq r_i(t|L,n) + \dot{r}_i(t|L,n)\Delta L, \tag{5.24}$$

where $\dot{r}_i(t|L,n) \triangleq \frac{\partial}{\partial L} r_i(t|L,n)$, and similarly for $x_i(t|L,n)$. Several terms in Eq. 5.21 can be simplified using this assumption including

$$\begin{aligned} \ln \frac{x_i(t|L+\Delta L, n)}{x_i(t|L,n)} &\simeq \ln\left(1 + \frac{\dot{x}_i(t|L,n)\Delta L}{x_i(t|L,n)}\right), \\ &\simeq \frac{\dot{x}_i(t|L,n)\Delta L}{x_i(t|L,n)}, \end{aligned} \tag{5.25}$$

where the last step used the approximation $\ln(1+x) \simeq x$ for small x. Thus, the sensitivity index Q in Eq. 5.21 is simplified in the suprathreshold case as

$$\begin{aligned} Q\left(L, \Delta L\right) = &\left(E_n \left\{ \sum_{i=1}^{M} \int_0^T \left[\frac{\dot{x}_i(t|L,n)\Delta L}{x_i(t|L,n)} \right] \dot{r}_i(t|L,n)\Delta L dt \right\} \right)^2 \Bigg/ \\ &\left(E_n \left\{ \sum_{i=1}^{M} \int_0^T \left[\frac{\dot{x}_i(t|L,n)\Delta L}{x_i(t|L,n)} \right]^2 r_i(t|L,n) dt \right\} + \right. \end{aligned} \tag{5.26}$$

$$+Var_n\left[\sum_{i=1}^{M}\int_0^T \frac{\dot{x}_i(t|L,n)\Delta L}{x_i(t|L,n)} r_i(t|L,n) - \dot{x}_i(t|L,n)\Delta L dt\right]\Bigg).$$

Characterization of processor performance can be simplified further by defining the *normalized sensitivity* $\delta'(L)$ of the processor (see Durlach and Braida, 1969; Braida and Durlach, 1988; Chapter 2) to be given by

$$[\delta'(L)]^2 \triangleq \frac{Q(L,\Delta L)}{(\Delta L)^2}. \tag{5.27}$$

Thus, the normalized-sensitivity squared $[\delta'(L)]^2$ of the general processor is given by

$$\begin{aligned}[\delta'(L)]^2 \;=\; & \left(E_n\left\{\sum_{i=1}^{M}\int_0^T\left[\frac{\dot{x}_i(t|L,n)}{x_i(t|L,n)}\right]\dot{r}_i(t|L,n)dt\right\}\right)^2 \Bigg/ \\ & \left[E_n\left\{\sum_{i=1}^{M}\int_0^T\left[\frac{\dot{x}_i(t|L,n)}{x_i(t|L,n)}\right]^2 r_i(t|L,n)dt\right\}\right. \\ & \left. +Var_n\left(\sum_{i=1}^{M}\left\{\int_0^T\left[\frac{\dot{x}_i(t|L,n)}{x_i(t|L,n)}\right] r_i(t|L,n) - \dot{x}_i(t|L,n)dt\right\}\right)\right].\end{aligned} \tag{5.28}$$

The partial derivatives with respect to level in Eq. 5.28 can be approximated by taking the difference between the AN discharge rate waveforms $r(t)$ at two incrementally different levels and dividing by the level difference (see Chapter 2).

Theoretical predictions of normalized sensitivity can be related to psychophysical performance in terms of ΔL_{JND}, which is defined as corresponding to $Q(L,\Delta L_{JND}) = 1$, and is therefore related to $\delta'(L)$ by

$$\Delta L_{JND}(L) = [\delta'(L)]^{-1}. \tag{5.29}$$

5.3.6 Specific processors

Several types of processors that are evaluated below are described in this section. Expressions for the normalized sensitivity of these processors in the suprathreshold conditions are also given to provide some insight into their performance.

Optimum processor

It has been shown above that the optimum processor given knowledge of each noise waveform corresponds to using expected responses that are equal to the rate waveforms for each noise waveform for each alternative [i.e., $x_i(t|L_1,n) = r_i(t|L_1,n)$ and $x_i(t|L_2,n) = r_i(t|L_2,n)$]. Evaluating Eq. 5.28 with the optimum expected responses, the normalized sensitivity squared for the suprathreshold conditions is given by

$$\begin{aligned}
[\delta'(L)]^2 &= \left(E_n\left\{\sum_{i=1}^{M}\int_0^T \frac{1}{r_i(t|L,n)}\left[\dot{r_i}(t|L,n)\right]^2 dt\right\}\right)^2 \Bigg/ \\
&\qquad \left(E_n\left\{\sum_{i=1}^{M}\int_0^T \frac{1}{r_i(t|L,n)}\left[\dot{r_i}(t|L,n)\right]^2 dt\right\}\right. \\
&\qquad\qquad \left. +Var_n\left\{\sum_{i=1}^{M}\left[\int_0^T \dot{r_i}(t|L,n) - \dot{r_i}(t|L,n)dt\right]\right\}\right), \\
&= E_n\left\{\sum_{i=1}^{M}\int_0^T \frac{1}{r_i(t|L,n)}\left[\dot{r_i}(t|L,n)\right]^2 dt\right\}, \\
&= E_n\left\{\sum_{i=1}^{M}\left[\delta_i'(L,n)\right]^2\right\}, \\
&= E_n\left\{\left[\delta'(L,n)\right]^2\right\}, \qquad (5.30)
\end{aligned}$$

where $[\delta_i']^2$ is the normalized sensitivity squared of the i^{th} AN fiber, and $[\delta']^2$ is the total normalized sensitivity squared for each noise based on the population of M AN fibers. The identical result can also be derived from the Cramér-Rao bound with a fixed noise waveform n (Siebert, 1970; Chapter 2). Thus, the optimum processor is unaffected by the random variation across noises (i.e., optimum performance in a random-noise experiment is the same as optimum performance in a fixed-noise experiment). This finding results from the processor using the AN responses to each noise individually. In the general discrimination experiment, the sensitivity for the optimum processor is given by

$$Q(L_1,L_2) = E_n\left\{\sum_{i=1}^{M} Q_i(L_1,L_2,n)\right\}, \qquad (5.31)$$

where Q_i is the sensitivity of the i^{th} AN fiber.

Noise-independent processor

The assumption that a listener uses the discharge rate waveform for each individual noise is unrealistic. In fact, it is unlikely that listeners changes their processors from noise to noise in a single experiment. The general equation for normalized sensitivity (Eq. 5.28) can be simplified based on the assumption that the processor uses the same expected responses for all noise waveforms [i.e., $x_i(t|L_1)$ and $x_i(t|L_2)$ are independent of n]. In this case, the *normalized sensitivity* squared is given by

$$[\delta'(L)]^2 = \left\{\sum_{i=1}^{M}\int_0^T \left[\frac{\dot{x}_i(t|L)}{x_i(t|L)}\right]\overline{\dot{r}_i}(t|L)dt\right\}^2 \Bigg/ \left(\sum_{i=1}^{M}\int_0^T \left[\frac{\dot{x}_i(t|L)}{x_i(t|L)}\right]^2 \overline{r_i}(t|L)dt + Var_n\left\{\sum_{i=1}^{M}\int_0^T \left[\frac{\dot{x}_i(t|L)}{x_i(t|L)}\right] r_i(t|L,n)dt\right\}\right), \quad (5.32)$$

where $\overline{r_i}(t|L) \triangleq E_n[r_i(t|L,n)]$ is the average rate waveform across all noises. For the fixed-noise experiment, the normalized sensitivity squared for each noise is given by

$$[\delta'(L,n)]^2 = \left\{\sum_{i=1}^{M}\int_0^T \left[\frac{\dot{x}_i(t|L)}{x_i(t|L)}\right]\dot{r}_i(t|L,n)dt\right\}^2 \Bigg/ \left\{\sum_{i=1}^{M}\int_0^T \left[\frac{\dot{x}_i(t|L)}{x_i(t|L)}\right]^2 r_i(t|L,n)dt\right\}. \quad (5.33)$$

A specific noise-independent processor that is evaluated below is based on the assumption that the processor uses the average rate waveform for each alternative across all noise waveforms [i.e., that the processor uses $x_i(t|L_1) = \overline{r_i}(t|L_1)$ and $x_i(t|L_2) = \overline{r_i}(t|L_2)$]. [1] This assumption states that the processor operates based on the common influence of the tone on the rate waveforms, but does not use the individual rate waveforms for the tone in each noise. In this case, the normalized sensitivity squared is given by

$$[\delta'(L)]^2 = \left\{\sum_{i=1}^{M}\int_0^T \frac{1}{\overline{r_i}(t|L)}\left[\overline{\dot{r}_i}(t|L)\right]^2 dt\right\}^2 \Bigg/ \left(\sum_{i=1}^{M}\int_0^T \frac{1}{\overline{r_i}(t|L)}\left[\overline{\dot{r}_i}(t|L)\right]^2 dt + Var_n\left\{\sum_{i=1}^{M}\int_0^T \left[\frac{\overline{\dot{r}_i}(t|L)}{\overline{r_i}(t|L)}\right] r_i(t|L,n)dt\right\}\right). \quad (5.34)$$

[1] Pilot studies with other noise-independent processors, such as tone alone or supra-threshold tone in noise, indicated that the primary results do not depend strongly on the exact choice of $x_i(t|L_1)$ and $x_i(t|L_2)$.

Time-independent processor

Another class of sub-optimum processor that is evaluated below is one in which the processor is assumed not to make use of any temporal information in the AN-discharge patterns. This assumption corresponds to the use of time-independent expected responses [$x_i(L_1, n)$ and $x_i(L_2, n)$]. In this case, the general processor (Eqs. 5.9 and 5.10) is given by

$$\begin{aligned} Y_i(\mathcal{T}^i) &\triangleq \sum_{j=1}^{K_i} \ln \frac{x_i(L_2,n)}{x_i(L_1,n)} + \int_0^T x_i(L_1,n) - x_i(L_2,n)dt, \\ &= \left[\ln \frac{x_i(L_2,n)}{x_i(L_1,n)}\right] K_i + T\left[x_i(L_1,n) - x_i(L_2,n)\right], \end{aligned} \tag{5.35}$$

and can be seen to only depend on the observed number of discharges K_i on each AN fiber. For any time-independent expected response, the *normalized sensitivity* squared is given by,

$$\begin{aligned} \left[\delta'(L)\right]^2 &= \left(E_n\left\{\sum_{i=1}^{M}\left[\frac{\dot{x}_i(L,n)}{x_i(L,n)}\right]\dot{\widetilde{r}}_i(L,n)\right\}\right)^2 \Bigg/ \\ &\quad \left[\frac{1}{T}E_n\left\{\sum_{i=1}^{M}\left[\frac{\dot{x}_i(L,n)}{x_i(L,n)}\right]^2 \widetilde{r}_i(L,n)\right\}\right. \\ &\quad \left. + Var_n\left(\sum_{i=1}^{M}\left\{\left[\frac{\dot{x}_i(L,n)}{x_i(L,n)}\right]\widetilde{r}_i(L,n) - \dot{\widetilde{x}}_i(L,n)\right\}\right)\right], \end{aligned} \tag{5.36}$$

where $\widetilde{r}_i(L) \triangleq \frac{1}{T}\int_0^T r_i(t|L)dt$ is the average discharge rate over time. For the fixed-noise case, the normalized sensitivity for each noise is given by

$$\left[\delta'(L,n)\right]^2 = T\left\{\sum_{i=1}^{M}\left[\frac{\dot{x}_i(L,n)}{x_i(L,n)}\right]\dot{\widetilde{r}}_i(L,n)\right\}^2 \Bigg/ \left\{\sum_{i=1}^{M}\left[\frac{\dot{x}_i(L,n)}{x_i(L,n)}\right]^2 \widetilde{r}_i(L,n)\right\}. \tag{5.37}$$

5.4 Application of Theory

5.4.1 Use of computational auditory-nerve models

The use of computational AN models, which are needed to describe the complex nonlinear responses of AN fibers to noise stimuli, precludes the computation of discharge-rate waveforms $r(t)$ for all M (roughly 30,000) AN fibers in Eq. 5.21. To reduce computation time, it

is assumed that it is roughly equivalent to compute the rate waveforms for a limited number (N_{CF}) of different model fibers, and to assume that there are M_i statistically identical, conditionally independent AN fibers represented by each model fiber, such that $M = \sum_{i=1}^{N_{CF}} M_i$. [2] Based on this assumption, Eq. 5.21 can be written as

$$\begin{aligned} Q(L_1, L_2) &= \left(E_n \left\{ \sum_{i=1}^{N_{CF}} M_i \int_0^T \ln \frac{x_i(t|L_2,n)}{x_i(t|L_1,n)} \left[r_i(t|L_2,n) - r_i(t|L_1,n) \right] dt \right\} \right)^2 \Bigg/ \\ &\quad \left(E_n \left\{ \sum_{i=1}^{N_{CF}} M_i \int_0^T \left[\ln \frac{x_i(t|L_2,n)}{x_i(t|L_1,n)} \right]^2 r_i(t|L_1,n) dt \right\} \right. \\ &\quad \left. + Var_n \left[\sum_{i=1}^{N_{CF}} M_i \int_0^T \ln \frac{x_i(t|L_2,n)}{x_i(t|L_1,n)} r_i(t|L_1,n) + x_i(t|L_1,n) - x_i(t|L_2,n) dt \right] \right). \end{aligned} \tag{5.38}$$

For the simple case in which $M_i = M_{CF}$ for each i (a typical assumption when each model fiber represents a different CF), Eq. 5.38 simplifies to

$$\begin{aligned} Q(L_1, L_2) &= \left(E_n \left\{ \sum_{i=1}^{N_{CF}} \int_0^T \ln \frac{x_i(t|L_2,n)}{x_i(t|L_1,n)} \left[r_i(t|L_2,n) - r_i(t|L_1,n) \right] dt \right\} \right)^2 \Bigg/ \\ &\quad \left(\frac{1}{M_{CF}} E_n \left\{ \sum_{i=1}^{N_{CF}} \int_0^T \left[\ln \frac{x_i(t|L_2,n)}{x_i(t|L_1,n)} \right]^2 r_i(t|L_1,n) dt \right\} \right. \\ &\quad \left. + Var_n \left[\sum_{i=1}^{N_{CF}} \int_0^T \ln \frac{x_i(t|L_2,n)}{x_i(t|L_1,n)} r_i(t|L_1,n) + x_i(t|L_1,n) - x_i(t|L_2,n) dt \right] \right), \end{aligned} \tag{5.39}$$

for the random-noise experiment. In a fixed-noise experiment, the sensitivity is given by

$$\begin{aligned} Q(L_1, L_2, n) &= M_{CF} \left\{ \sum_{i=1}^{N_{CF}} \int_0^T \ln \frac{x_i(t|L_2,n)}{x_i(t|L_1,n)} \left[r_i(t|L_2,n) - r_i(t|L_1,n) \right] dt \right\}^2 \Bigg/ \\ &\quad \left\{ \sum_{i=1}^{N_{CF}} \int_0^T \left[\ln \frac{x_i(t|L_2,n)}{x_i(t|L_1,n)} \right]^2 r_i(t|L_1,n) dt \right\}, \end{aligned} \tag{5.40}$$

[2] Note that this is also an interesting theoretical view of the population of AN fibers within the auditory system. One can hypothesize that the auditory system benefits in two ways from the distribution of AN fibers: (1) multiple AN fibers with different CF's, thresholds, and dynamic ranges allow the auditory system to encode sound over a wide range of frequencies and levels, and (2) multiple conditionally independent AN fibers with identical characteristics allow the auditory system to reduce the limitations imposed by physiological response variability through redundancy.

and the benefit from having M_{CF} statistically identical AN fibers at each CF is to increase the sensitivity Q by a factor of M_{CF}. However, for the random-noise experiment (Eq. 5.39), the benefit from having M_{CF} statistically identical AN fibers at each CF can range from a factor of M_{CF} [when the AN variance dominates, i.e., $\frac{1}{M_{CF}} E_n \{\cdots\} \gg Var_n (\cdots)$] to a factor of 1 [when the stimulus variance dominates, i.e., $\frac{1}{M_{CF}} E_n \{\cdots\} \ll Var_n (\cdots)$]. Note that the AN (internal) variance is reduced by a factor of $\frac{1}{M_{CF}}$ by combining over M_{CF} fibers, while no reduction in the stimulus (external) variance occurs from combining over multiple fibers.

An illustration of the use of a computational AN model to make predictions of sensitivity for several of the models and experiments described above is shown in Fig. 5-2. The nonlinear computational AN model used in the present study is described in Appendix A; however, the basic implementation issues that are important for the present predictions are discussed below. The predictions were made with 60 high-spontaneous-rate (HSR, 60 sp/sec) model fibers that had characteristic frequencies (CF) ranging from 200 to 20,000 Hz.

Fig. 5-2 shows sensitivity Q for the detection of a 1049-Hz tone, which corresponds to the model CF that was closest to 1 kHz. Sensitivity is shown for detection of the tone in quiet and in the presence of noise maskers with 100-Hz and 10-kHz bandwidths. The bandlimited maskers were centered geometrically at the tone frequency, and were presented with a spectrum level of N_0=20 dB SPL and a duration of T=200 ms. The tone was presented at a level of L =40 dB SPL, which roughly corresponds to human threshold ($Q = 1$) for both masker bandwidths (Fletcher, 1940; Bos and de Boer, 1966), and a duration of T =200 ms.

The left column of Fig. 5-2 shows sensitivity based on $M_{CF} = 1$ AN fiber at each model CF, while the right column corresponds to $M_{CF} = 200$. The human AN contains about 30,000 AN fibers (Rasmussen, 1940), with CF's distributed close to logarithmically from 20-20,000 Hz (Greenwood, 1990). Liberman (1978) has shown in cat that the HSR population makes up roughly 60% of the total AN-fiber population. Thus, $M_{CF} = 200$ statistically identical AN fibers for each of the 60 model CF's ranging from 200-20000 Hz was used to represent the total population of HSR fibers in the human AN, given that CF's below 200 Hz are not likely to respond to a 1-kHz, 40-dB SPL tone. Predictions are compared between the all-information model (based on all AN discharge times) and the rate-place model (based only on the set of AN discharge counts). For both maskers, predictions are made for the

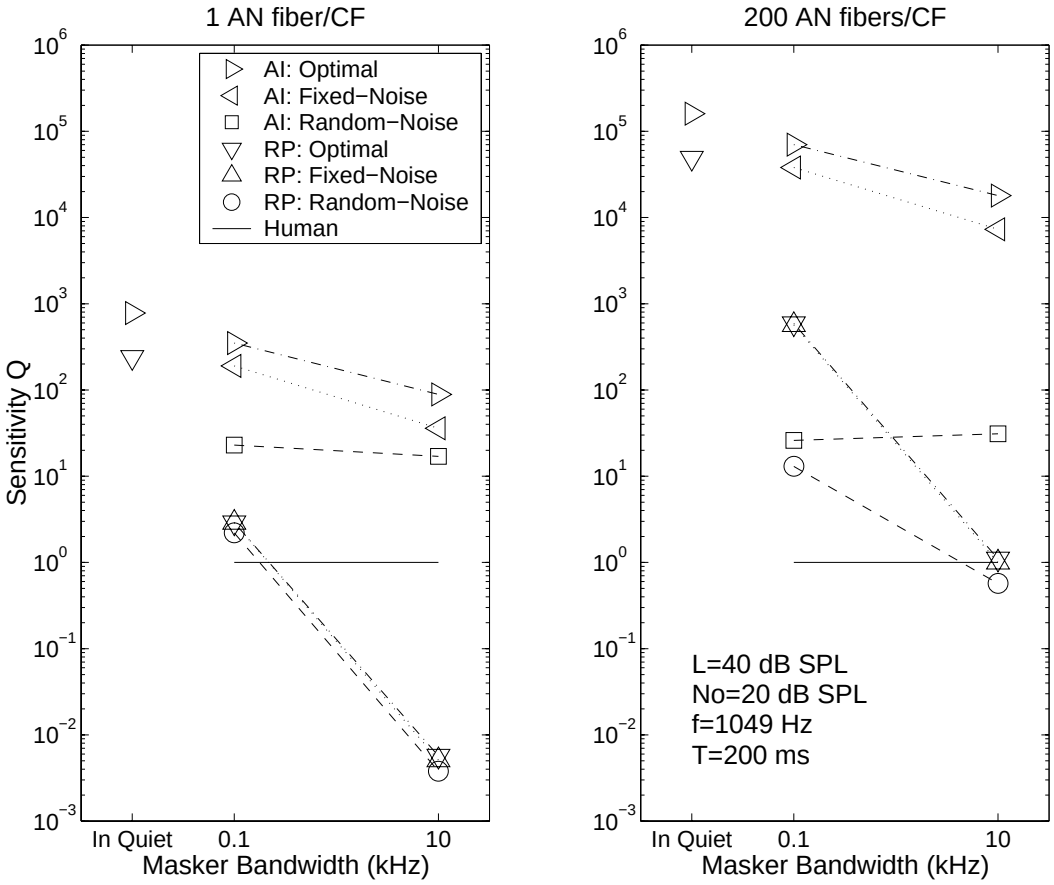

Figure 5-2: Sensitivity Q for detection of a 1049-Hz tone in quiet and in noise maskers with narrow and broad bandwidths predicted using a computational AN model. The left panel represents predictions based on M_{CF}=1 high-spontaneous-rate (HSR: 60 sp/sec) AN fiber for each model characteristic frequency (CF), while the right panel shows predictions for M_{CF}=200 statistically identical AN fibers for each CF, which represents the total HSR fiber population (see text). Predictions are compared based on the set of AN discharge times (all-information: AI) and based on the set of AN discharge counts (rate-place: RP). Masking predictions are made for the realistic processor (Eqs. 5.9 and 5.10) and for the optimum processor (Eqs. 5.6 and 5.8) based on 50 noise waveforms. The realistic-processor predictions are made for a random-noise experiment (Eq. 5.39) and a fixed-noise experiment (Eqs. 5.40 and 5.23). The tone was 1049 Hz, 40 dB SPL, 200 ms (20-ms rise/fall), which corresponds roughly to human threshold for both bandwidths (Q = 1, solid line). The AN response extended 25 ms after the stimulus offset to include the transient response (see Chapter 2). The noise maskers had a spectrum level of 40 dB SPL, and a duration equal to the tone.

realistic processor (Eqs. 5.9 and 5.10) that only uses the average tone-in-noise response, and for the optimum processor (Eqs. 5.6 and 5.8) that uses each noise waveform individually. Predictions for the realistic processor are made for the random-noise experiment (Eq. 5.39) and for the fixed-noise experiment (Eqs. 5.40 and 5.23).

The predictions in Fig. 5-2 demonstrate several effects that are important for understanding the effect of noise maskers on psychophysical performance limits. In all cases, the sensitivity Q decreased when a noise masker was presented. The sensitivity for the all-information model was always higher than the corresponding rate-place sensitivity. Optimal sensitivity was always greater than or equal to fixed-noise sensitivity, which was greater than or equal to random-noise sensitivity.

For the $M_{CF} = 1$ predictions shown in the left column, the optimal and fixed-noise sensitivities were reduced for the 10-kHz masker compared with the 100-Hz masker, especially for the rate-place model. For the all-information predictions, the reduction in sensitivity that results from using random- instead of fixed-noise maskers was much larger for the narrowband masker than for the broadband masker. This result is consistent with the large random fluctuations associated with narrowband noise stimuli. The effect of random-noise variation was minimal for both masker bandwidths for the rate-place model. The random-noise sensitivity for the all-information model was essentially the same for both masker bandwidths, consistent with equal human detection performance for both bandwidths for this tone frequency (Fletcher, 1940; Bos and de Boer, 1966). In contrast, the rate-place sensitivity was greatly reduced for the 10-kHz masker compared with the 100-Hz masker. Both the all-information and rate-place models demonstrated greater sensitivity than humans for the 100-Hz random noise, while only the all-information sensitivity was greater than human sensitivity for the 10-kHz random noise.

The predictions with $M_{CF} = 200$ statistically identical AN fibers for each model CF represent the total HSR population of AN fibers. For independent AN fibers (such as are assumed in the optimum model and in the fixed-noise and in-quiet experiments), the benefit from having multiple AN fibers at each CF is to increase the sensitivity Q by a factor of M_{CF}=200 (Eq. 5.40). In contrast, the benefit in the random-noise experiment is less than a factor of M_{CF}=200 (Eq. 5.39). For the all-information model, the benefit is less

than a factor of two for both masker bandwidths because the external stimulus variance is dominating the denominator of Eq. 5.39. The benefit of M_{CF}=200 AN fibers for each model CF in the rate-place model is a factor of six for the narrowband masker, and is a factor of 150 for the broadband masker. This difference in the benefit of adding multiple AN fibers is consistent with a larger external stimulus variability for the narrowband masker than for the broadband masker. Thus, the use of random-noise maskers has a strong and complex effect on the benefit that can be achieved by combining information across AN fibers. The reduced benefit results from the correlation of AN fibers due to the common influence of the random stimulus noise.[3]

5.4.2 Distribution of information across characteristic frequency

The sensitivity evaluated for each model CF provides a rough estimate of the distribution of information across the population of AN fibers. Note that this may not represent the total amount of information, depending on the correlation across CF's. The total information available is not simply the sum of the information across CF's due to interactions between CF's resulting from the common source of variability from the noise stimulus. This distribution of information simply represents the amount of information that each CF in isolation could contribute to discrimination.

The sensitivity for the i^{th} CF is given by

$$\begin{aligned} Q_i(L_1,L_2) &= \left(E_n \left\{ \int_0^T \ln \frac{x_i(t|L_2,n)}{x_i(t|L_1,n)} \left[r_i(t|L_2,n) - r_i(t|L_1,n) \right] dt \right\} \right)^2 \Bigg/ \\ &\quad \left(\frac{1}{M_i} E_n \left\{ \int_0^T \left[\ln \frac{x_i(t|L_2,n)}{x_i(t|L_1,n)} \right]^2 r_i(t|L_1,n) dt \right\} + \right. \end{aligned} \tag{5.41}$$

[3]Several pilot simulations were performed to investigate whether the small benefit from combining over multiple AN fibers resulted from the assumption that the total HSR population could be represented by 200 statistically identical AN fibers at each model CF. In a suprathreshold discrimination in noise experiment, model CF's were spaced 0.1 mm apart on the BM to represent single inner-hair-cell (IHC) spacing, and $M_{CF} = 10$ AN fibers were assumed to be represented by each model CF. The predicted sensitivity for these conditions was very similar to the sensitivity predicted based on M_{CF}=200 AN fibers for each of the 60 model CF's. Thus, the representation of the total AN population by 60 model CF's spaced from 200-20,000 Hz appears to be a reasonable approximation given the computational time required to perform the accurate 1-IHC spacing of model CF's.

$$+Var_n\left[\int_0^T \ln\frac{x_i(t|L_2,n)}{x_i(t|L_1,n)} r_i(t|L_1,n) + x_i(t|L_1,n) - x_i(t|L_2,n)dt\right]\Bigg).$$

In the fixed-noise case, the sensitivity of the i^{th} CF and n^{th} noise waveform is given by

$$\begin{aligned} Q_i(L_1,L_2,n) &= M_i\left\{\int_0^T \ln\frac{x_i(t|L_2,n)}{x_i(t|L_1,n)}\left[r_i(t|L_2,n) - r_i(t|L_1,n)\right]dt\right\}^2 \Bigg/ \\ &\qquad \left\{\int_0^T \left[\ln\frac{x_i(t|L_2,n)}{x_i(t|L_1,n)}\right]^2 r_i(t|L_1,n)dt\right\}. \end{aligned} \tag{5.42}$$

Note that in a fixed-noise experiment, the total sensitivity for the general processor given in Eq. 5.40 is *not* equal to the sum of the individual sensitivities on each fiber given in Eq. 5.42, as might be expected for an experiment in which the stimulus variability was removed. This is due to the fact that the average discharge rates in noise that are used by the processor provide weights to apply to the temporal patterns of discharges, as well as to the contributions from each fiber. For general expected-response waveforms, the weights assigned to each CF may not be optimum, and therefore the familiar result that the sensitivities from independent sources can be added does not hold in this case because the combination of information across independent sources is non-optimal.

The distribution of information across the population of AN fibers can be used to evaluate whether only the CF's near the frequency of a tone contain useful information for detection of the tone in noise, which is a common assumption in interpreting psychophysical masking experiments (Moore, 1995a). Figure 5-3 shows the sensitivity Q for individual CF's plotted as a function of CF for the rate-place (top panel) and all-information (bottom panel) models. Sensitivity of the realistic processor for detection in quiet (solid curve) is compared to detection in a narrowband (100 Hz) random-noise masker and to detection in a broadband (10 kHz) noise masker based on Eq. 5.41. The left column shows predictions based on M_{CF}=1 AN fiber for each model CF, while the right column represents M_{CF}=200.

For detection in quiet, a broad range of CF's contribute significant information, with the CF equal to the tone frequency contributing the most significant information. The sensitivity curve is shifted upward by a factor of 200 for the M_{CF}=200 predictions, as expected. When

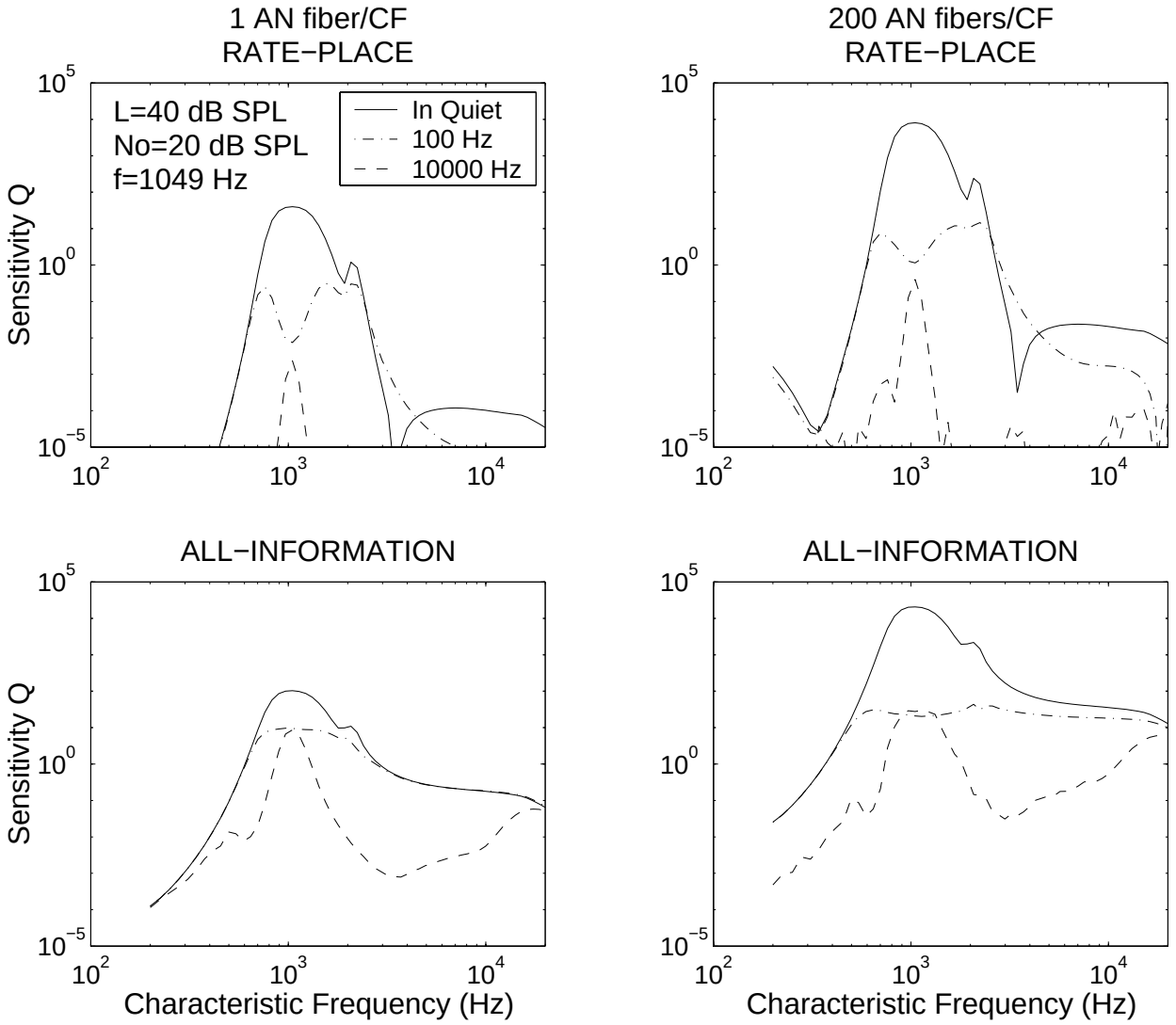

Figure 5-3: Sensitivity Q for detection of a tone in quiet and in random-noise maskers with bandwidths of 100 Hz and 10 kHz is plotted as a function of CF for the rate-place (top panel) and all-information (bottom panel) models based on high-spontaneous-rate (60 sp/sec) AN fibers. The tone was 1049 Hz, 40 dB SPL, 200 ms (20-ms rise/fall), which corresponds roughly to human threshold for both bandwidths ($Q = 1$, solid line). The noise maskers had a spectrum level of 40 dB SPL, and a duration equal to the tone. Masked detection predictions are for the realistic processor that uses knowledge of the average tone-in-noise responses over 50 noise waveforms.

the random narrowband of noise is added to the tone, the rate-place information in the CF's near the tone frequency is most degraded due to saturation of the rate-level curve, while the information at CF's well above and below the tone frequency are typically unaffected. The on-frequency CF is operating on a very shallow part of the rate-level curve in the presence of the noise masker, while the CF's near the edges of the noise are operating on a steeper part of the rate-level curve, and thus provide more information. Thus, the situation arises in which there is more rate-place information off-frequency than on-frequency for detection of a tone in a narrowband of noise.

The all-information predictions for $M_{CF} = 1$ are similar; however, there is less of a detrimental effect of saturation at CF, and thus the all-information profile is much flatter than the rate-place information profile for detection in narrowband noise. The all-information profile for M_{CF}=200 is much broader than for M_{CF}=1, with significant information contributed by CF's from 600 Hz to 20 kHz. The dramatic change in shape of the all-information profile when multiple statistically identical AN fibers are added for each CF results from variation in the correlation between same-CF fibers as a function of CF. For CF's near the tone frequency, the noise masker causes the same-CF fibers to be highly correlated, and thus there is only a small benefit of adding additional fibers. For CF's well away from the tone frequency, the noise masker has a small effect and the AN randomness has a large effect, and thus there is a large benefit from adding multiple same-CF AN fibers.

There is a small amount of rate-place information at the CF equal to the tone frequency that remains in the presence of the broadband noise masker because the rate-level function does not completely saturate in the presence of the noise. For the broadband noise, the information off-frequency is greatly reduced, and thus the information at the CF equal to the tone frequency is most dominant. The situation for the all-information model is similar, with a slightly broader range of CF's near the tone frequency contributing significant information. Thus, for detection in broadband noise, detection is based primarily on information near the CF equal to the tone-frequency (i.e., on-frequency information).

5.4.3 Analyzing temporal information

Detection of tones in noise is often assumed to depend only on the long-term power spectra of the signal and masker (Moore, 1995a). However, Fig. 5-2 demonstrated that temporal information is available for the detection of low-frequency tones in noise. Similar predictions for 8-kHz tones (not shown) indicate that there is also significant temporal information available for the detection of high-frequency tones in noise, despite the lack of strong phase-locking to the temporal fine-time structure above 2 kHz in the AN (Johnson, 1980; see Appendix A). In order to visualize the type of temporal information available for a single CF, Eq. 5.41 can be evaluated as a function of time as

$$\begin{aligned} Q_i(t, L_1, L_2) \;=\; & \left(E_n \left\{ \ln \frac{x_i(t|L_2, n)}{x_i(t|L_1, n)} \left[r_i(t|L_2, n) - r_i(t|L_1, n) \right] \right\} \right)^2 \Bigg/ \\ & \left(\frac{1}{M_i} E_n \left\{ \left[\ln \frac{x_i(t|L_2, n)}{x_i(t|L_1, n)} \right]^2 r_i(t|L_1, n) \right\} \right. \\ & \left. + Var_n \left[\ln \frac{x_i(t|L_2, n)}{x_i(t|L_1, n)} r_i(t|L_1, n) + x_i(t|L_1, n) - x_i(t|L_2, n) \right] \right). \end{aligned} \tag{5.43}$$

Equation 5.43 represents the sensitivity $Q(t, L_1, L_2)$ of a processor that relies solely on a very small temporal window (e.g., a single sample point in this illustration). However, the total sensitivity based on the entire stimulus duration is not simply the integral of $Q(t, L_1, L_2)$ across time (Eq. 5.42) due to potential correlation between temporal segments.

The type of temporal information available to the realistic processor for the detection of a low- or high-frequency tone in random noise is illustrated in Fig. 5-4. The average-noise-alone response used by the realistic processor is shown in the top panel, while the average-tone-in-noise response is shown in the middle panel. The resulting temporal information available is shown in the bottom panel by plotting sensitivity Q as a function of time.

The tone level shown in Fig. 5-4 is L=60 dB SPL, while the noise masker spectrum level is No=40 dB SPL. Thus, the 1010-Hz tone condition is near human threshold (Bos and de Boer, 1966). A clear difference is seen between the average noise-alone and tone-in-noise responses. The temporal information that is available for the low-frequency tone follows the fine-time structure of the tone response due to strong AN phase-locking below 2 kHz. Both

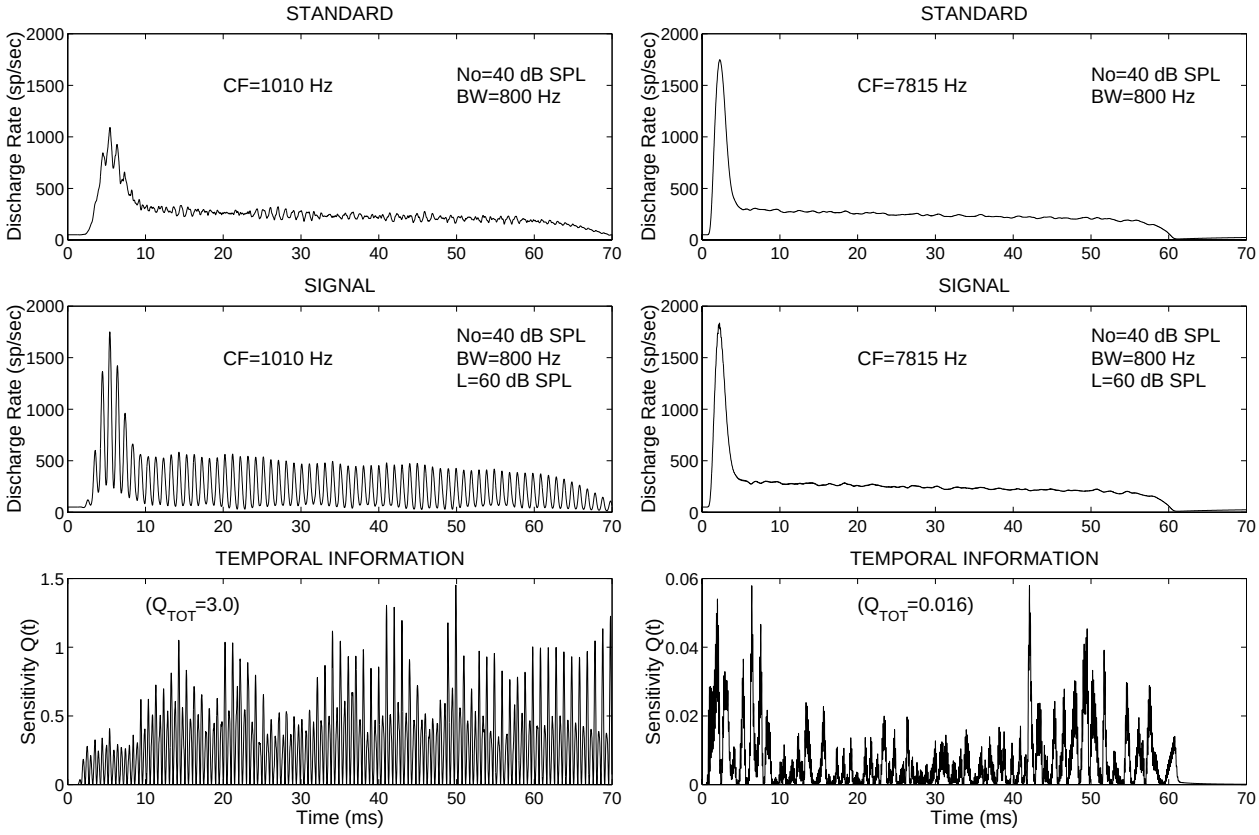

Figure 5-4: An illustration of the temporal information available to the realistic processor. Average noise-alone (top panel) and tone-in-noise (middle panel) responses (based on 100 noise waveforms) are shown for detection of a low-frequency (1010 Hz,left column) and high-frequency (7815 Hz, right column), 60-dB SPL tone presented at CF. The temporal information for the high-spontaneous-rate (60 sp/sec) fiber is shown in the bottom panel by plotting sensitivity Q as a function of time (Eq. 5.43).

the upward- and downward-going portions of each tone-response period provide information.

The high-frequency condition (7815 Hz) shown in Fig. 5-4 is the same tone level (L=60 dB SPL) as the low-frequency condition, and thus represents a sub-human-threshold condition, due to larger filter bandwidths at high frequencies. The average-noise-alone and tone-in-noise responses show very little fine-time phase locking due to very low synchrony coefficients for the AN model at high frequencies. Each condition shows a large onset response, followed by a relatively flat steady-state response with small envelope fluctuations. These slow fluctuations differ between the signal and standard responses, and thus the temporal information primarily represents envelope information for high frequency tones in noise. Note that the high-frequency condition has much less total sensitivity Q than the low-frequency condition, consistent with higher human thresholds at high frequencies.

5.4.4 Estimating psychophysical thresholds

In order to relate model predictions to psychophysical performance in a general discrimination task, threshold $L_{thr}(L_1)$ is defined as the value of L_2 such that $Q(L_1, L_2 = L_{thr}) = 1$ for a given L_1. Psychophysical discrimination performance is often reported in terms of the just-noticeable difference (JND) in level, $\Delta L_{JND}(L_1) = L_{thr} - L_1$, or in terms of the detection threshold L_{thr} with $L_1 = -\infty$. For a suprathreshold condition, Eq. 5.29 relates $\Delta L_{JND}(L_1)$ to the normalized sensitivity $\delta'_L(L_1)$ predicted from the AN model using Eq. 5.28.

For a general discrimination experiment, threshold is more difficult to estimate using a computational approach. The sensitivity $Q(L_1, L_2)$ must be calculated (Eq. 5.39) for various values of L_2 to determine threshold L_{thr}. Figure 5-5 illustrates the estimation of thresholds for a detection-in-noise task. The significant difference between the random- and fixed-noise sensitivities for the realistic all-information processor observed in Fig. 5-2 results in a much lower threshold for the fixed-noise experiment (0 dB SPL) than for the random-noise condition (24 dB SPL). The optimum threshold (-6 dB SPL) is below the threshold for the realistic processor based on restricted knowledge. Human threshold for this condition is roughly 40 dB SPL, and thus the all-information threshold for the random-noise experiment is well below human threshold, while the rate-place threshold (43 dB SPL) is only slightly above human threshold.

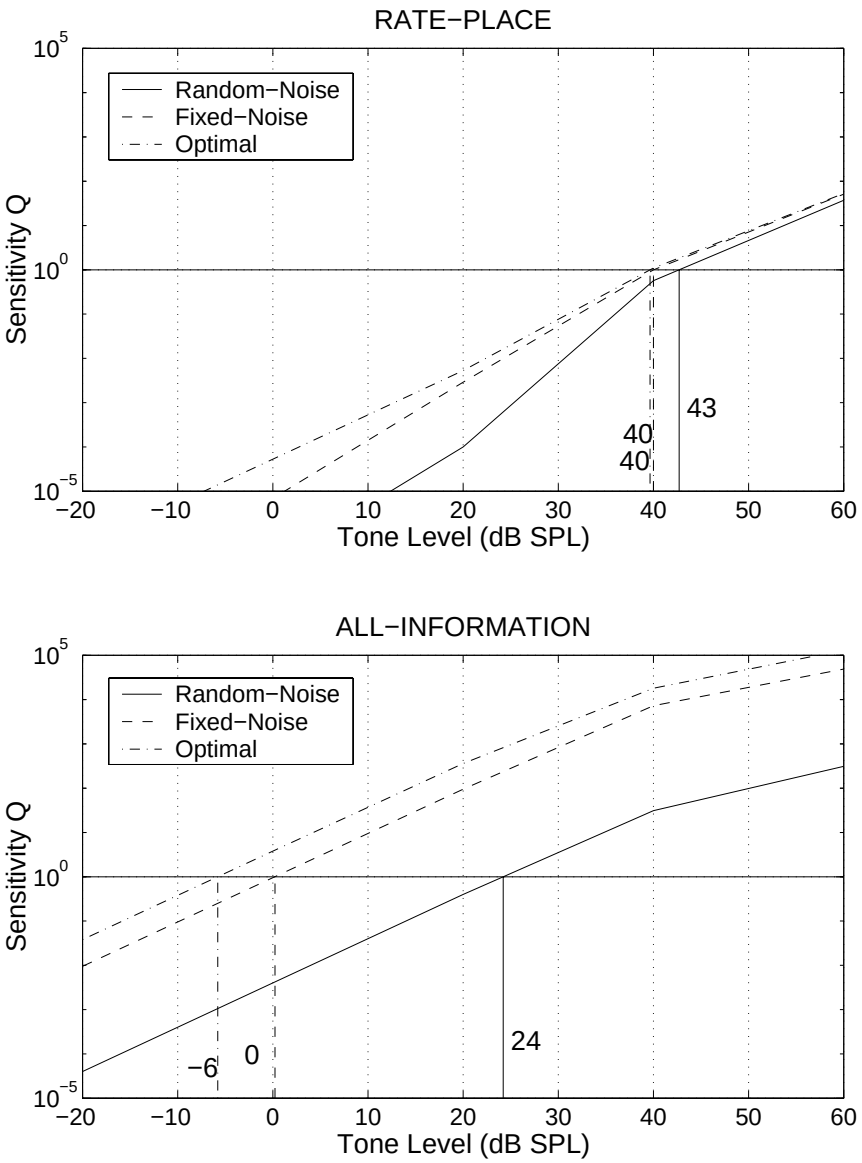

Figure 5-5: Method for determining detection threshold. The sensitivity Q is calculated from Eq. 5.39 for the realistic processor in random- and fixed-noise experiments, as well as for the optimal processor (based on 50 noise waveforms). Threshold is defined as the tone level for which Q=1, and is indicated by the numbers next to the vertical lines for each model. Thresholds are compared for the rate-place (top panel) and all-information (bottom panel) models. M_{CF}=200 high-spontaneous-rate (60 sp/sec) AN fibers for each of the 60 model CF's ranging from 200 to 20,000 Hz. The tone signal was 1049 Hz and 200 ms (20-ms rise/fall). The noise masker had a bandwidth of 10 kHz centered geometrically at the tone frequency, a spectrum level of 20 dB SPL, and the same duration as the tone. The AN response used in Eq. 5.39 included 25 ms after the nominal stimulus offset to include the transient response (see Chapter 2).

5.5 Discussion

The primary goal of this chapter of the dissertation was to describe and demonstrate the theory developed for predicting psychophysical performance limits in experiments that use noise maskers. This theory was based on separating the contribution of auditory-nerve response variation (internal noise) and stimulus variation (external noise) to the variance in a decision variable that can be calculated based only on the average tone-in-noise responses. Predictions for this realistic processor were made for experiments in which the noise waveforms were randomly varied from interval to interval, as well as for experiments in which the effect of external stimulus variation was removed by presenting the same noise waveform in each observation interval. Predictions were compared for processors that rely on the set of AN discharge times (all information) and for processors that rely only on the set of discharge counts (rate place) in order to evaluate the significance of temporal information for discrimination in noise. The sensitivity of individual model CF's can be calculated in order to examine the distribution of information across the AN population. The type of temporal information available can be illustrated by evaluating the sensitivity for a processor that relies only on information from a narrow temporal window. These types of analyses should prove useful for quantifying the effects of noise maskers on the signal information encoded in AN responses, and thus in evaluating the appropriateness of several assumptions that are typically used in psychophysics to interpret masking experiments.

5.5.1 Auditory detection in noise

A secondary goal of the present chapter was to address several issues related to the assumptions in the power-spectrum model of masking (Moore, 1995a). While the present chapter does not represent a thorough evaluation of these assumptions, the predictions described above provide some useful insight into the AN information available for the detection of tones in noise.

It is often assumed that detection in noise depends only on the long-term spectra of the signal and masker (Moore, 1995a). Figure 5-3 demonstrates that there is significant temporal information for detecting tones in noise within AN fibers for which rate-place information is

"masked" (e.g., CF equal to the tone frequency). An illustration of the types of temporal information available for detection of a tone in noise at both low and high frequencies is provided in Fig. 5-4. Fine-time temporal information is available for detecting low-frequency tones in noise, while temporal envelope information is available for detecting high-frequency tones in noise.

Another common assumption is that useful information is limited to CF's near the tone frequency for detection of a tone centered spectrally within a noise masker. The information profiles shown in Fig. 5-3 demonstrate that significant off-frequency information exists for detection of tones in narrowband noise maskers. For the rate-place model, this off-frequency information is much more significant than the on-frequency information. For detection of tones in broadband noise, information within CF's near the tone frequency dominates performance.

The predictions in Fig. 5-2 suggest that the all-information performance based on the total HSR population of AN fibers is consistent with human detection of a 1-kHz tone being equal for 100-Hz and 10-kHz random noise maskers (Fletcher, 1940; Bos and de Boer, 1966). However, all-information sensitivity Q is more than an order of magnitude greater than human sensitivity. Rate-place sensitivity based on the total HSR population is an order of magnitude greater than human sensitivity for the 100-Hz masker; however, it is a factor of two less than human sensitivity for the 10-kHz masker. Thus, the large decrease in sensitivity based on average-rate information in the HSR population of AN fibers is not consistent with human performance. There is a small set of low-spontaneous-rate (LSR) AN fibers [roughly 16% of the total AN population (Liberman, 1978)] that have higher thresholds and larger dynamic ranges than the HSR fibers. Thus, it has been suggested that the LSR fibers may be used for detection in noise at high masker levels (Young and Barta, 1986).

Finally, the predictions in Figs. 5-2 and 5-3 demonstrate that correlation between AN-fiber responses to random noise stimuli influence predictions of sensitivity in complex ways. The information profiles in Fig. 5-3 change their shape when information is combined across multiple statistically identical AN fibers at each model CF. The AN responses of same-CF fibers are much more highly correlated for CF's near the tone frequency than for off-frequency fibers. Thus, the benefit from combining across multiple AN fibers is less near the tone

frequency than for CF's away from the tone frequency. Figures 5-2 and 5-3 both indicate that the same-CF correlation has a much greater influence on the all-information model than on the rate-place model. When the total HSR population of AN fibers is considered, the difference between the random- and fixed-noise sensitivities is much larger for the all-information model than for the rate-place model or for human performance. It is not clear whether this difference results from the all-information model performing too poorly in the random-noise experiment when the AN fibers are correlated or from the all-information model performing much too well when AN fibers are independent. The fact that the random-noise sensitivity for the all-information model is much closer to human sensitivity, both in absolute value and in the trend across masker bandwidth, suggests the latter.

Thus, it appears that the correlation across AN fibers is important to consider for the many psychophysical tasks in which random-noise stimuli are used. The ability to combine information across multiple AN fibers depends strongly on this across-fiber correlation. This is most certainly a significant issue for the peripheral auditory system that conveys sensory information to the brain via 30,000 primary auditory-nerve fibers, and therefore deserves future investigation. The influence of correlation most likely becomes even more complex at higher levels of the auditory system, where several neurons may be innervated by the same projecting neurons from other auditory centers (Jenison, 2000).

5.5.2 Limitations of analysis

Sub-optimum processor

The analysis presented here allows the prediction of performance for any discrimination experiment based on a specific processor that only relies on the ensemble properties of AN responses in the presence of random noise. Thus, the unrealistic assumption that the processor uses each noise waveform individually is avoided. However, a limitation of this approach is that we have not shown that this processor is the optimum processor, given that only the first-order statistical properties (i.e., average) of the AN noise responses are known. While the ability to specify the optimum processor is theoretically pleasing, evaluating the changes in performance of this explicitly-stated *smart* processor as a function of stimulus

parameters allows the issues of interest in this dissertation to be addressed.

Estimation of expected value and variance

The equations derived in the above analysis depend on the expected value E_n and variance Var_n of terms across the ensemble of noise waveforms. The present evaluation of these equations relies on approximating the ensemble expected values and variances from a finite set of N noise waveforms (N is given in each figure caption). The accuracy of these approximations improves with N; however, the computation time required for the predictions also increases. Thus, it is desired to use as small an N as provides reasonable estimates of E_n and Var_n.

The masking predictions in the present study (Fig. 5-2) were made by approximating the expected values and variances based on 50 noise waveforms. The same 50 noise waveforms were used to determine the *a priori* information used by the realistic processor [i.e., $x_i(t|L_1) = \overline{r_i}(t|L_1)$ and $x_i(t|L_2) = \overline{r_i}(t|L_2)$] and to evaluate performance (Eq. 5.21). [4] While calculating the AN responses to 50 noises for each alternative and for each of the 60 model CF's is reasonable for the few conditions examined in the present chapter, it is too large to allow useful examination of entire psychophysical experiments, such as are addressed in the following chapter.

The AN responses used for the predictions of sensitivity Q in Fig. 5-2 were examined in order to evaluate whether 10 noise waveforms could be used to provide reasonably accurate predictions. The 50 AN responses were divided into five sets of 10, and sensitivity Q was calculated based on each set of 10 noise responses for both the 100-Hz and the 10-kHz noise maskers. The standard deviation of Q across the sets of 10 noises and the slope of Q as a function of tone level L (Fig. 5-5) were used to determine the error in the threshold estimate that would result from values of Q that were one standard deviation away from the mean. It was determined that all-information thresholds would be within ± 2.3 dB of their true value and rate-place thresholds would be within ± 1.4 dB of their true value for both low and high frequencies. Thus, the predictions in the following chapter were made based on 10 noise waveforms; however, the use of 15 or 20 noise waveforms would result in significantly more

[4] It would be preferable in future studies to use a different set of noise waveforms to determine the processor's *a priori* information than the set of noises used to evaluate performance.

accurate predictions, and thus may be warranted for future simulations.

Assumption of known tone phase

The analyses described above for level discrimination and detection of a tone in the presence of a background noise assumed a fixed tone phase. While this assumption is not desired in the long run, it is not likely to significantly affect the current performance predictions. In addition, this assumption will likely be removed in future analyses (e.g., monaural, across-frequency coincidence detection does not rely on knowledge of the tone phase on each trial, see Chapter 4). Siebert (1970) evaluated the significance of assuming a fixed phase for pure-tone frequency discrimination, and showed that performance was only degraded by a factor of $\sqrt{2}$ when the tone phase was randomized from trial to trial. He described the small effect as arising from the use of temporal comparisons that depend on either one or two random (Poisson) discharge times. In the case of a fixed and known phase, the processor has an absolute temporal reference, and thus the temporal precision of each discharge time includes the variance of only a single Poisson event. In the random-phase case, the processor must rely on relative temporal comparisons, such that the precision of any discharge time includes the variance of two random variables. The decrease by factor of two in the variance of temporal comparisons accounts for the decrease in performance by a factor of $\sqrt{2}$. The known-phase assumption eliminates a large amount of computation time in the evaluation of the equations described above, and is suggested to have minimal significance based on Siebert's (1970) analysis. However, the significance of this assumption should be evaluated specifically for detection experiments in future studies.

5.6 Acknowledgments

We thank Xuedong Zhang and Ian Bruce for their collaboration on the auditory-nerve model. Supported by NIH Grants T32DC00038 and R01DC00100, and NSF Grant 9983567. The simulations in this study were performed on computers provided by the Scientific Computing and Visualization group at Boston University.

Chapter 6

The influence of the cochlear amplifier on auditory-filter estimates using the notched-noise method

6.1 Abstract

Many psychophysical methods estimate auditory tuning in human listeners by measuring the detection of tones in various noise maskers. However, the cochlear amplifier dynamically alters basilar-membrane tuning in the normal auditory system based on the spectral and temporal configuration of the stimulus. The present study used several versions of a phenomenological AN model to evaluate the effects of compression and suppression on estimates of auditory filters using the notched-noise method (Patterson, 1976; Glasberg and Moore, 1990). The computational AN model included the effects of level-dependent tuning, level-dependent phase, compression, suppression, and fast nonlinear dynamics on the responses of high-, medium-, and low-spontaneous-rate AN fibers. Quantitative methods used statistical decision theory to evaluate psychophysical performance limits imposed by the random nature of AN discharges and random-noise stimuli. Performance was predicted based on AN information in discharge counts and in discharge times for the detection of a tone in notched-noise maskers. Standard methods based on the power-spectrum model of masking were used to estimate auditory filters from the detection predictions. Both compression and suppression are shown to influence psychophysical methods for estimating auditory frequency selectivity. Implications of the dependence of auditory-nerve tuning on both the signal and masker are discussed in terms of psychophysical methods for estimating auditory filters. These include differences between simultaneous and nonsimultaneous masking, as well as the influence of asymmetric suppression growth on the estimates of auditory-filter asymmetry.

6.2 Introduction

The ability of the auditory system to partially resolve frequency components in a complex stimulus has been widely used as the basis for many fundamental theories of auditory perception (e.g., von Helmholtz, 1863; Fletcher, 1940). Thus, much effort has gone into developing techniques for the measurement of the frequency selectivity of the auditory system. Fletcher (1940) observed that the detection threshold for a tone in band-limited noise increases as the bandwidth of the noise increases until a "critical bandwidth", beyond which detection threshold was roughly constant. This observation led to the concept of the "auditory filter," which forms the basis for many modern psychophysical methods for estimating auditory tuning (for a review see Moore, 1995a).

Most psychophysical methods for measuring auditory frequency selectivity are based on the power-spectrum model of masking, which assumes that the detection of a tone in noise is based on the signal and noise energy that passes through a single linear filter. The notched-noise method is a common technique of estimating auditory filter shapes by measuring detection thresholds for a tone in the presence of two bandlimited noise maskers that are above and below the tone frequency (Patterson, 1976; Glasberg and Moore, 1990; Rosen *et al.*, 1998). Detection is measured as a function of the separation between the two noise bands (i.e., the notch width), and the power-spectrum model is used to estimate a filter shape that accounts for the rate of decrease in threshold as notch width is increased. Asymmetric notches are often used to evaluate the upper and lower sides of the auditory filter independently. Filter estimates based on the notched-noise method are successful in predicting detection performance in a variety of masking conditions (e.g., Derleth and Dau, 2000). However, the observation that cochlear tuning becomes broader as stimulus level increases in association with the cochlear amplifier (Rhode, 1971; Ruggero *et al.*, 1997) complicates the interpretation of filter estimates based on the power-spectrum model.

A consistent finding is that psychophysical estimates of auditory-filter bandwidth increase as a function of level, primarily due to a decrease in the slope of the low-frequency side as level is increased (Glasberg and Moore, 1990; Rosen *et al.*, 1998). In addition, the most consistent psychophysical finding related to hearing impairment, besides an increase in

threshold, is broadened auditory filters for impaired subjects compared with normal-hearing listeners (Moore, 1995b). This result is consistent with the cochlear amplifier being damaged or absent in impaired ears. Moore *et al.* (1999b) have shown that estimates of equivalent-rectangular bandwidth (ERB) using the notched-noise method are highly correlated with several psychophysical measures of compression, which is also thought to be related to the cochlear amplifier. Thus, it appears that psychophysical measures of auditory frequency selectivity are related to the health of the cochlear amplifier; however, there is continued debate as to the proper method for estimating auditory filter shapes as a function of level (Lutfi and Patterson, 1984; Moore and Glasberg, 1987; Glasberg and Moore, 1990; Rosen and Baker, 1994; Moore, 1995a; see Rosen *et al.*, 1998 for a review). This debate is not surprising given that current methods for estimating the nonlinear tuning of the auditory system use the power-spectrum model, which is based on linear-filter theory. The central issue in this debate has been whether the filter shape depends on the input to or the output of the auditory filter. Glasberg and Moore (1990) suggest that the pre-filtered stimulus level determines the filter shape, and thus they measure detection using a fixed noise spectrum level. Rosen *et al.* suggest that the post-filtered stimulus determines the auditory tuning, and thus estimate auditory filters using a fixed probe-tone level.

While no method based on linear-filter theory is appropriate for estimating nonlinear tuning in all conditions, it is important to understand how to interpret psychophysical estimates of auditory-filter shapes in relation to the many nonlinear response properties of the auditory periphery. Basilar-membrane (BM) tuning has been shown to broaden with increases in level, and to demonstrate compressive magnitude responses and nonlinear phase responses (Ruggero *et al.*, 1997). It was demonstrated in Chapter 4 that compression of magnitude responses and nonlinear phase responses are significant for the encoding of sound level within a narrow range of characteristic frequencies (CF's); however, the AN model that was used was limited to pure-tone stimuli. Two-tone suppression is another nonlinear response property of the auditory periphery, and refers to the ability of an off-frequency tone to suppress BM and AN responses to CF-tones (Sachs and Kiang, 1968; Delgutte, 1990b; Ruggero *et al.*, 1992). Delgutte (1990a) demonstrated that AN suppression plays a large role in masking of signals by off-frequency stimuli, a finding that was supported by similar psychophysical

experiments (Oxenham and Plack, 1998). Suppression has been hypothesized to play a role in auditory tuning based on sharper psychophysical estimates of frequency selectivity using non-simultaneous masking compared to simultaneous masking (e.g., Houtgast, 1977; Moore, 1978; Moore and Glasberg, 1981,1986; Moore *et al.*, 1987). The variation in dynamic range of AN discharge rate across different spontaneous-rate populations has been hypothesized to be related to compressive BM responses (Sachs and Abbas, 1974; Liberman, 1978; Winter and Palmer, 1991).

These nonlinear physiological response properties have typically been studied in isolation, e.g., it is not clear from physiological two-tone suppression studies alone that the ability of an off-frequency, non-excitatory tone to suppress the response to a CF tone is related to auditory tuning. However, much evidence has now been collected suggesting that many of the observed nonlinear AN response properties are likely to result from the same mechanism, the cochlear amplifier (Sachs and Abbas, 1974; Sewell, 1984; Patuzzi *et al.*, 1989; Ruggero and Rich, 1991; Ruggero *et al.*, 1992). A phenomenological AN model with nonlinear tuning (Zhang *et al.*, 2000) is used in the present study to relate many of the response properties associated with the cochlear amplifier to psychophysical estimates of auditory tuning.

Auditory filters are estimated using the notched-noise method by quantifying the information available in average-rate and temporal discharge patterns in the population of AN fibers. A theoretical approach developed in Chapter 5 is used to quantify the contributions of the stochastic AN-discharge variation (internal noise) and stimulus variation (external noise) on performance limits for the detection of tones in noise maskers. A major benefit of this approach for evaluating the notched-noise method is that the AN-fiber bandwidths in the model are known, and can be compared directly to the auditory-filter estimates. Different versions of the computational AN model are used to demonstrate the contributions of compression and suppression to estimates of auditory tuning using the notched-noise method.

6.3 Methods

6.3.1 Computational auditory-nerve model

The implementation and response properties of the computational auditory-nerve (AN) model that are important for the present study are described in Appendix A, and the complete model is described by Zhang *et al.* (2000). Predictions from four versions of the AN model (Fig. 6-1) are compared below in order to evaluate the role of several response properties associated with the cochlear amplifier, including level-dependent tuning, compression, and suppression (see Appendix A for details). The *linear-sharp* version of the AN model has low thresholds and linear excitatory filters with narrow bandwidths. The *impaired* version of the AN model has high thresholds and linear filters with broad bandwidths. Two nonlinear versions of the AN model are used that have excitatory filters with tuning that matches the linear-sharp model at low levels and the impaired model at high levels. The two nonlinear models differ from one another in the spectral content of the stimulus that controls the level-dependent tuning. The *nonlinear AN model without suppression* includes controls the level-dependent gain and bandwidth by the stimulus energy within the low-level excitatory filter, and thus demonstrates compression but not suppression. The *nonlinear AN model with suppression* has level-dependent gain and bandwidth that are varied according to the stimulus energy within a bandpass filter that is five times as broad as the low-level excitatory filter (e.g., Geisler and Sinex, 1980). Thus, this version of the model demonstrates both compression (i.e., stimulus energy that passes through the excitatory filter acts to compress the AN response by broadening the excitatory filter) and suppression (i.e., stimulus energy that passes through the broadband filter but not the excitatory filter acts to suppress the AN response by broadening the excitatory filter).

The computational AN model described in Appendix A was used to generate a population response for each of the three spontaneous-rate (SR) groups described by Liberman (1978). A set of 60 model characteristic frequencies (CF's) were used, similar to Chapter 5. The CF's ranged from 200 Hz to 20 kHz, and were spaced according to a human cochlear map (Greenwood, 1990). This spacing of fiber CF's corresponds to roughly 0.5 mm separation on BM, and is estimated as about one half of a human ERB (Glasberg and Moore, 1990). The

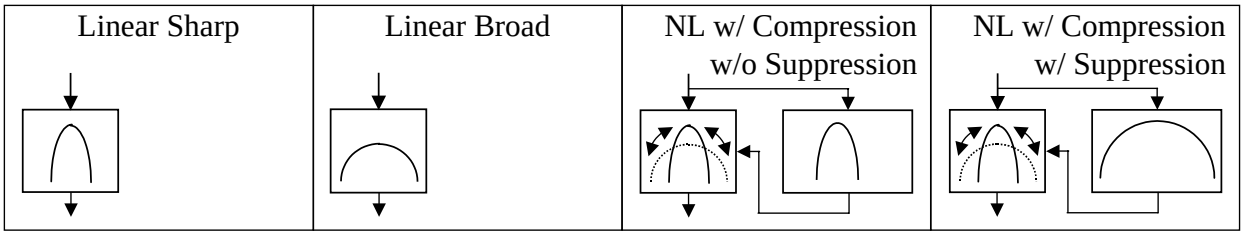

Figure 6-1: Four versions of the phenomenological AN model used to isolate the effects of compression and suppression on tuning.

tone frequency was always chosen to be equal to one of the model CF's, and thus can differ slightly from the tone frequency used in the human data from previous studies. Populations of high-, medium-, and low-SR fibers were simulated by assuming 200 HSR, 75 MSR, and 50 LSR fibers per CF. This represents all human AN fibers within the frequency range from 200 Hz to 20 kHz (Chapter 5), and is consistent with the 61%, 23%, and 16% distributions reported by Liberman (1978). All AN fibers were assumed to have stochastic responses that were conditionally independent Poisson processes given the stimulus.

6.3.2 Statistical decision theory

The present study uses the statistical-decision-theory (SDT) analysis developed in Chapter 5 to quantify performance limits due to both random variations in the noise stimuli and in the AN responses. The effect of noise maskers on predicted performance is compared for processors based on rate-place information and on all information (both temporal and average rate), as shown in Fig. 2 of Chapter 2. In the rate-place analysis, the observations are the population of spike counts $\{K_i\}_{i=1,\cdots,M}$ produced by homogeneous Poisson processes with rates equal to the average rates $\bar{r}_i$ produced by the AN model. The all-information analysis is based on observing the population of spike times $\{t_1^i, \cdots, t_{Ki}^i\}_{i=1,\cdots,M}$ produced by non-homogeneous Poisson processes with time- varying rate waveforms $r_i(t)$ produced by the AN model. This section provides an overview of the detection-in-noise analysis that was used in the current study, while the details of this approach are described in Chapter 5.

The form of a general processor that evaluates each discharge time t_j^i from the i^{th} AN fiber based on *a priori* information that consists of the average signal-plus-noise $x(t|SN)$ and noise-alone $x(t|N)$ responses was described in Chapter 5, and is given by

$$Y_i(\mathcal{T}) \triangleq \sum_{j=1}^{K_i} \ln \frac{x_i(t_j|SN)}{x_i(t_j|N)} + \int_0^T [x_i(t|N) - x_i(t|SN)]dt. \tag{6.1}$$

The performance of this *realistic* (sub-optimal) processor based on all M AN fibers is given by the sensitivity index [similar to $(d')^2$]

$$\begin{aligned} Q = & \left(E_n \left\{ \sum_{i=1}^{M} \int_0^T \ln \frac{x_i(t|SN,n)}{x_i(t|N,n)} \left[r_i(t|SN,n) - r_i(t|N,n) \right] dt \right\} \right)^2 \Bigg/ \\ & \left(E_n \left\{ \sum_{i=1}^{M} \int_0^T \left[\ln \frac{x_i(t|SN,n)}{x_i(t|N,n)} \right]^2 r_i(t|N,n) dt \right\} \right. \\ & \left. + Var_n \left[\sum_{i=1}^{M} \int_0^T \ln \frac{x_i(t|SN,n)}{x_i(t|N,n)} r_i(t|N,n) + x_i(t|N,n) - x_i(t|SN,n) dt \right] \right), \end{aligned} \tag{6.2}$$

where E_n and Var_n represent the expected value and variance across the set of random noise waveforms n, and the first term in the denominator of Eq. 6.2 represents the *AN variance* and the second term represents the *stimulus variance*. Detection threshold is defined as the signal level for which the sensitivity index $Q = 1$. Performance is compared for two types of experiments using noise stimuli, as described in Chapter 5. A *random noise* experiment uses different noise waveforms in each observation interval, while a *fixed noise* experiment uses identical noise waveform in both observation intervals, and thus has no stimulus variance.

Performance was predicted based on the high-, medium-, and low-spontaneous-rate (SR) populations of AN fibers individually, as well as based on the total AN-fiber population. The predictions for the total AN population were made by including all AN fibers in the summations in Eq. 6.2, which accounts for any potential correlation between AN fibers of different SR. Note that the sensitivity indices Q for individual fibers cannot be simply added to obtain the total sensitivity because the fibers' responses are potentially correlated when stimulated with random-noise stimuli. Also, the information from different AN fibers is not necessarily combined optimally for the *realistic* (sub-optimal) processor used in the present

predictions (see Chapter 5).

6.3.3 Notched-noise method for estimating auditory filters

The present study used the procedure described by Glasberg and Moore (1990) to estimate auditory-filter shapes based on notched-noise stimuli and the power-spectrum model. Thresholds were predicted for the detection of a $f = 1049$ Hz tone in the presence of two noise bands that were above and below the tone frequency, and had bandwidths equal to $0.4f$. The separation between the closest edge of each noise band and the tone frequency is referred to in terms of the normalized frequency separation $\frac{\Delta f}{f}$. The tone was always centered between the two noise bands, i.e., only symmetric notches were used in the present study, and thus the total notch width of the noise masker was $2\Delta f$. Detection thresholds were measured for normalized frequency separations of 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, and 0.6. A fixed noise spectrum level of $N_0 = 20$ dB SPL was used for all conditions in the present study. This spectrum level was chosen in order to evaluate whether compression and suppression influence psychophysical estimates of auditory filters at a spectrum level near the lowest levels that are typically used in the notched-noise method.

The method described by Glasberg and Moore (1990) was used to fit rounded-exponential filters, roex(p,r), to the predicted detection thresholds in the present study, where p is a parameter that describes the slope of the filter, and r is a parameter that determines the dynamic range of the filter. The fitting procedure was used without the corrections described by Glasberg and Moore (1990) for earphone frequency response or for the transmission characteristics of the middle and external ear. The estimated auditory filters that result from the fitting procedure are described in terms of the slope p, the dynamic range r, the equivalent-rectangular bandwidth ERB, and a parameter K that represents the signal-to-noise ratio at the output of the filter that results in detection threshold (i.e., K is inversely related to the efficiency of the processor). The goodness of fit of the power-spectrum model is quantified in terms of the root-mean-squared (rms) difference between the AN-model detection thresholds and the predicted thresholds based on the auditory filter.

6.4 Results

Detection thresholds were predicted for the rate-place and all-information models as a function of the notch width of the noise masker. This experiment is identical to the method often used to estimate auditory-filter shapes (e.g., Glasberg and Moore, 1990). Insight into the significance of several nonlinear physiological mechanisms associated with the cochlear amplifier can be garnered by comparing estimated auditory filters for different versions of the AN model to estimates from human data.

The simplest comparison is for versions of the AN model with linear tuning, because the equivalent-rectangular bandwidths (ERB) used in the model are fixed and known, and can be compared to the estimates of ERB from the notched-noise method. Predicted performance for the *linear-sharp* version of the AN model are shown in Fig. 6-2 for the three SR populations (open symbols) as well as for the total AN population (TOT: filled symbols), and are based on a random-noise experiment. The rate-place detection thresholds for the total-AN population do not correspond to one SR group for all notch widths. Rather, rate-place performance is dependent on the LSR population for small notch widths and the HSR population for large notch widths. All-information thresholds are always well below the rate-place thresholds, especially for notch widths larger than $\Delta f/f$=0.0. Human data is shown for the normal ear of a unilaterally-impaired subject based on a noise spectrum level of $N_0 = 50$ dB SPL (Glasberg and Moore, 1986). The rate-place thresholds shown for $N_0 = 20$ dB SPL are roughly 40 dB less than the human thresholds, but match the general shape of the dependence of human thresholds on notch width.

The power-spectrum model was used to fit auditory filters to the model predictions based on the total AN population as well as to the human data. The fit of the power-spectrum model to the rate-place predictions based on the notch widths ranging from 0.0 to 0.6 (solid lines) is quite good, as is the fit to the human data. In contrast, the fit to the all-information thresholds based on all notch widths is poor. The high predicted all-information threshold for the 0.0 notch width is the primary cause of the poor fit, as demonstrated by the strong improvement in the fit when the 0.0 notch width is excluded. The cause of the relatively high all-information threshold at 0.0 notch width is discussed below. A summary of the

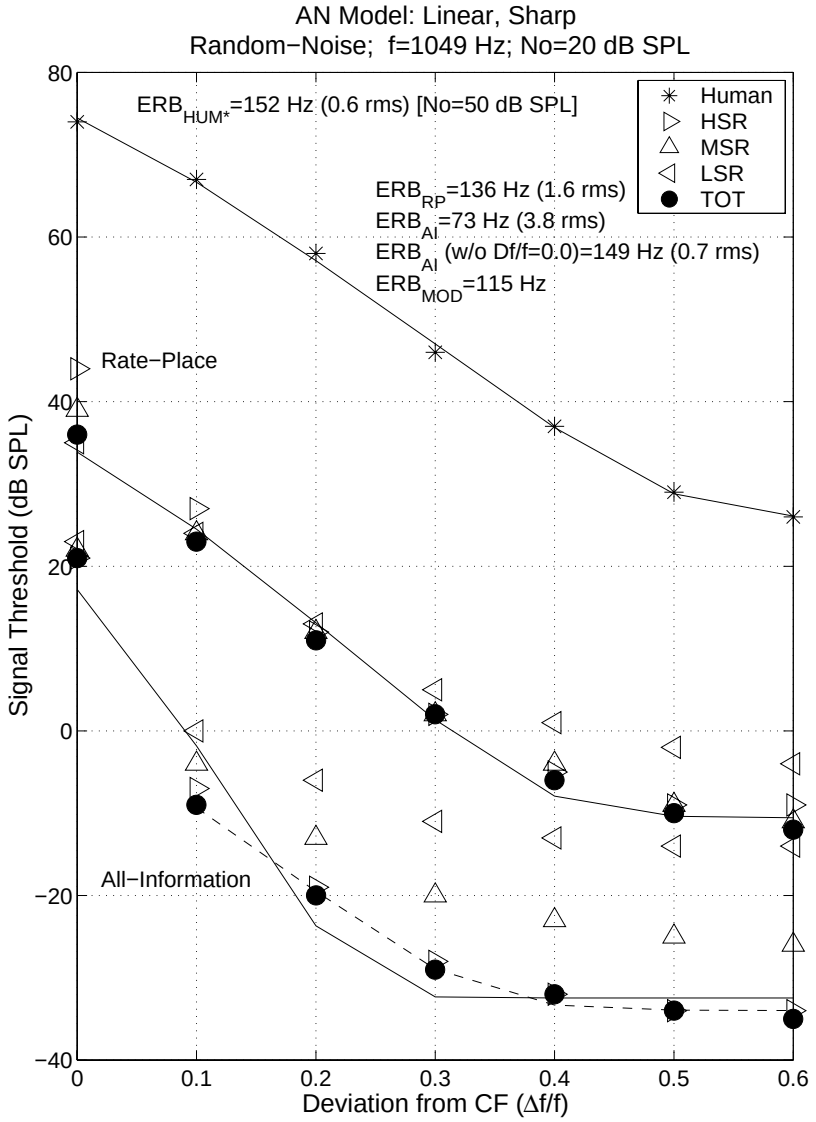

Figure 6-2: Detection of a tone in random notched noise based on the *linear AN model with sharp tuning*. Threshold is plotted as a function of normalized notch width, where Δf is the deviation of the closest edge of each noise band from the tone frequency (i.e., $2\Delta f$ is the notch width). Predicted rate-place (RP) and all-information (AI) performance is shown for each of the three spontaneous-rate (SR) populations, as well as for the combination of all three SR groups (TOT). Model predictions are for a 1049-Hz, 200-ms (20-ms rise/fall) tone presented simultaneously with noise bands that had a bandwidth of $0.4f$ and a spectrum level of N_0=20 dB SPL. Fitted thresholds (solid and dashed lines), equivalent-rectangular bandwidth (ERB) estimates, and root-mean-squared (rms) errors from the power-spectrum model are shown for the RP model and for the AI model based on all seven notch widths (solid lines), and for the AI model based on the six notch widths excluding 0.0 (dashed line). The ERB of the 1049-Hz CF in the AN model was 115 Hz. Human detection data and power-spectrum-model predictions are shown for a 1-kHz, 200-ms (10-ms rise/fall) tone and noise-spectrum level of $N_0 = 50$ dB SPL (subject TD: Glasberg and Moore, 1986).

parameters used by the power-spectrum model to fit the random-noise rate-place and all-information predictions based on all notch widths is given in Table 6.1, while Table 6.2 provides a summary of the fits based on the exclusion of the 0.0 notch width.

The estimated equivalent-rectangular bandwidth (ERB) for the rate-place model was 136 Hz, which is only slightly larger than the ERB (115 Hz) used for the 1049-Hz CF in the model [Glasberg and Moore's (1990) low-level ERB]. The ERB for the all-information model was 73 Hz, and was very sharp due to the high threshold for the 0.0 notch width. When the 0.0 notch width was excluded, the all-information ERB was 149 Hz and was much closer to the estimated rate-place ERB. The ERB for the human data shown was 152 Hz, which matched well with the ERB of 155 Hz for a 1-kHz tone and a N_0=50 dB SPL spectrum level obtained from Glasberg and Moore's (1990) description of the dependence of auditory-filter shape on level. The human ERB for a 1-kHz tone and a spectrum level of N_0=20 dB SPL obtained from Glasberg and Moore's (1990) equations was 125 Hz.

Predicted detection thresholds for the *impaired* version of the AN model are shown in Fig. 6-3. This version of the AN model has linear filters with bandwidths that are broader than the linear-sharp model by a factor that is related to the amount of cochlear-amplifier gain at each CF (See Appendix A). The 1049-Hz CF filter in the impaired model has an ERB of 248 Hz. The power-spectrum model fits to the rate-place thresholds based on the total AN population are again quite good. The all-information fits are not as good, especially when the 0.0 notch width is included. The ERB for the rate-place model was 227 Hz, which is close to the model-filter ERB at 1049 Hz. The all-information ERB is much narrower than the model ERB when all notch widths are included and when the 0.0 notch width is excluded. Human data from the impaired ear of the listener shown in Fig. 6-2 are also shown for comparison. The rate of decrease in threshold as notch width is increased is much slower for the impaired rate-place and human thresholds, and is the cause of the broadened ERB estimates.

The predicted thresholds for the *nonlinear AN model with suppression* are shown with the fits from the power-spectrum model in Fig. 6-4. The human data from the normal ear of the subject shown in Fig. 6-2 is shown for comparison. The fits from the power spectrum model are not as good for the nonlinear version of the AN model as for the linear AN model.

Table 6.1: Auditory-filter parameters estimated from the random-noise thresholds predicted by the AN-model for notch widths from 0.0 to 0.6. The power-spectrum model was used to fit a roex(p,r) filter to the AN thresholds. Estimates for both the rate-place and all-information thresholds from the four versions of the AN model are shown. ERB: equivalent-rectangular bandwidth; rms: root-mean-squared error; p: filter slope; r: filter dynamic range; K: signal-to-noise at the filter output required for detection threshold.

Model	ERB (Hz)	rms (dB)	p	r	K (dB)	Model Filter ERB(Hz)
Rate-Place						
Linear, Sharp	135.8	1.59	30.9	-52.3	-7.5	115
Nonlinear, wo/Suppression	160.2	1.49	26.2	-48.8	-8.1	115-248
Nonlinear, w/Suppression	254.7	2.62	16.5	-37.1	-13.3	115-248
Impaired	226.8	0.38	18.5	-35.5	-6.9	248
All-Information						
Linear, Sharp	73.3	3.78	57.3	-60.2	-21.5	115
Nonlinear, wo/Suppression	85.8	5.03	48.9	-50.2	-20.5	115-248
Nonlinear, w/Suppression	272.3	4.84	15.4	-42.5	-31.4	115-248
Impaired	91.3	2.08	46.0	-44.6	-19.4	248

Table 6.2: Auditory-filter parameters estimated from the random-noise thresholds predicted by the AN-model for notch widths from 0.1 to 0.6. The power-spectrum model was used to fit a roex(p,r) filter to the AN thresholds. Estimates for both the rate-place and all-information thresholds from the four versions of the AN model are shown. ERB: equivalent-rectangular bandwidth; rms: root-mean-squared error; p: filter slope; r: filter dynamic range; K: signal-to-noise at the filter output required for detection threshold.

Model	ERB (Hz)	rms (dB)	p	r	K (dB)	Model Filter ERB(Hz)
Rate-Place						
Linear, Sharp	156.8	0.85	26.8	-48.5	-12.0	115
Nonlinear, wo/Suppression	168.1	1.53	25.0	-47.6	-9.6	115-248
Nonlinear, w/Suppression	323.6	1.95	13.0	-86.9	-18.2	115-248
Impaired	233.0	0.35	18.1	-35.1	-7.5	248
All-Information						
Linear, Sharp	148.6	0.72	28.3	-40.6	-42.6	115
Nonlinear, wo/Suppression	299.3	2.47	14.0	-72.9	-45.0	115-248
Nonlinear, w/Suppression	371.1	2.69	11.3	-142.4	-39.1	115-248
Impaired	162.8	0.92	26.0	-32.0	-32.9	248

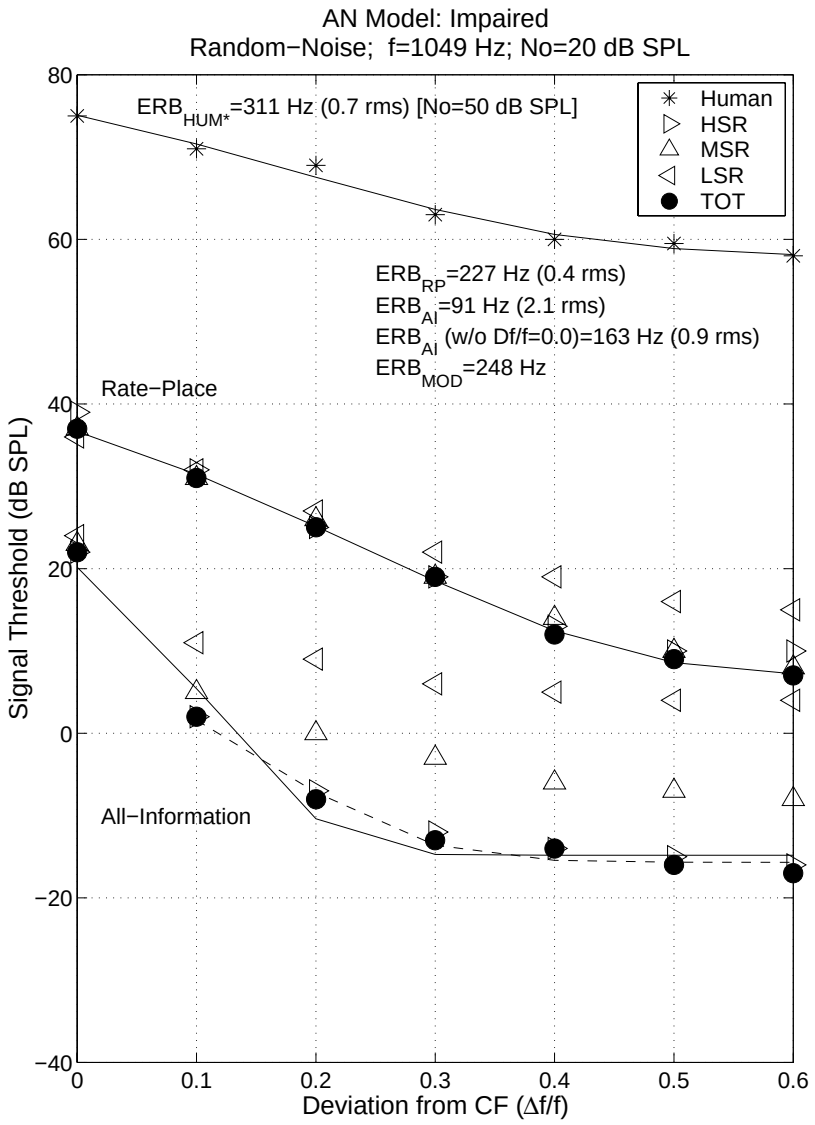

Figure 6-3: Detection of a tone in random notched noise based on the *impaired AN model*. Same symbols as Fig. 6-2. The ERB of the 1049-Hz CF in the AN model was 248 Hz. Impaired human detection data and power-spectrum-model predictions are shown for a 1-kHz, 200-ms (10-ms rise/fall) tone and noise-spectrum level of $N_0 = 50$ dB SPL (subject TD, impaired ear: Glasberg and Moore, 1986).

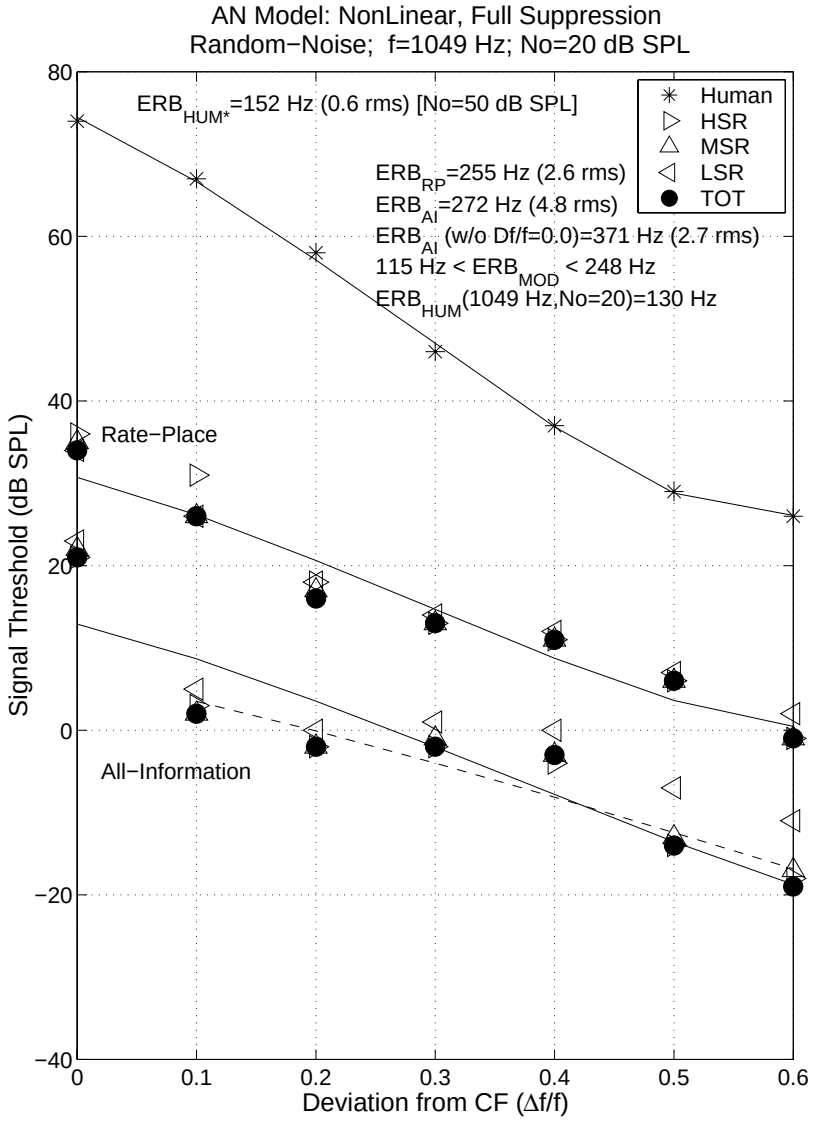

Figure 6-4: Detection of a tone in random notched noise based on the *nonlinear AN model with suppression*. Same symbols as Fig. 6-2. The ERB of the 1049-Hz CF in the AN model varied nonlinearly, and was between 115 and 248 Hz. Human detection data and power-spectrum-model predictions are shown for a 1-kHz, 200-ms (10-ms rise/fall) tone and noise-spectrum level of $N_0 = 50$ dB SPL (subject TD: Glasberg and Moore, 1986).

The estimated ERB's for the rate-place (255 Hz) and all-information (272) models are much broader than the ERB's estimated from human data. The ERB for the listener shown was 152 Hz (f =1 kHz and N_0=50 dB SPL), while the ERB based on the level-dependent estimates provided by Glasberg and Moore (1990) was 130 Hz ($f = 1049Hz$ and N_0=20 dB SPL). The model-filter ERB for the 1049-Hz CF varies nonlinearly within the range from 115 Hz at low levels to 248 Hz at high levels. The rate-place predictions decrease at a rate similar to the human data up to a notch width of 0.2; however, as notch width increases further the rate-place thresholds decrease much more slowly than the human data. The poor fit to the all-information thresholds is due to both the threshold at 0.0 notch width, which is 20 dB higher than for the 0.1 notch width, and the roughly equal thresholds for notch widths between 0.1 and 0.4.

Figure 6-5 shows the predicted thresholds and fits from the power-spectrum model for the *nonlinear AN model without suppression*. The normal-hearing human data is shown for comparison. The fit to the rate-place predictions is fairly good, especially for notch widths smaller than 0.5. The estimated ERB for the rate-place model was 160 Hz, which is significantly smaller than the ERB estimated from the AN model when suppression was included. The fits to the all-information model are both poor, due primarily to the relatively flat thresholds between notch widths of 0.2 to 0.4. The all-information ERB was 86 Hz when all notch widths were included, and was much broader (299 Hz) when the 0.0 notch width was excluded.

Predicted rate-place thresholds for the four versions of the AN model are compared in Fig. 6-6 for the random-noise experiment. Both the normal and impaired human data are shown for comparison. The fits from the power-spectrum model based on all notch widths are shown as the lines in Fig. 6-6, and the estimated ERB's are shown in the legend. Table 6.1 summarizes the estimates and parameters from the power-spectrum model fits to the four versions of the AN model.

The highest predicted rate-place threshold for each notch width was for the impaired version of the AN model, while the lowest threshold was for the linear-sharp AN model, except for the 0.0 notch width (Fig. 6-6). Thresholds from the two nonlinear versions of the AN model (with and without suppression) were essentially the same for notch widths

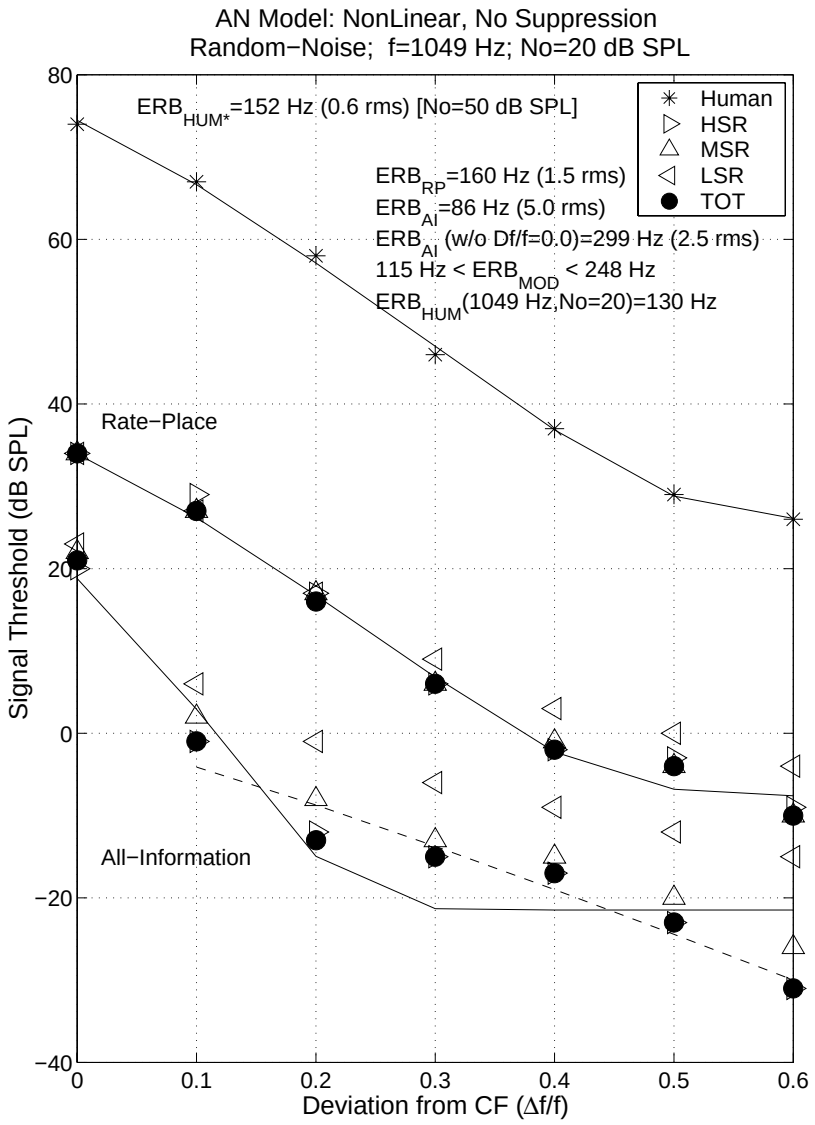

Figure 6-5: Detection of a tone in random notched noise based on the *nonlinear AN model without suppression*. Same symbols as Fig. 6-2. The ERB of the 1049-Hz CF in the AN model varied nonlinearly, and was between 115 and 248 Hz. Human detection data and power-spectrum-model predictions are shown for a 1-kHz, 200-ms (10-ms rise/fall) tone and noise-spectrum level of $N_0 = 50$ dB SPL (subject TD: Glasberg and Moore, 1986).

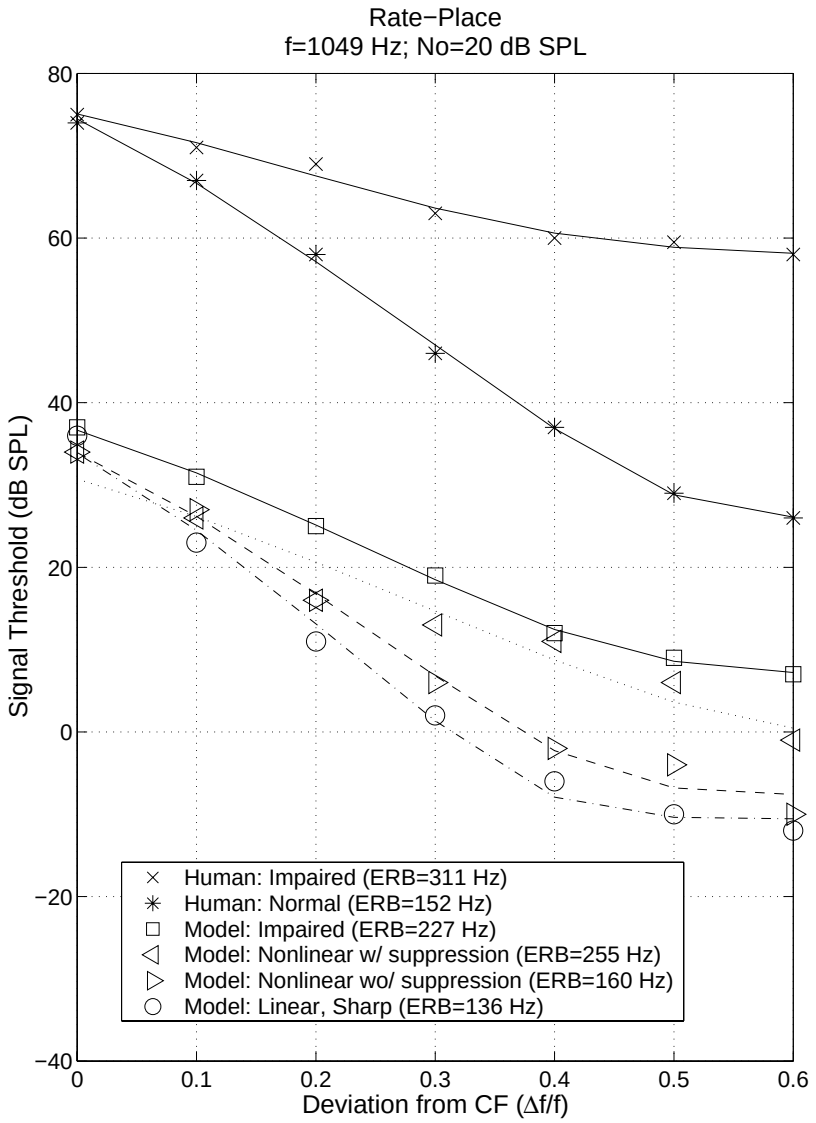

Figure 6-6: Predicted tone detection in random notched noise based on rate-place (RP) information for the four versions of the AN model. Predicted performance is based on the combination of the three spontaneous-rate populations (TOT curves in Figs. 6-2-6-5. Model predictions are for a 1049-Hz, 200-ms (20-ms rise/fall) tone presented simultaneously with noise bands that had a bandwidth of $0.4f$ and a spectrum level of N_0=20 dB SPL. Fitted thresholds (solid and dashed lines) and equivalent-rectangular bandwidth (ERB) estimates from the power-spectrum model are shown based on all seven notch widths. Human detection data and power-spectrum-model predictions are shown for a 1-kHz, 200-ms (10-ms rise/fall) tone and noise-spectrum level of N_0 = 50 dB SPL (subject TD, normal and impaired ear: Glasberg and Moore, 1986).

of 0.2 and less, as expected for conditions in which the majority of the noise energy is near the frequency of the tone. However, the thresholds for the nonlinear AN model with suppression are up to 10 dB higher than the nonlinear model without suppression for broader notch widths. Thus, suppression has a significant effect for broad notch widths, and acts to increase the estimated ERB using the notched noise method.

The thresholds from the nonlinear AN model without suppression are consistently higher than the linear-sharp model, except for the 0.0 notch width for which the linear-sharp threshold is 2-dB higher than the nonlinear thresholds. The higher thresholds for the nonlinear model without suppression at notch widths greater than 0.2 (i.e., notch widths for which the tone level at threshold is less than 20 dB SPL) likely result from broader filters in the nonlinear model. However, given that the tone level is below 20 dB SPL, which is the threshold for at-CF compression in tone responses, the broadened tuning in the nonlinear model without suppression must be due to the noise energy that passes through the excitatory filter. Thus, compression of AN responses also acts to increase the estimated ERB compared with the linear-sharp AN model.

The broader ERB estimated for the nonlinear AN model with suppression compared with the impaired model results from the relatively flat thresholds for notch widths between 0.2 and 0.4, even though these thresholds are lower than the impaired thresholds for all notch widths. For notch widths of 0.2 or less, thresholds from the nonlinear models were significantly less than for the impaired model, suggesting that the excitatory filters in the nonlinear models were significantly narrower than the impaired model for this noise spectrum level. Thus, effects from off-frequency nonlinearities can produce ERB estimates that are broader than the excitatory filters.

The predicted all-information thresholds from the four AN-model versions are compared in Fig. 6-7. Several effects in the all-information predictions are similar to the rate-place predictions. The lowest all-information thresholds for each notch width except 0.0 were from the linear-sharp AN model. The thresholds for the nonlinear model with suppression were always higher than the thresholds for the nonlinear model without suppression, which were always higher than the linear-sharp thresholds. However, in contrast to the rate-place model, the thresholds for the nonlinear model with suppression were higher than the thresholds for

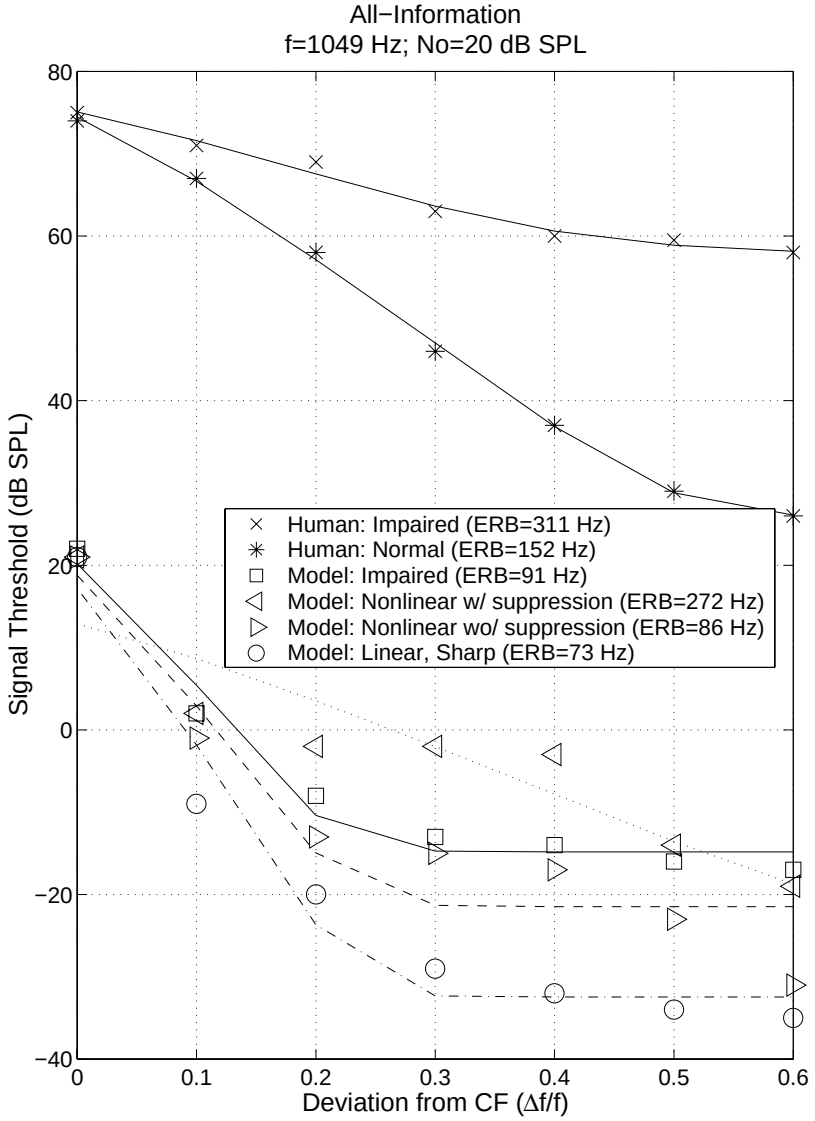

Figure 6-7: Predicted tone detection in random notched noise based on all-information (AI) information for the four versions of the AN model. Same symbols as Fig. 6-6.

the impaired model for notch widths between 0.2 and 0.5. Both of the linear versions of the AN model had an orderly progression of thresholds as notch width increased. In contrast, both nonlinear versions of the model demonstrated relatively flat thresholds between 0.2 and 0.4 notch widths, and a decrease in threshold for larger notch widths. The all-information thresholds for the 0.0 notch width are essentially the same for all versions of the AN model, consistent with external stimulus variability dominating all-information performance for this notch width, as discussed below.

The effect of random- versus fixed-noise maskers on the predicted detection thresholds is considered in Fig. 6-8. A summary of the power-spectrum-model fits to the fixed-noise threshold predictions from the four AN-model versions is provided in Table 6.3 (based on all notch widths) and in Table 6.4 (based on all notch width except 0.0). Rate-place and all-information detection thresholds for the linear-sharp AN model are shown in Fig. 6-8, in addition to the fits from the power-spectrum model. Predicted thresholds for the random-noise experiment are always higher or equal to the thresholds for the fixed-noise experiment, consistent with the potential detrimental effect of external stimulus variability on performance. There is a significant, but small difference between random- and fixed-noise thresholds for the rate-place model for notch widths less than 0.4. However, the estimated ERB for the two experiments is essentially identical for the linear-sharp AN model. The random-noise effect on estimate of ERB is much larger for the all-information model. The random-noise threshold for the 0.0 notch width is more than 20 dB higher than for the fixed-noise experiment, and thus the random-noise ERB for the all-information model is roughly one-half of the ERB estimate for the fixed-noise experiment. The random- and fixed-noise thresholds for the all-information model are very similar for all notch widths other than 0.0.

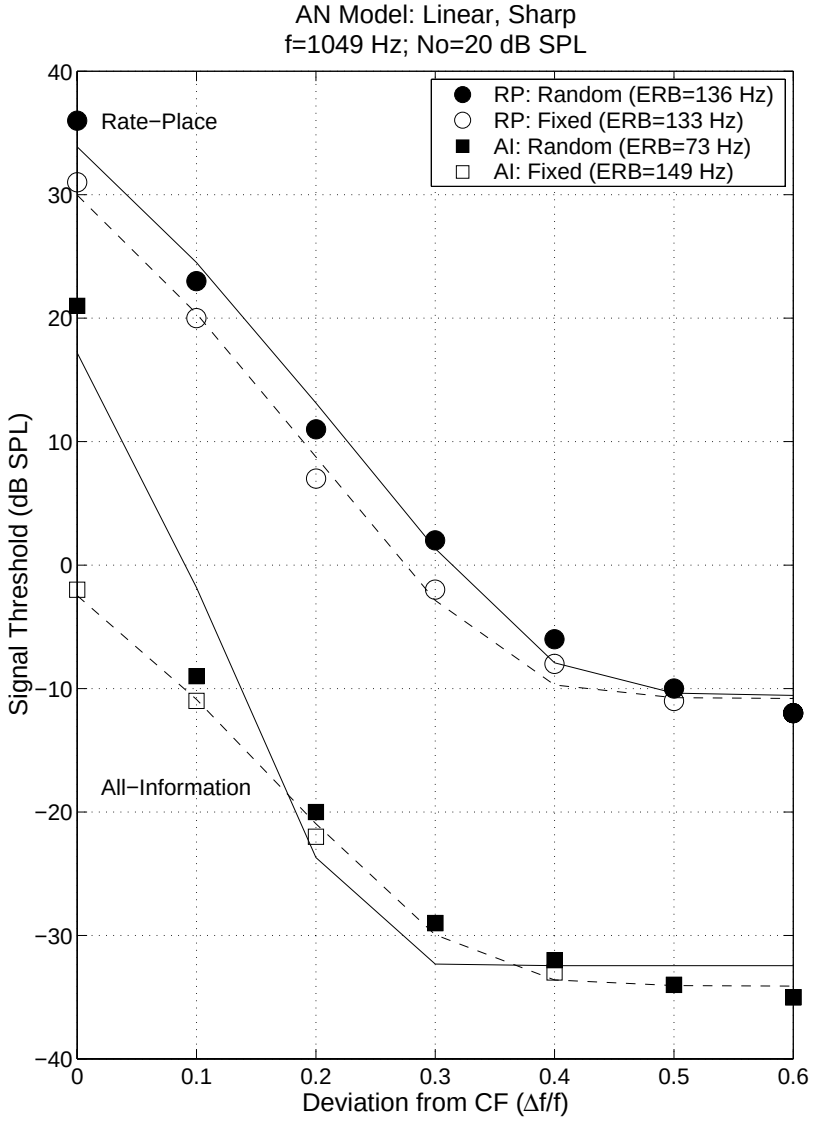

Figure 6-8: Detection of a tone in random- and fixed-noise as a function of notch width based on the *linear AN model with sharp tuning*. Predicted rate-place (RP) and all-information (AI) performance is based on the combination of the three spontaneous-rate populations (TOT curves in Fig. 6-2). Fitted thresholds (lines) and equivalent-rectangular bandwidth (ERB) estimates from the power-spectrum model are shown for random- (solid lines) and fixed-noise (dashed lines) experiments based on all seven notch widths. The ERB of the 1049-Hz CF in the AN model was 115 Hz.

Table 6.3: Auditory-filter parameters estimated from the fixed-noise thresholds predicted by the AN-model for notch widths from 0.0 to 0.6. The power-spectrum model was used to fit a roex(p,r) filter to the AN thresholds. Estimates for both the rate-place and all-information thresholds from the four versions of the AN model are shown. ERB: equivalent-rectangular bandwidth; rms: root-mean-squared error; p: filter slope; r: filter dynamic range; K: signal-to-noise at the filter output required for detection threshold.

Model	ERB (Hz)	rms (dB)	p	r	K (dB)	Model Filter ERB(Hz)
Rate-Place						
Linear, Sharp	133.0	1.16	31.5	-48.8	-11.3	115
Nonlinear, wo/Suppression	171.0	1.38	24.5	-46.9	-14.6	115-248
Nonlinear, w/Suppression	211.5	1.53	19.9	-38.2	-17.7	115-248
Impaired	247.5	0.25	17.1	-28.7	-13.6	248
All-Information						
Linear, Sharp	149.2	0.70	28.2	-39.1	-44.2	115
Nonlinear, wo/Suppression	198.3	1.01	21.2	-36.5	-46.4	115-248
Nonlinear, w/Suppression	229.4	1.58	18.4	-30.9	-47.7	115-248
Impaired	334.3	0.23	13.4	-19.4	-46.8	248

Table 6.4: Auditory-filter parameters estimated from the fixed-noise thresholds predicted by the AN-model for notch widths from 0.1 to 0.6. The power-spectrum model was used to fit a roex(p,r) filter to the AN thresholds. Estimates for both the rate-place and all-information thresholds from the four versions of the AN model are shown. ERB: equivalent-rectangular bandwidth; rms: root-mean-squared error; p: filter slope; r: filter dynamic range; K: signal-to-noise at the filter output required for detection threshold.

Model	ERB (Hz)	rms (dB)	p	r	K (dB)	Model Filter ERB(Hz)
Rate-Place						
Linear, Sharp	145.8	1.00	28.8	-46.4	-14.0	115
Nonlinear, wo/Suppression	154.6	1.00	27.1	-49.5	-11.5	115-248
Nonlinear, w/Suppression	200.2	1.61	21.0	-39.1	-16.5	115-248
Impaired	238.6	0.17	17.8	-29.3	-12.9	248
All-Information						
Linear, Sharp	158.4	0.66	26.5	-37.8	-45.7	115
Nonlinear, wo/Suppression	207.9	0.21	20.2	-35.8	-47.4	115-248
Nonlinear, w/Suppression	275.1	1.47	15.4	-28.9	-50.9	115-248
Impaired	319.3	1.05	13.9	-19.9	-46.2	248

6.5 Discussion

6.5.1 Detection in notched-noise based on auditory-nerve information

Auditory frequency selectivity is typically evaluated in terms of the power-spectrum model of masking, which is based on the signal and noise energy through a set of auditory filters. While this interpretation is likely to represent a reasonable description of the overall processing of the auditory system in certain situations, it may not be appropriate for individual auditory-nerve fibers, which often have a limited dynamic range due several saturating nonlinearities between the basilar membrane and the auditory nerve (Patuzzi and Robertson, 1988). It has been hypothesized that the wide dynamic range of human hearing is encoded across different spontaneous-rate (SR) populations of AN fibers (Delgutte, 1987; Winslow and Sachs, 1988; Viemeister, 1988a,b; Winter and Palmer, 1991).

The present study predicted detection thresholds for a tone in a notched noise masker based on the HSR, MSR, and LSR populations of AN fibers for the same conditions that are typically used to estimate auditory frequency selectivity (e.g., Glasberg and Moore, 1990). The predicted detection thresholds indicate that performance often varies across the different SR populations (Figs. 6-2 - 6-5). For the model based on average discharge rate, different SR populations can determine overall performance at different notch widths (e.g., LSR fibers for small notch widths and HSR fibers at large notch widths in Fig. 6-2), even at the low noise spectrum level used in the present study ($N_0 = 20$ dB SP). The reliance of total rate-place performance on the LSR population is stronger at higher noise spectrum levels and higher signal frequencies (not shown). Thus, the detection thresholds that are used to estimate a single auditory filter depend on different AN fibers as a function of notch width, and it is unlikely that auditory filters based on individual AN fibers would consistently correspond with those measured psychophysically.

The predicted thresholds for the rate-place model based on the total AN population are well fit by the power-spectrum model. The rms errors for the fits to the rate-place thresholds (Tables 6.1 - 6.4) are consistent with the range of rms errors reported for human listeners

using the notched-noise method. Shailer *et al.* (1990) reported rms errors that were typically between 0.9 and 2.2, and were always less than or equal to 2.7. The estimated equivalent-rectangular bandwidths (ERBs) for the rate-place model are very close to the ERB used for the 1049-Hz CF in the two linear versions of the AN model. This result is reassuring, given that the notched-noise method was designed based on the assumption of linear filters.

In contrast to the rate-place model, the fits to the all-information predictions for the random-noise experiment were poor. The rms errors for the all-information fits based on all notch widths (Table 6.1) were larger than are typically reported for human listeners. The primary source of error for the all-information fits came from the high threshold for the 0.0 notch width in the random-noise conditions, and when this notch width was excluded (Table 6.2), the rms errors were much improved and within the range of human data. The all-information threshold for the 0.0 notch width was much lower for the fixed-noise experiments, while the thresholds for the other notch widths were typically reduced less in the fixed-noise experiment. Thus, the fits to the fixed-noise all-information thresholds (Tables 6.3 and 6.4) were much better than to the random-noise thresholds, with rms errors within the range of human fits.

The high threshold for the 0.0 notch width in the random-noise conditions resulted in much too narrow filter estimates for the all-information model. The estimated all-information ERB was 73 Hz for the linear-sharp model that had a filter bandwidth of 115 Hz for the 1049-Hz CF, and 91 Hz for the impaired model that had an filter bandwidth of 248 Hz. The estimated all-information ERB's for the two linear versions of the AN model were much broader when either the 0.0 notch width was excluded from the fits, or when the power-spectrum model was fit to the fixed-noise thresholds.

The source of the high relative threshold for the 0.0 notch width is likely due to the efficient use of fine-time information by the all-information model for notch widths other than 0.0 (i.e., when there is no random noise energy at the tone frequency). It has been demonstrated by Siebert (1970) and in Chapter 2 that the amount of information in the fine-time patterns of AN discharges is much greater than is needed to explain performance by human listeners. This is consistent with the extremely low all-information thresholds for broad notch widths. When random noise energy exists at the tone frequency, the fine-time

information about the tone is much less reliable, and the all-information threshold for the 0.0 notch width is 30 dB higher than the threshold for the 0.1 notch width (Fig. 6-8). In the fixed-noise experiment where the external stimulus variability has been removed, the all-information model is able to make use of the fine-time information about the presence of the tone, and the threshold for the 0.0 notch width is less than 10 dB higher than the 0.1 notch width (Fig. 6-8). Thus, the transition between the less reliable fine-time information for the 0.0 notch width and the more reliable fine-time information for the other notch widths results in estimated ERB's that are much too narrow for the all-information model. A temporal model that did not process fine-time information efficiently would be expected to be less influenced by random-noise variations, and would likely produce more realistic detection thresholds and auditory-filter estimates.

Thus, the present analysis highlights the importance of considering the effect of external stimulus variability on detection-in-noise performance, in addition to the effect from AN-response variability. The present predictions for random- and fixed-noise experiments suggest that fine-time information for detecting tones in noise is most reliable when there is no noise energy in the same frequency region as the tone, or when the noise is not random; however, fine-time information is much less reliable when random noise is present with energy at the tone frequency. Thus, it is likely that human listeners do not make efficient use of fine time information to detect a tone in noise, because the sharp decrease in predicted all-information threshold between the 0.0 and 0.1 notch widths is not seen in human data.

6.5.2 The influence of cochlear nonlinearity on auditory-filter estimates

The computational AN model in the present study was used to quantitatively test the significance of several nonlinear response properties associated with the cochlear amplifier for estimates of auditory filters. The phenomenological model represents a unified view of how each of these nonlinear response properties relates to the cochlear amplifier mechanism. The model was implemented by matching the response properties of the model to measurements of cochlear compression and two-tone suppression tuning curves from the available physio-

logical data, often from different species. While the model response properties quantitatively match the available physiological data fairly well, it should not be assumed that the model represents the exact amount of compression or suppression in humans, which can not be measured directly. Thus, the strength of the current model is in demonstrating the effects of the presence or absence of these response properties, rather than predicting the exact size of the effects. Based on the current model, the phenomenon of compression is discussed as representing the property that stimulus energy that passes through the excitatory filter acts to compress the AN response by broadening the excitatory filter. The phenomenon of suppression is discussed as representing the property that stimulus energy well away from the CF, which may not excite the AN fiber, can suppress the response of the AN fiber to excitatory stimulus energy. The off-frequency stimulus energy acts to broaden the excitatory filter in the same way as near-CF energy due to the breadth of the suppression filter in the model control path.

The equivalent-rectangular bandwidths (ERB's) estimated for the AN model with both compression and suppression were always broader than for the model with only compression (i.e., without suppression) (Tables 6.1 - 6.4). Similarly, the estimated ERB's for the model with only compression were always broader than the model with linear, sharp filters. These relations were true for both random- and fixed-noise predictions, and for fits based on either all of the notch widths or excluding the 0.0 notch width. While the size of the effect varied across the conditions represented in Tables 6.1 - 6.4, the present predictions demonstrate that both compression and suppression can influence estimates of auditory filters based on the notched-noise method. The effects of both compression and suppression consistently result in broader estimates of tuning than the low-level filters in the AN model. This result is consistent with the finding that frequency selectivity measured psychophysically using nonsimultaneous masking is typically sharper than when measured using simultaneous masking (e.g., Houtgast, 1977; Moore, 1978; Moore and Glasberg, 1981,1986; Moore *et al.*, 1987). This finding has been interpreted as resulting from the lack of suppression from a non-simultaneous masker (see Moore, 1995a for a review).

The ERB's from the nonlinear AN model with suppression were always broader than from the impaired model in random-noise conditions, even though the "actual" filters are

narrower. This finding provides an example where it is not just the masking noise in the excitatory bandwidth of the AN filter that determines the detection threshold, but that stimulus energy below the CF can add additional excitatory masking energy in the form of nonlinear distortion (i.e., that stimulus energy near 500 Hz creates masking energy at the 1000-Hz CF due to distortion from rectifying nonlinearities in the control path). This distortion is present in both nonlinear model versions, but not in either of the linear models. The nonlinear filters are not broader than the impaired filters, but the detection of the tone is influenced by noise away from CF. Distortion can result in flat thresholds for notch widths from roughly 0.2 to 0.4 (Figs. 6-4 and 6-5), because these are conditions in which there is significant masker energy at 500 Hz. The effect is much stronger in the random-noise predictions, and is small in the fixed-noise predictions. This suggests that the distortion provides masking primarily when random distortion energy is produced at CF, and both rate-place and all-information thresholds are not strongly masked by the distortion if the distortion energy is not random. The effects of distortion are very strong for the all-information model, and have less influence on the rate-place model. This result is due to the fact that the random distortion energy at 1000 Hz greatly reduces the reliability of the fine-time information that the all-information model relies upon so strongly for notch widths greater than 0.0, i.e., conditions for which there is no noise energy at 1 kHz. The presence of nonlinear distortion in the AN model is an example of a case in which the notched-noise method is not measuring the actual filters, since detection thresholds are increased due to distortion from off-frequency masking energy.

The AN model used in the present study was implemented based on physiological compression and suppression data from chinchilla, guinea pig, and cat. The strength of the nonlinear responses properties in the AN model, including compression, suppression, and distortion, may not accurately represent the strength of human nonlinearity; however, the main effects demonstrated by the present predictions depend primarily on the presence or absence of these nonlinear properties. Future studies based on the present approach can quantitatively evaluate the strength of human nonlinearity by relating physiological response properties in the AN model to human psychophysical performance. The similar physiological and psychophysical masking studies by Delgutte (1990a) and Oxenham and Plack (1998) provide a good basis for relating the strength of compression and suppression in cats and

humans. Physiological and psychophysical measures of basilar-membrane compression (e.g., Ruggero *et al.*, 1997; Cooper and Rhode, 1997; Oxenham and Plack, 1997; Plack and Oxenham, 1998) provide useful data for relating the strength of compression in chinchilla, guinea pig, and humans.

6.5.3 Implications for the measurement of auditory tuning

It is not surprising that mechanisms that broaden the tuning of AN fibers would result in broader auditory filter estimates. The present study demonstrates this quantitatively by linking several physiological phenomena to psychophysically measured ERB's that are based on the detection of a tone in notched noise using a population of AN fibers. The primary benefit of such a demonstration is to provide a quantitative framework in which to discuss the significance of the cochlear amplifier on auditory tuning. This study also provides insight into appropriate interpretations of results from current psychophysical methods for estimating auditory filters.

The dependence of auditory tuning on stimulus configuration

In attempting to measure auditory tuning psychophysically, it is critical to understand how tuning varies as a function of the spectral and temporal content of the stimulus. The present predictions indicate that both the signal and masker influence the tuning of the auditory system, and thus it is improper to discuss *the* tuning at a given CF without specifying the stimulus configuration. This concept has been discussed previously in terms of the broadening of auditory tuning with level; however, there is continued debate regarding how to define level. Current psychophysical measurement techniques assume that tuning depends on either the level of the stimulus prior to filtering (e.g., Moore and Glasberg, 1990) or on the level of the filtered stimulus (e.g., Rosen and Baker, 1994; Rosen *et al.*, 1998). The current AN model based on physiological responses suggests that neither of these views are correct, and that auditory tuning depends on both signal and masker energy that passes through the suppression filter (i.e., both excitatory energy near CF and suppressive energy well away from the signal frequency). Thus, psychophysical experiments that use noise to

mask off-frequency information can change the tuning of the filter they are attempting to measure based on their measurement technique. An example of this effect is the difference in frequency selectivity measured using simultaneous masking (for which the signal and masker energy influence the tuning of the filter used to process the signal) and nonsimultaneous masking (for which only the signal energy influences BM tuning due to the fast dynamics of the cochlear amplifier) (e.g., Houtgast, 1977; Moore, 1978; Moore and Glasberg, 1981,1986; Moore *et al.*, 1987).

Interpretation of current psychophysical methods

The dependence of auditory tuning of both signal and masker configuration has important implications for comparing different methods of estimating auditory filters. Techniques such as the critical-band method developed by Fletcher (1953) have been criticized based on methodological issues, such as the varying influence of random-noise fluctuations as a function of bandwidth (see Chapter 5), and the assumption of rectangular filters. However, a more fundamental issue issue is whether the filter being estimated is the same in each experiment, given that different methods use different stimulus configurations. The critical-band method uses noise bands centered at the tone frequency and measures detection as a function of bandwidth (Fletcher, 1940). The notched-noise methods uses noise bands that surround the tone frequency, but only have masker energy at the tone frequency for one of many notch widths measured. The critical-ratio method measures detection in broadband noise. Each of these methods invoke the physiological mechanisms of compression and suppression in vastly different ways. Given that compression and suppression have been shown in the present study to alter auditory tuning, it is likely that these three common methods for estimating auditory filters are not measuring the same filter. Thus, differences in filter estimates across methods must be evaluated both in terms of methodological issues, but also in terms of whether the tuning that is being measured is the same across methods (e.g., Lentz *et al.*, 1999).

Beyond the issue of whether different methods are measuring the same filter, the present study suggests that tuning is likely to vary across the different conditions used within a single method to estimate auditory filter shapes (also see Sinex and Havey, 1986; Delgutte,

1990a). For example, the notched noise method method fits detection thresholds as a function of notch width based on the power spectrum model, which assumes a single filter was responsible for each detection threshold. However, given that the influence of compression and suppression depends on the spectral content of the masker, it is likely that auditory tuning varies across the notch widths used to measure a single auditory filter. Future studies using the modeling approach presented here can evaluate the significance of this effect quantitatively by calculating the bandwidth (averaged across time) of the model CF at the tone frequency as a function of the notch widths used in the notched-noise method. It is likely that the estimated filter from the notched-noise method represents some sort of average filter across notch width.

A consistent result from the notched-noise method is that estimated auditory filters become more asymmetric as stimulus level increases (Moore and Glasberg, 1987; Moore 1995a). This conclusion is based on the use of asymmetric notched-noise maskers in order to evaluate the relative contribution of masking energy above and below the tone frequency. At high levels, detection thresholds decrease more slowly when the edge of the low-frequency noise band is closer to the tone frequency than is the edge of the high-frequency noise band. This result has been interpreted using the power-spectrum model of masking as indicating that the auditory filter at high levels has a shallower slope on the low-frequency side than on the high-frequency side. However, the results from the present study suggest the influence of suppression above and below the tone frequency must also be considered. Delgutte (1990b) systematically measured the growth of two-tone suppression in AN fibers as a function of suppressor level, and found that suppression was asymmetric above and below CF. Suppression thresholds were higher below CF than above CF, but more significantly, the growth-rate of suppression was much higher below CF than above CF (up to a factor of 10 in the same AN fiber). Thus, much stronger suppression was observed below CF than above CF at high levels. This suggests that the asymmetric notched-noise maskers used to evaluate the two sides of the auditory filter would produce more suppression, and thus a broader filter, when the low-frequency noise band is closer to the tone frequency than when the high-frequency noise band is closer. The same pattern of detection thresholds measured for human listeners with asymmetric notched-noise maskers could result from symmetric

excitatory filters and asymmetric growth of suppression, and thus the asymmetry of auditory tuning can not be measured using the current notched-noise method. Thus, the reported changes in auditory-filter asymmetry as a function of level may be due to the asymmetry in the growth of suppression above and below CF, and not on actual changes in auditory-filter asymmetry.

Despite the complications that cochlear nonlinearity place on the interpretation of psychophysical measures of auditory tuning, the estimates derived from these methods have proven to be useful in predicting a variety of masking conditions (Moore, 1995a). Derleth and Dau (2000) successfully predicted masking patterns for mid-level tone and narrowband signals and maskers using a modulation-filter-bank model (Dau *et al.*, 1997a) that contained a linear peripheral filter bank based on the notched noise method (Patterson *et al.*, 1987). Derleth and Dau suggested that a linear filter bank was successful in predicting masking patterns at mid levels because the effects of cochlear nonlinearities were likely included in the filter bandwidths estimated based on the notched-noise method, and that these filter bandwidths can be considered to represent "effective" auditory filters. The present study suggests that the auditory filters estimated from the notched-noise method may represent a general approximation of the average tuning that results from a tone stimulus in the presence of mid-level masking energy near (both above and below) the frequency of tone.

However, the present predictions suggest that the filter estimates from the notched-noise method are significantly broader than the low-level tuning of AN fibers for two reasons: 1) the tone level is typically above the compression threshold, resulting in a broader filter, and 2) the noise used to mask off-frequency information has broadened the filter through suppression. Recent non-invasive measures of human cochlear tuning based on the phase of stimulus-frequency otoacoustic emissions (Shera and Guinan, 2000) have also suggested that psychophysical measures of auditory tuning overestimate filter bandwidths. The phenomenological model used in the present study suggests that stimulus conditions for which any of the noise-masker energy passes through the suppression filter result in auditory tuning that is broader than the low level tuning. The threshold for compression of the response to a CF-tone is about 20 dB SPL in the model, which is consistent with physiological measures of basilar-membrane compression (Ruggero, *et al.*, 1997). Thus, a noise masker that produces

roughly 20 dB SPL or more of overall level through the suppression filter would be expected to suppress the response to a CF-tone. If the suppression filter had a bandwidth that was on the order of 100 Hz (likely to be an underestimate given that current estimates of the auditory filter bandwidth at 1-kHz are between 100 and 150 Hz), then a noise spectrum level greater than 0 dB SPL would produce more than 20 dB SPL overall level through the suppression filter, and thus produce suppression. Even lower noise spectrum levels would produce suppression at higher frequencies for which excitatory and suppressive tuning is broader. Studies that estimate auditory-filter shapes are rarely based on masking conditions that use noise spectrum levels less than 20 dB SPL. Thus, it is likely that most psychophysical estimates of auditory filters include effects of suppression, and therefore overestimate the bandwidths of AN threshold tuning curves. Psychophysical estimates of auditory tuning typically demonstrate smaller changes in bandwidth with level than would be predicted based on the amount of cochlear amplifier gain reported from measures of BM compression. This effect may result because psychophysical measures of tuning are not made at levels low enough to avoid the effects of compression and suppression. A consequence of this limitation is that the difference between the frequency-selective capabilities of normal and impaired listeners may be underestimated using current psychophysical methods. The largest difference between estimates of frequency selectivity for normal and impaired listeners would be expected to occur for levels just above threshold, as suggested by Rosen *et al.* (1998).

Different methods are appropriate for difference stimulus conditions

The present study suggests that the most appropriate method for estimating auditory tuning depends on the stimulus configuration of interest. Simultaneous masking methods such as the notched-noise method may be appropriate for representing the average tuning of the auditory system in the presence of both signal and masker. However, simultaneous masking results in overestimates of filter bandwidth when tuning in the presence of the signal alone is of interest, due to the effects of compression and suppression in response to the noise masker. Nonsimultaneous masking can be used to estimate filters that represent the tuning in response to the signal alone, and allow the masking noise to be used as purely as an excitatory masker. However, the amount of excitatory masking produced by a nonsimultaneous masker

is dependent on the tuning that results from the particular masker used, and thus may also be influenced by the effects of compression and suppression in response to the noise masker alone. A similar idea was used to model the additivity of nonsimultaneous masking based on independent effects of compression for the signal and maskers (e.g., Oxenham and Moore, 1994). Thus, estimates of auditory tuning using nonsimultaneous masking represent filters that depend primarily on the signal, but may be broader than low-level filters due to the effects of compression in response to the signal.

6.6 Conclusions

1. The cochlear amplifier alters the tuning of the normal auditory system based on the spectral and temporal configuration of the stimulus (e.g., both signal and masker). Thus, no single set of linear filters derived from the power-spectrum model of masking is appropriate to represent the frequency selectivity of the auditory system for all stimulus conditions.

2. Estimates of auditory frequency selectivity using simultaneous masking represent some sort of average tuning in the presence of both the signal and the masker, and likely overestimate the bandwidth of tuning in response to the signal alone due to compression and suppression.

3. Nonsimultaneous masking is likely to provide a more accurate estimate of tuning in response to the signal alone. However, compression and suppression may affect the signal and masker responses independently, and thus must be considered in interpreting estimates of frequency selectivity based on nonsimultaneous-masking.

4. The typically reported increase in auditory-filter asymmetry as level is increased is consistent with the asymmetric growth of suppression above and below CF. The use of asymmetric notched-noise maskers to estimate auditory-filter asymmetry is likely too be strongly biased by the effects of suppression, especially in simultaneous-masking conditions.

6.7 Acknowledgments

We thank Xuedong Zhang and Ian Bruce for their collaboration on the auditory-nerve model. Supported by NIH Grants T32DC00038 and R01DC00100, and NSF Grant 9983567. The simulations in this study were performed on computers provided by the Scientific Computing and Visualization group at Boston University.

Chapter 7

Summary and Comments

The motivation for this dissertation was the fact that there are still many situations for which hearing aids have a limited benefit, despite a rapidly advancing knowledge of peripheral auditory processing. There is now strong evidence that many nonlinear response properties observed in the normal auditory periphery originate from a single mechanism, referred to as the cochlear amplifier, which is damaged or absent in many forms of sensorineural hearing loss. Thus, the primary goal of this dissertation was to quantitatively relate physiological response properties associated with the cochlear amplifier to psychophysical performance. The observation that hearing-impaired listeners with amplification often demonstrate near-normal performance in simple listening tasks necessitates the ability to quantitatively relate physiological response properties to psychophysical performance in more complex tasks than previous methods allowed. Thus, a main focus of this dissertation was generalizing methods from statistical decision theory to allow current AN models that include the most significant properties associated with the cochlear amplifier to be used to analyze a broader range of psychophysical experiments.

The general modeling approach developed in this dissertation allows 1) any AN model to be used, 2) optimal performance based on temporal and/or average rate information to be predicted, 3) performance based on sub-optimal processors to be predicted, and 4) psychophysical experiments that use random-noise stimuli to be analyzed. Two psychophysical tasks were evaluated in this dissertation for which the cochlear amplifier is significant. To evaluate differences between normal-hearing and hearing-impaired listeners, the set of phys-

iological properties associated with the cochlear amplifier must be considered as a whole. However, the modeling approach used in this dissertation allows additional insight to be gained by parsing the relative contributions of individual response properties to psychophysical performance. This general modeling approach is applicable to a wide range of issues related to hearing, and can easily be used with future AN models that incorporate more detailed descriptions of AN response properties. In addition, these methods are applicable to any sensory system for which a description of the variation in neural discharge statistics with stimulus parameters is available. The sections that follow discuss specific issues addressed in this dissertation.

7.1 Separating the statistical decision theory analysis from the AN model

Chapter 2 described a method for using an arbitrary AN model (i.e., computational or analytical) with statistical decision theory (SDT) to analyze the temporal and average-rate information available in the AN for any discrimination task in which a single parameter is varied. Siebert (1965, 1968, 1970) and Colburn (1969, 1973, 1977a,b, 1981) analyzed various discrimination experiments using SDT with analytical AN models; however, their SDT analysis was tied directly to the specific AN models they used. Chapter 2 described the parts of Siebert's (1970) analysis using the Cramér-Rao bound that are independent of the AN model when Poisson discharge statistics are assumed. The result was an expression that related the time-varying discharge rates $r_i(t)$ provided by the AN model to the just-noticeable difference in the stimulus parameter of interest for the psychophysical experiment. It was also demonstrated in Chapter 2 that the same result is obtained using a likelihood-ratio test. This analysis provided insight into the expression for the amount of information available in each AN fiber, i.e., that this expression is closely related to the common sensitivity index d'. The expression for the theoretical performance limits can be evaluated using computational AN models by using a simple method for approximating the partial derivative of the discharge rate with respect to the stimulus parameter of interest.

The separation of the SDT analysis from the AN model makes this approach extremely general in that it can be used with any AN model given that Poisson discharge statistics are assumed. As future AN models are developed with more physiological detail, this approach can be used without a rederivation of the SDT analysis. The ability to use computational models allows AN responses to arbitrary stimuli to be predicted, and thus a much broader range of psychophysical tasks to be evaluated. Computational AN models have been used previously to develop theories for the psychophysical significance of complex physiological phenomena; however, these models have typically only been tested against physiological data. The current approach provides a link between complex physiological properties that have been characterized in animals and psychophysical measures of performance in humans. A specific benefit of using computational AN models is the ability to make performance-limit predictions for short-duration stimuli because neural-adaptation and transient responses are incorporated in the computational AN model. Chapter 2 highlighted several extensions of the stimulus parameter space over which frequency- and level-discrimination predictions could be made compared with Siebert's (1968, 1970) analytical AN model.

7.2 Analysis of psychophysical experiments with random variation of a single parameter

The analysis developed in Chapter 3 represents a first step in extending the SDT analysis used by Siebert and Colburn to include psychophysical tasks that use random stimuli. The SDT analysis was extended to include one-parameter discrimination tasks in which another single parameter is randomized. The randomization of certain stimulus parameters is often used in psychophysical experiments to limit the cues available to the listener in order to test theories of auditory processing [(e.g., profile-analysis experiments in which overall level is randomized to test energy-detector models (Green, 1988)].

The primary theoretical result of this analysis was an intuitive equation that expressed the information available in the presence of random stimulus variation as the difference between the average fixed-stimulus information and the amount of information that was lost due

to the random stimulus variation. This expression was used to relate optimal performance based on temporal and/or average-rate information in the AN population to the time-varying discharge rates provided by any AN model. The results in Chapter 3 show that frequency-discrimination predictions from both the all-information (based on AN discharge times) and the rate-place (based on AN discharge counts) models are unaffected by random level variation, consistent with human performance. Significant temporal information exists in individual AN fibers that is unaffected by random level variation. In contrast, the rate-place model must rely on the AN population response to avoid the influence of random level variation, i.e., by comparing AN fibers with characteristic frequencies (CF's) above and below the tone frequency. The results from Chapter 3 provide insight into the appropriate interpretation of random-level frequency discrimination experiments, and provide additional constraints on rate-place models of auditory encoding (discussed in further detail below).

7.3 Analysis of psychophysical experiments with random-noise stimuli

Many influential psychophysical experiments that have been used to develop theories for auditory processing have used random-noise stimuli. However, the interpretation of these experiments is often based on assumptions about the effect of noise maskers on signal information in the auditory system. One effect of noise maskers that is often not considered is the influence of random stimulus variations, which can limit psychophysical performance by contributing to the variance of AN discharge times and counts.

Previous psychophysical studies have evaluated the influence of random stimulus variations (external noise) by comparing performance with random- and fixed-noise maskers (de Boer, 1966; Spiegel and Green, 1981; Siegel and Colburn, 1983, 1989; Buus, 1990; Dau *et al.*, 1996b). Fixed-noise performance is interpreted as representing only the influence of physiological variations (internal noise), while the random-noise results represent the influence of both internal and external noise. These results have been used with several additional assumptions to estimate the ratio of internal and external variance in the listener's decision

process, without making any specific assumptions about the processor or the physiological source of internal noise (e.g., Spiegel and Green, 1981; Siegel and Colburn, 1983, 1989). Other psychophysical studies have evaluated external- and/or internal-noise effects on the performance of specific decision models (e.g., de Boer, 1966; Dau *et al.*, 1996b). In these studies, the internal noise is estimated by matching model and human performance in conditions for which internal noise is assumed to dominate, but is never compared with known physiological sources of internal noise. Colburn (1969, 1977a,b) and Siebert (1973) quantitatively compared the effects of physiological internal noise from AN discharges and external stimulus variations on psychophysical performance; however, their use of analytical AN models required simplistic assumptions about AN responses to noise. The new theoretical approach developed in Chapter 5 allows computational AN models to be used with statistical decision theory to quantify the effects of both internal and external noise on temporal and/or average-rate information in the AN.

The fundamental theoretical result from Chapter 5 is an equation for the sensitivity of a general processor that discriminates two stimulus alternatives in a psychophysical experiment with random-noise stimuli based on the population of AN discharges. The equation relates the time-varying discharge rates provided by the AN model to the sensitivity in terms of the mean-difference between the decision variable in the two stimulus alternatives and the variance. The total variance is separated into the contributions of the random nature of AN discharges (internal noise) and the random stimulus fluctuations (external noise). This separation allows predictions to be made for both random- and fixed-noise conditions. Performance can be evaluated based on different assumptions about the *a priori* information used by the processor, e.g., whether information about individual noise waveforms is used or only information about the noise-ensemble properties (more realistic for human listeners). In addition, Chapter 5 provides equations that quantify the total information available in individual AN fibers, as well as the information available as a function of time.

Chapter 5 provided several examples of the application of this analysis to the detection of a tone in a bandlimited noise masker. The effect of external noise was shown to be larger for a narrowband masker than for a broadband masker, as expected, for both the all-information and rate-place models. The correlation across AN fibers that results from

random-noise stimuli was shown to significantly decrease the benefit obtained by combining temporal or average-rate information across the population of AN fibers in many conditions. This strong effect of correlation occurs for conditions in which the external-noise variance is similar to or larger than the internal-noise variance. An interesting example of this effect was seen in the distribution of AN information across CF (information profile) for a narrowband masker. The shape of the information profile changed dramatically based on the number of AN fibers that were assumed to be represented by each model CF. The benefit of combining across N AN fibers with CF's well outside the noise spectrum (i.e., dominated by internal noise) demonstrated an increase in sensitivity that was close to the expected factor of $\sqrt{N}$ that comes from independent fibers. A much smaller benefit was observed for CF's within the noise spectrum because the responses of these AN fibers were dominated by the external noise. The complex effects of random-noise stimuli on temporal and average-rate information in the AN warrant future study. This new theoretical approach represents a significant extension of previous approaches, and greatly expands the types of psychophysical experiments that can be evaluated in terms of physiological response properties. In addition, the present approach provides a method for evaluating many of the assumptions made in psychophysical studies that estimate the relative amounts of internal and external noise because the effects of internal and external noise on AN information can be quantified directly.

7.4 Signal processing capabilities of monaural coincidence counters

Coincidence detection has long been considered to be important for binaural processing in the auditory system (e.g., Jeffress, 1948; reviewed by Colburn, 1996). Recently, monaural, across-frequency coincidence detection has been proposed by Carney (1994) as a physiologically realistic mechanism to decode the level-dependent spatio-temporal patterns in AN discharges that result from nonlinear phase cues. Coincidence detection is a general neural mechanism in that any neuron that receives sub-threshold inputs and has a limited integration window can be viewed as a coincidence detector. In addition, several neural

response types in the antero-ventral cochlear nucleus have properties that are consistent with monaural coincidence detection. Chapter 4 presented a simple monaural coincidence-counting model that received two AN inputs, similar to the binaural coincidence model used by Colburn (1969, 1977a,b). Expressions for the expected value and variance of the number of coincidence counts were derived based on the Poisson statistics of the AN discharge times. These expressions were given in terms of the time-varying discharge rates provided by the AN model, and thus could be used with any AN model, i.e., analytical or computational.

Predicted performance based on a population of coincidence counters was derived in Chapter 4 using an analytical AN model. Monaural coincidence detection was shown to be a robust mechanism for decoding both temporal (nonlinear phase and synchrony) and average-rate information. The coincidence-counting mechanism was shown to encode changes in stimulus level over the entire dynamic range of hearing at both low and high frequencies, and to provide a good match to the trends and absolute performance levels of human listeners. A close match to human performance levels implies that all processing at higher levels in the system must be optimal. However, the absolute performance levels based on the implementation of the coincidence mechanism in this dissertation are likely to be conservative. Coincidence performance could be improved by assuming a coincidence window that was wider than 10-μs; the analysis in Chapter 4 only requires the coincidence window to be significantly narrower than the period of AN-response oscillation. In addition, coincidence performance could be improved by increasing the number of coincidence counters by assuming that each AN fiber innervates more than one coincidence counter. This assumption would violate the independence assumption used to calculate coincidence performance, but would increase the amount of information available in the coincidence population. The present study provides quantitative support that coincidence detection is important for robust monaural encoding of sound level, and provides simple methods to analyze coincidence performance for a wide variety of psychophysical tasks.

7.5 Limitations of AN models used in this dissertation

The previous sections in this chapter have primarily focused on the significance of the methodological issues addressed in this dissertation, and were essentially independent of the details of the AN models used. The following sections of this chapter address issues that do depend on the AN models, and thus it is appropriate to first discuss several limitations of the AN models used in this dissertation.

The AN models are based on physiological data from various species by necessity, and it is assumed that human physiological responses do not differ from these other species in significant qualitative or quantitative ways. This assumption is typically difficult to test; however, the general modeling approach used in this dissertation may provide new methods for testing this assumption, as discussed in the future work section.

The strength of compression in the AN models at high frequencies was based on data from guinea pig and chinchilla. The 60-dB of cochlear-amplifier gain produced "straight" rate-level curves for the low-spontaneous-rate (LSR) AN fibers in the model, consistent with their observation in guinea pig (Winter and Palmer, 1991). Straight rate-level curves have not been observed in cat (Sachs and Abbas, 1974), consistent the inability of LSR fibers with CF's within 1% of 8 kHz to provide robust encoding of high sound levels in cat (Winslow and Sachs, 1988). These results suggest that cats may have less gain from the cochlear amplifier; however, large nonlinear phase shifts (up to $\pi/2$, and similar to high-CF BM responses from other species) have been observed in low-CF AN fibers in cat (Anderson *et al.*, 1971). Results from human temporal bone studies suggest that OHC loss is correlated with 50-60 dB increases in threshold (Schuknecht, 1994). Also, human psychophysical studies have estimated BM compression similar to that observed in chinchilla (Oxenham and Plack, 1997). Thus, further studies are necessary to verify the appropriate strength of compression at high frequencies to use in the human AN model; however, the decrease in strength of compression with decreases in CF is supported by psychophysical studies (Hicks and Bacon, 1999a).

The two-tone suppression tuning curves of the computational AN model match well to tuning curves from cat (Delgutte, 1990b). However, the growth of suppression in the model

is symmetric above and below CF, in contrast to the steeper growth rate below CF observed in cats (Delgutte, 1990b). Thus, while the qualitative effect of broadband suppression is captured by the AN model, the strength of suppression effects predicted by the model should be viewed with some caution.

Several response properties are not included in the AN models used in this dissertation. The effects of external- and middle-ear filtering (see Rosowski, 1996 for a review) are not considered. The AN models are primarily based on physiological data collected from animals under anesthesia. Thus, the effects of the olivocochlear efferent system are not considered in the AN models, but could potentially have played a role in the psychophysical data that were compared to model predictions. While it is clear that the efferent system can modulate the strength of the cochlear amplifier (Murugasu and Russell, 1996; see Guinan, 1996 for a review), there is currently no model for how this system works as a function of arbitrary stimuli that is sufficient to address the psychophysical significance of the efferent system. Scharf *et al.*, (1994) reported that a human patient with unilaterally sectioned olivocochlear efferents demonstrated no effect of the efferent system on performance in experiments on tone detection, level discrimination, and frequency discrimination in quiet and in noise (i.e., the experiments evaluated in this dissertation).

Several types of complex irregularities have been reported at high levels in AN responses to tones (e.g., Kiang *et al.*, 1969; Liberman and Kiang, 1984; Kiang, 1984, 1990; Ruggero *et al.*, 1996) and clicks (Lin and Guinan, 2000). These irregularities include sharp dips in rate-level curves, rapid phase transitions, and bimodal times of excitation within a stimulus period (i.e., peak splitting). These effects are typically concurrent, and occur most often between 80-100 dB SPL. While there is now clear evidence that these complexities occur in many species, the effects vary within and across species and there is currently no systematic view of how these properties vary across CF (Ruggero *et al.*, 1996) or in response to complex stimuli. A two-drive model of tone response phase reversals has been developed by Goldstein (1990, 1991, 1995), and extended to clicks by Lin and Goldstein (1995); however, this model does not include level-dependent phase shifts associated with broadened tuning, and has not been tested for arbitrary stimuli. The psychophysical consequences of these irregularities have not been considered in detail; however, the general modeling approach developed in

this dissertation provides a method to evaluate their significance. Depending on the assumed processor, these complex response properties could potentially provide either a benefit or a distraction for level encoding. The AN models used in this dissertation do not include these irregularities, but do include many significant response properties associated with the cochlear amplifier. The properties included in the AN models occur both below and above the levels at which the complex behavior is observed. Thus, the sources of information predicted by these models at high levels (e.g., relative phase cues across CF) are likely to be present in the normal auditory system in addition to the irregularities. Human listeners demonstrate roughly constant level-discrimination performance across the level range over which these irregularities occur (Florentine *et al.*, 1987). Thus, it is unclear whether these irregularities in AN responses are psychophysically significant; however, future work with models that better represent high-level AN responses is required to evaluate this issue quantitatively.

7.6 Effects of compression and suppression on psychophysical estimates of auditory filters

The theoretical approach developed in Chapter 5 was used to quantitatively relate physiological response properties to a fundamental method used in psychophysics to estimate auditory frequency selectivity, the notched-noise method. The computational AN model used in Chapter 6 (Appendix A, and Zhang *et al.*, 2000) provided a phenomenological tool to evaluate the relative effects of compression and suppression on psychophysical estimates of auditory filters. Auditory filters estimated from notched-noise detection thresholds were compared from four versions of the AN model that differed in how their tuning characteristics depended on the stimulus. It was consistently found that equivalent-rectangular bandwidths (ERB's) from the model with both compression and suppression were broader than ERB's from the model with compression and without suppression, which were broader than ERB's from the model with linear sharp tuning.

This result suggests that both compression and suppression play a role in the notched-noise method, and that psychophysical estimates of human auditory filters are likely to

be broader than AN-threshold tuning. The increased strength of compression and suppression with CF suggests that the degree to which AN-threshold tuning is overestimated may be larger at high CF's, consistent with results from recent measurements of otoacoustic-emissions (Shera and Guinan, 1999b, 2000). The cochlear amplifier controls BM tuning based on the spectral and temporal configuration of the stimulus (both signal and masker). This physiological property suggests that the excitatory filter being estimated in the notched-noise method is likely to be changing as the notch width of the masker is varied. Thus, auditory filters estimated from simultaneous masking experiments (e.g., Glasberg and Moore, 1990; Rosen *et al.*, 1998) are likely to represent a measure of average tuning in the presence of both the signal and masker, and likely overestimate the tuning bandwidth in response to the signal alone. The fast dynamics of BM tuning suggest that suppression of the signal by the masker does not occur in non-simultaneous masking experiments, consistent with sharper frequency selectivity estimated in such conditions (e.g., Moore and Glasberg, 1986; Moore *et al,*, 1987). This result suggests that estimates of frequency selectivity from non-simultaneous masking experiments represent a measure of tuning in the presence of the signal alone. The asymmetric growth of suppression in AN fibers (Delgutte, 1990b) suggests that the increase in asymmetry of psychophysically estimated AN filters with level is, at least partially, due to the effects of suppression rather than changes in asymmetry of the excitatory filter.

This analysis of the notched-noise method represents one example of the application of the general modeling approach developed in this dissertation to psychophysical tasks that use random-noise stimuli. While this experiment was chosen due to its relative simplicity, it is clear that many complex issues influence basic psychophysical masking studies. This modeling approach provides a quantitative method to investigate many of these issues in future studies.

7.7 Temporal and average-rate information in the auditory nerve

A general approach used throughout this dissertation was to quantify the total information from all AN fibers in a given population based on the set of discharge times (all-information model) or the set of discharge counts (rate-place model). The ability of average-rate and temporal information in the AN to account for human performance has been compared in terms of both trends versus stimulus parameters and absolute performance levels.

Absolute performance levels based on rate-place information were typically closer to human performance than the all-information predictions. Rate-place performance for both frequency discrimination and detection in noise was generally close to human performance, while all-information performance was at least two orders of magnitude better. Level-discrimination performance was generally more similar between the rate-place and all information models, and was roughly an order of magnitude better than human performance.

Despite the better match between rate-place and human absolute performance levels, there were several discrepancies between the trends in human and rate-place performance that are difficult to resolve. Rate-place predictions did not match the trend in human frequency-discrimination performance as a function of frequency for frequencies above 2 kHz. The small effect of random level variation on human frequency-discrimination performance with a high-frequency noise masker was inconsistent with the large effect predicted for the rate-place model when the noise was assumed to eliminate all of the information in CF's above the tone. This is a common assumption that is used to interpret noise-masking experiments; however, this assumption should be tested in future studies with the theoretical method developed in Chapter 5 for analyzing random-noise experiments. The trend in level-discrimination performance based on rate-place information in AN fibers with a narrow range of CF's was inconsistent with robust human performance at low frequencies.

All-information performance was always predicted to be significantly better than human performance, but often showed similar trends to human performance, especially in the conditions where the rate-place trends were inconsistent with human performance. All-information frequency-discrimination performance matched the trend in human performance as a func-

tion of frequency up to 10 kHz, and was predicted to be unaffected by random level variation in the presence of high-frequency noise. Similarly, all-information level-discrimination performance based on a narrow range of CF's matched the robust human performance at high levels and low frequencies, because nonlinear phase cues were the only source of information in the AN model about changes in level.

The trends in all-information performance that did not match human performance typically resulted from the assumption that the AN fine-time information is used optimally. The much too rapid improvement in all-information frequency-discrimination performance as a function of duration results from the ability of the optimal processor to make temporal comparisons between all AN discharges. The large difference between predicted all-information performance in random- and fixed-noise maskers with energy at the tone frequency is inconsistent with human performance (e.g., both noise bandwidths considered in Chapter 5, and the 0.0 notch-width condition in Chapter 6). There is a large amount of fine-time information available in the absence of random-noise energy at the tone frequency (e.g., fixed-noise or off-frequency masking conditions); however, the fine-time information is much less reliable in on-frequency masking conditions. However, significant temporal information exists in the presence of random on-frequency maskers because all-information performance is always significantly better than rate-place and human performance.

Thus, all predictions from this dissertation suggest that if monaural temporal information is used to encode frequency or level, then the AN discharge times are processed by an inefficient (sub-optimal) mechanism. Similar conclusions were reached by Siebert (1970) for monaural frequency-discrimination at low-frequencies and by Colburn (1969, 1973, 1977a,b) for the processing of interaural-time differences in the binaural auditory system. The predictions in Chapter 4 demonstrated that a coincidence-counting mechanism uses the information in the set of AN discharge times inefficiently, but maintains the trends in all-information level-discrimination performance. Predicted performance based on this general, physiologically realistic processor was shown to match both the trends and absolute levels of human level-discrimination performance as a function of stimulus level at both low and high frequencies.

However, the coincidence-detection mechanism used in this dissertation could not ac-

count for human frequency-discrimination performance as a function of both frequency and duration (not shown). Predicted performance based on a population of coincidence counters correctly matched the trends in human performance versus duration (similar to the rate-place model), but demonstrated a much shallower U-shape dependence on frequency than human performance. When performance was predicted based on the set of coincidence times, an appropriate U-shaped frequency dependence was predicted; however, the improvement in performance as a function of duration was much too steep (similar to the all-information model). Thus, the same rate-place/all-information dilemma exists for frequency-discrimination performance based on the information in a population of coincidence detectors as exists based on the information in the AN population. It was suggested in Chapter 2 that a restricted temporal mechanism that made comparisons of all AN discharges within a narrow window (roughly 10-20 ms) might account for the dependence of frequency-discrimination performance as a function of both frequency and duration. This type of restricted temporal mechanism is slightly more general than a first-order-interval processor, which has been predicted to account for both the frequency and duration dependence of human frequency-discrimination performance (Goldstein and Srulovicz, 1977).

Thus, the predictions from this dissertation suggest that temporal information is significant for frequency-discrimination at both low and high frequencies, but only significant for level-discrimination at low frequencies. This discrepancy raises the question of whether the same mechanism is used to process AN discharge times for the encoding of both level and frequency. Results from this dissertation suggest that different inefficient (sub-optimal) processors of AN discharge times are required to account for both level and frequency discrimination; however, future studies on the signal processing capabilities of inefficient temporal processors are required to investigate this issue more thoroughly.

7.8 Psychophysical significance of the cochlear amplifier

The general modeling approach developed in this dissertation was applied to two psychophysical tasks for which the cochlear amplifier is significant. These initial applications provide some insight into benefits that the cochlear amplifier provides to normal-hearing listeners, and the general difficulty in measuring significant deficits in performance of hearing-impaired listeners on simple tasks. The discussion of hearing impairment in this dissertation is applicable to forms of sensorineural hearing loss that involve damage or loss of the cochlear amplifier. While this type of loss is likely to be common in humans (Pickles, 1988; Patuzzi *et al.*, 1989; Ruggero and Rich, 1991; van Tasell, 1993; Moore, 1995b), there are many additional types of hearing loss (e.g., conductive, sensorineural based on IHC loss, retro-cochlear) that may occur in isolation or in combination with damage to the cochlear-amplifier.

7.8.1 Normal hearing

Several potential benefits of the cochlear amplifier were apparent from the results of this dissertation. The predictions from Chapter 4 suggest that the cochlear amplifier produces a wide dynamic range in AN fibers with a narrow range of CF's, a property that is useful for wideband complex stimuli (e.g., speech) for which spread of excitation is not beneficial. Both nonlinear gain and nonlinear phase cues from the cochlear amplifier are required for the robust encoding of sound level within a narrow range of CF's. Nonlinear phase cues contributed the only useful information at high levels and low frequencies, and produced robust level encoding at low frequencies. However, the nonlinear phase cues were not useful for encoding tone level at high frequencies due to the decrease in AN phase-locking at high frequencies. Nonlinear gain cues provided the only useful information at high levels and high frequencies, and accounted for robust encoding of sound level at high frequencies. However, the nonlinear gain cues do not encode high sound levels at low frequencies due to the reduced cochlear-amplifier in the apex of the cochlea. The ability of a narrow-CF region to produce Weber's Law over the entire dynamic range of hearing is often assumed

in psychophysical models that are successful in explaining many psychophysical phenomena (e.g., Florentine and Buus, 1981; Dau *et al.*, 1996a,b, 1997a,b). The present study provides a physiological basis for this assumption based on the effects of the cochlear amplifier on both the temporal and average-rate information in the AN. Loss of the benefit provided by the cochlear amplifier for extending the dynamic range over which changes in sound level can be encoded in a narrow range of CF's may be related to the difficulties that hearing-impaired listeners have in some complex environments. An additional potential benefit of the cochlear amplifier is the dynamic variation in tuning at each CF, which may be related to the severe difficulties hearing-impaired listeners have in understanding speech in temporally and spectrally varying backgrounds (e.g., Moore, 1995b; Peters *et al.*, 1998; Moore *et al.*, 1999a). This issue of dynamic tuning in the normal auditory system is discussed further below.

Several conditions were evaluated in this dissertation for which the cochlear amplifier provided little or no benefit to psychophysical performance. The predictions from Chapter 4 suggest that the cochlear amplifier has only a small effect on level-discrimination performance in quiet, i.e., when AN information in the entire range of CF's is considered. The only effects of the cochlear amplifier on predicted performance were subtle, but were consistent with significant effects often reported for human performance. The "near-miss" to Weber's Law reported in human data at low frequencies was only predicted by the AN-model with nonlinear phase cues, while the linear AN model predicted Weber's Law across a wide level range. The nonmonotonic behavior of human level-discrimination performance at high frequencies (i.e., the "mid-level bump", Florentine *et al.*, 1987) was only predicted by the AN model with nonlinear gain responses. In addition, predictions from Chapter 6 (as well as other predictions not shown) suggest that the cochlear amplifier may have little effect on the detection of tones in broadband noise. These small predicted effects of the cochlear amplifier are consistent with the lack of significant deficits in hearing-impaired performance on simple tasks (e.g., Glasberg *et al.*, 1987; Florentine, *et al.*, 1993; Moore, 1995b).

While the cochlear amplifier is often viewed in terms of the potential benefit it provides, there are conditions for which the cochlear amplifier can diminish psychophysical performance. The mid-level bump predicted in Chapter 4 represents diminished level-

discrimination performance that results from the compressive BM magnitude responses associated with the cochlear amplifier. These predictions are consistent with reports that hearing-impaired listeners can demonstrate better-than-normal performance in some conditions when comparisons are made at equal-sensation-levels (e.g., Florentine *et al.*, 1993).

7.8.2 Differences in normal and impaired auditory systems

Despite all that is known about the physiological mechanisms associated with hearing loss, there is still not a thorough understanding of why hearing aids are of limited benefit to persons with sensorineural hearing loss. Moore (1995b) has described two opposing views of why hearing impaired listeners have difficulty understanding speech even with amplification. Some researchers believe that audibility is the primary factor, while others believe that deficits in supra-threshold discrimination play a major role. There are also strongly opposing views of whether this problem can be solved. While new hearing-aid algorithms continue to be developed, such as automatic gain control and frequency selective gain (Moore, 1995b), Plomp (1978) has described a pessimistic view of the limited benefit of hearing aids due to difficulty in correcting for distortion losses that can often accompany linear attenuation losses. A more detailed understanding of the psychophysical significance of the physiological differences between normal and impaired auditory systems will provide valuable guidance for determining which approaches have the most potential to improve the ability of hearing-impaired listeners.

The preliminary applications of the modeling approach developed in this dissertation suggest a hypothesis for the significance of the cochlear amplifier in terms of comparing normal-hearing and hearing-impaired performance on psychophysical tasks. The cochlear amplifier appears to benefit normal-hearing listeners by enhancing the signal information in local frequency (and temporal) regions of complex stimuli in several distinct ways. While hearing-impaired listeners lose the benefit of these specific enhancements, they are often able to perform at near-normal levels because significant signal information exists in forms that are not degraded by hearing loss. Only in situations in which all forms of signal information are degraded is there a measurable effect of hearing loss, e.g., with complex stimuli in noise.

The results from this dissertation suggest a more complex view of the differences be-

tween normal-hearing and hearing-impaired auditory systems than is often used to interpret psychophysical studies. The computational AN model used in Chapters 5 and 6 provided an implementation of how the cochlear amplifier dynamically alters the tuning within the normal peripheral auditory system based on the spectral and temporal configuration of the stimulus. This phenomenological model highlights that the difference between a normal-hearing listener with a cochlear amplifier and a hearing-impaired listener without a cochlear amplifier is not simply a difference between one system in two states, i.e., with low or high thresholds, narrow or broad tuning, or compressive or linear magnitude responses. Rather, the difference is between two completely different systems, one that is static with insensitive and broad tuning, and one that has fast, dynamic tuning that is continuously changing (cycle by cycle at low frequencies) in response to the stimulus. This view is different than the view that has arisen from the power-spectrum model of masking, i.e., that the primary differences in the impaired system are higher thresholds and broader tuning.

The difference between these two views is likely to be most significant for rapidly changing complex stimuli, and may be less significant for steady-state stimuli. Peters *et al.* (1998) and Moore *et al.* (1999a) have reported that the difference between the ability of normal-hearing and hearing-impaired listeners to understand speech is much larger in the presence of temporally and spectrally varying backgrounds than for stationary broadband backgrounds, even when advanced amplification algorithms were provided to the hearing-impaired listeners. This result has been interpreted as representing a deficit in the ability of hearing-impaired listeners to "listen in the dips" (e.g., Peters *et al.*, 1998; Moore *et al.*, 1999a). This deficit has been suggested to result from degraded frequency selectivity and reduced temporal resolution, which have been accounted for in terms of broadened tuning and loss of compression, respectively (see Moore, 1995b, and Moore and Oxenham, 1998, for reviews). Thus, the absence of the dynamic tuning provided by the cochlear amplifier appears to be significant for stimulus conditions in which current amplification algorithms have the most limited benefit, and the present modeling approach provides a method to investigate these effects further in future studies.

7.8.3 Implications for hearing-aid strategies

The present study provides some insight into the potential benefits and limitations of hearing aids, which ideally would reproduce the normal information in the AN of an impaired ear. The phenomenological models used in this dissertation suggest that it is impossible to perfectly replicate the normal information in all AN fibers because the dynamic processing of the cochlear amplifier depends on the CF in addition to the spectral and temporal configuration of the stimulus. This suggestion is consistent with Plomp's (1978) view of a distortion loss that cannot be overcome by hearing aids. In terms of the cochlear amplifier, Plomp's distortion loss can be considered to correspond to an absence of nonlinearity in the impaired ear, rather than an additional source of nonlinearity. A more optimistic view of the potential benefit of hearing aids is suggested by the predictions from this dissertation that there are many types of information in the AN that can produce near-normal performance without the cochlear amplifier. Thus, even though much of the processing provided by the cochlear amplifier to enhance local signal information in the normal auditory system cannot be recreated exactly with preprocessing, the real requirement for a hearing-impaired listener may be to enhance the local signal information to a significant level based on any form of information, even if it is not the exact amount or type of signal information used by the healthy ear.

The fast, stimulus-dependent tuning associated with the cochlear amplifier suggests that preprocessors should also vary their processing based on the temporal and spectral configuration of the stimulus. There are advanced hearing-aid algorithms that perform this type of processing, e.g., fast-acting, multiple-band compression algorithms; however these algorithms have a limited benefit for speech in temporally and spectrally varying backgrounds (e.g., Moore *et al.*, 1999a; reviewed by Moore, 1995b). Compression has been used in hearing aid algorithms to overcome the effect of loudness recruitment, which is thought to be associated with the lack of compression in damaged cochleae (Moore, 1995b). However, the cochlear amplifier performs compression in the normal auditory system through broadband-stimulus-dependent changes in tuning that produce compression for near CF energy, level-dependent phase shifts, and broadened tuning with increases in level. Thus, a large amount of poten-

tially important processing that occurs in the normal auditory system is not provided by the magnitude compression schemes that have been implemented so far, even the fast-acting, multiple-band algorithms. For example, none of the hearing-aid algorithms to date attempt to provide impaired listeners with the nonlinear phase cues that were predicted in Chapter 4 to be important for processing complex stimuli in the normal auditory system.

Hearing aid strategies that attempt to implement fast-acting, multi-band compression using a scheme more similar to the cochlear amplifier, i.e., based on broadband-stimulus-dependent tuning, have the potential to provide additional benefit to hearing-impaired listeners beyond current magnitude compression algorithms. Much work is required to quantify whether such processing strategies will prove useful given the physiological limitations of not being able to access each CF directly based on acoustic hearing aids.

Methods for quantitatively relating physiological response properties to psychophysical performance, such as those presented in this dissertation, will improve our understanding of the processing of the cochlear amplifier in the normal ear and its relation to psychophysical masking studies. This knowledge will be beneficial for identifying useful forms of signal information that could be enhanced by preprocessing schemes to improve hearing-impaired performance. There is great potential for digital-signal processing technology, which is now being incorporated into hearing aids, to allow the types of dynamic algorithms that are necessary to recreate some of the nonlinear cues provided by the cochlear amplifier.

7.9 Future Work

A general benefit of any model that was developed based on a set of data is the ability to make predictions for new experiments. This benefit includes the ability to test quantitatively current theories that have only been expressed qualitatively based on experimental results, and to test new ideas that have not yet been tested experimentally. The current modeling approach is particularly useful because of the ability to suggest useful psychophysical experiments based on physiological data, and vice versa. There are several directions in which the quantitative approach developed in this dissertation has the potential to improve the current understanding of normal and impaired hearing.

Physiological response properties that have been characterized in experimental animals can be related to humans by using the approach developed in this dissertation to provide a link between physiological and psychophysical studies of similar phenomena. As discussed in Chapter 6, the physiological study of the effects of AN suppression in cat on psychophysical masking by Delgutte (1990a) and the human psychophysical masking study by Oxenham and Plack (1998) could be compared with the current approach to quantitatively relate the strength of suppression in cats and in humans. Similarly, the strength of compression in humans could be evaluated by using an AN model that matched physiological measurements of BM compression in chinchilla (e.g., Ruggero *et al.*, 1997; Cooper and Rhode, 1997) and predicting performance based on AN information for psychophysical tasks that estimate compression in human listeners (e.g., Oxenham and Plack, 1997; Plack and Oxenham, 1998).

Much discussion of the significance of hearing impairment on speech perception has focused on the predicted smearing of the internal representation of vowel spectra resulting from broadened tuning. This idea has produced a wealth of studies (both psychophysical and physiological) on steady-state vowel perception, or on discrimination using vowel-like harmonic tone complexes, with normal and impaired subjects. Psychophysical studies using these relatively simple stimuli generally show a significant difference between hearing-impaired and normal listeners (e.g., Danaher *et al.*, 1973; Danaher and Pickett, 1975; van Tasell, 1980; Hannley and Dorman, 1983; Turner and van Tasell, 1984; Turner and Holte, 1987; Leek *et al.*, 1987; Leek and Summers, 1996). There have also been several physiological studies that have examined the internal representation of vowel spectra or harmonic tone complexes in AN fibers of normal and impaired ears (e.g., Sachs and Young, 1979,1980; Young and Sachs, 1979; Horst *et al.*, 1990; Hienz *et al.*, 1996; Miller *et al.*, 1997,1999; Wong *et al.*, 1998). While many of these studies have focused on the effect of broadened filters, there are several reasons that perception of changes in F2 or F3 would be expected to be impaired with the loss of the cochlear amplifier. Suppression has been hypothesized to enhance the representation of spectral peaks in complex sounds, such as steady-state vowels (Young and Sachs, 1979; Sachs and Young, 1980). Synchrony capture, which refers to the nonlinear dominance of spectral peaks near CF on AN responses to complex stimuli, appears to overcome the broadened tuning at high stimulus levels (Horst *et al.*, 1990; Miller *et al.*, 1997,1999). Miller *et al.* (1997,1999)

have shown that responses to steady-state vowels in AN fibers that have CF's near the F2 formant frequency are captured by the F2 stimulus component in normal cochleae, but that impaired cochleae have AN responses that are much more significantly influenced by F1 and other stimulus components. This finding correlates with psychophysical experiments that demonstrate a much larger masking effect of F2 by F1 in hearing-impaired listeners than in normal-hearing listeners (Danaher, 1973; Danaher and Pickett, 1975; Hannley and Dorman, 1983; Turner and Holte, 1987). The influence of suppression, synchrony capture, and the other cochlear-amplifier properties on the AN information available for discrimination of simple vowel-like stimuli in noise (e.g., Leek *et al.*, 1987) could be evaluated with the present approach. Comparisons between the model predictions and the wealth of physiological and psychophysical data provide powerful tests for whether performance deficits in various tasks should be expected due to peripheral impairment.

The present approach has the potential to suggest new physiological experiments or metrics that may be useful in studying the encoding of sound in the auditory periphery. For example, the significant information contributed by nonlinear phase cues at low frequencies (Chapter 4) is not apparent in current physiological metrics that are used to compare the encoding of complex sounds in normal and impaired AN responses (e.g., Miller *et al.*, 1997,1999). Average-rate and synchronized-rate (the discharge rate associated with each stimulus frequency component) metrics have typically been used to quantify AN information in these studies. The present study suggests that the phase of synchronized-rate responses would demonstrate a significant difference between the normal and impaired AN. The present study also suggests that average-rate representations based on a population of coincidence detectors would be enhanced compared with the AN-rate representation due to the processing of nonlinear phase cues. Cochlear-nucleus neurons with several response types have been shown to have responses consistent with a coincidence-detection mechanism (Carney, 1990; Joris *et al.*, 1994). Thus, the present study provides a potential basis for the observation of enhanced rate representations of speech in the cochlear nucleus at high levels (Blackburn and Sachs, 1990; May *et al.*, 1998). A relation between the expected value and the variance of discharge counts from a coincidence detector was described in Chapter 4, and depended on the shape of the temporal coincidence window. This relation suggests physiological metrics that

may be useful in rate-based studies of neurons that are believed to operate as coincidence detectors. It should be noted that temporal information is also enhanced in the cochlear nucleus (Joris *et al.*, 1994), perhaps based on coincidence detection, and may be processed further to enhance representations at higher levels of the auditory system. Future theoretical analyses of coincidence detection may suggest additional physiological metrics that would be useful in studies of temporal information at higher levels of the auditory system.

As discussed above, restricted (sub-optimal) temporal processors are necessary to account for human performance based on the information in AN discharge times. Coincidence detection may be a robust mechanism for detecting tones in noise because the discharge times on adjacent AN fibers will become more coincident as tone level is increased, independent of whether the average discharge rate of the AN fibers is saturated by the noise. Thus, future studies based on restricted temporal processors may provide better matches to the trends and absolute performance levels of human listeners in psychophysical masking tasks than the optimal performance quantified in the present study.

It is currently difficult to clinically identify the specific etiology of an impaired listener. Future applications of the present approach with AN model versions that are impaired in different ways (e.g., conductive loss, IHC loss, reduced OHC function) can suggest and test hypotheses for potential audiological tests that could isolate specific types of hearing loss. This knowledge is crucial to the proper applications of hearing aids for individual listeners. The present approach can also be used in order to evaluate psychophysical tasks that are used to simulate specific hearing losses. Frequency-shaped noise maskers are used to reduce the audibility of stimuli for a normal-hearing subject to match the audiogram of a hearing-impaired subject (e.g., Florentine and Buus, 1984; Florentine, 1992; Florentine *et al.*, 1993; Buus *et al.*, 1995). This approach is typically successful in predicting many aspects of hearing impairment; however, not all deficits demonstrated by impaired listeners are predicted by the noise-masked normal listeners. The modeling approach could be used to parse the physiological response properties that are responsible for the differences between noise-masked-normal and impaired performance.

The present approach provides a quantitative framework that could be used for the development and testing of hearing-aid algorithms. Predicted performance based on an

impaired version of the AN model with and without the hearing aid could be compared to demonstrate the potential usefulness of the algorithm for improving hearing-impaired performance. Bruce *et al.* (1999,2000) are pursuing this idea by comparing physiological metrics of AN information from normal and impaired versions of the computational model used in Chapters 5 and 6 (Zhang *et al.*, 2000) to data reported by Miller *et al.* (1997) from normal and impaired AN responses in cat. The general approach developed in this dissertation provides a quantitative method to relate normal and impaired physiological responses to psychophysical performance.

Appendix A

Nonlinear computational auditory-nerve model

The computational auditory-nerve (AN) model used for the predictions in Chapters 5 and 6 is similar to the phenomenological model developed by Zhang *et al.* (2000). Thus, only the implementation issues and response properties important for the present predictions are discussed below, while the details of the implementation are described by Zhang *et al.* (2000). The computational model describes many of the important AN-response properties associated with the cochlear amplifier, including broadened tuning with increases in level and the associated compressive-magnitude and nonlinear-phase responses, and two-tone suppression properties. The present AN model is an improvement of the original Carney (1993) model, with the major advance being the addition of two-tone suppression properties and different spontaneous-rate fibers. The predictions in Chapters 5 and 6 use the computational model to produce a population response; however, the description below is limited to a single characteristic frequency (CF).

The block diagram in Fig. A-1 shows the basic implementation of the AN model. Level-dependent tuning in the model is achieved with a third-order nonlinear narrowband filter followed by a broader first-order linear narrowband filter, both of which have the same CF. The bandwidth of the nonlinear filter is controlled via the time constant $\tau(t)$, which is varied according to the output of a control path, and the linear-filter has a bandwidth that is equal to the maximum bandwidth of the nonlinear filter (corresponding to τ_{min}).

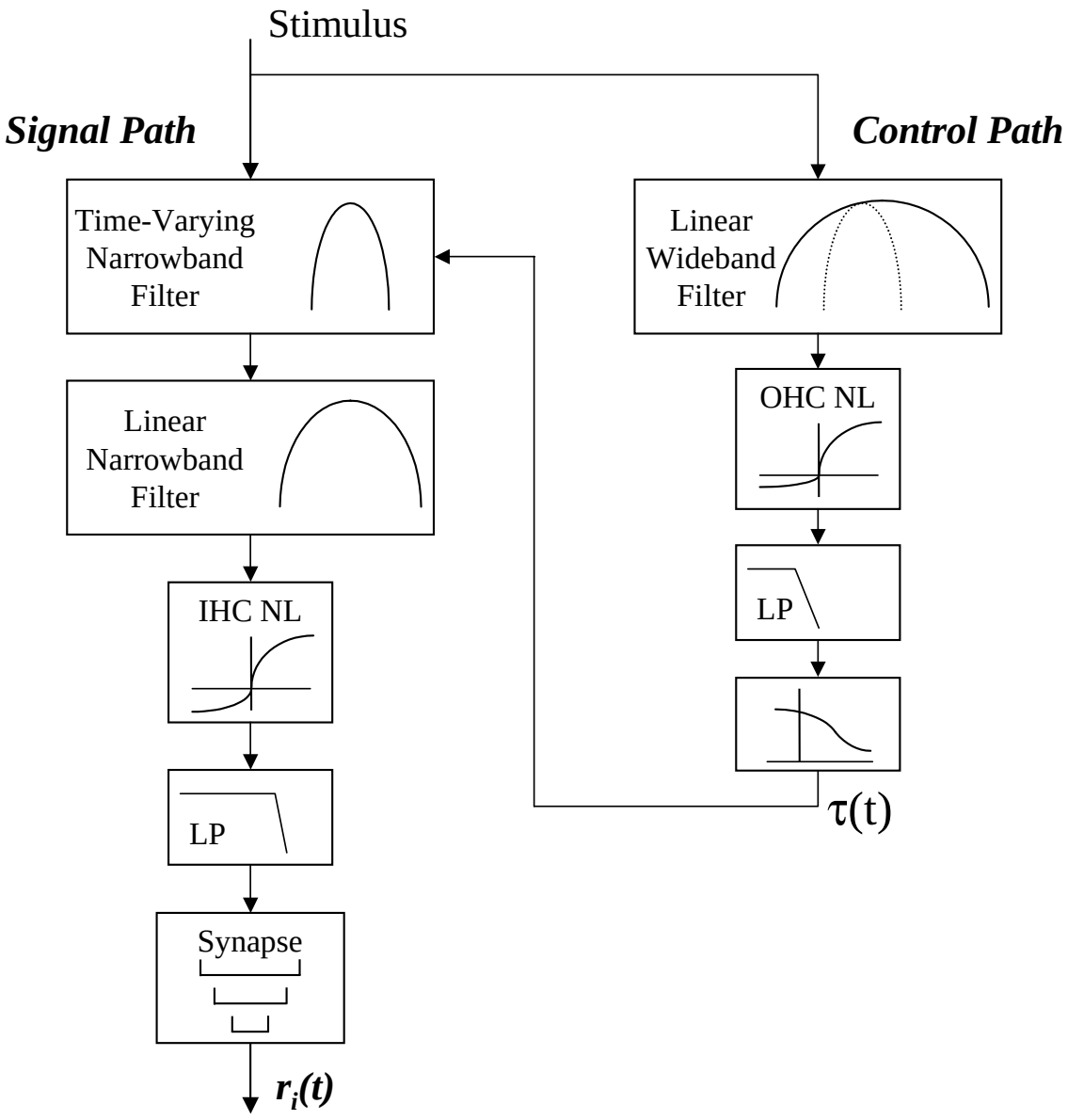

Figure A-1: Block diagram of computational nonlinear auditory-nerve model.

The control path has a third-order, linear wideband filter (discussed in more detail below), followed by an asymmetric (7:1), saturating nonlinearity. The nonlinearity, which represents the transduction properties of the outer-hair cells (OHC), is followed by a 5th-order lowpass filter with an 800-Hz cutoff frequency. The lowpass filter determines the dynamics of the cochlear amplifier, which appear to be extremely fast (time constant of roughly 200 μs) based on BM responses to clicks (Recio *et al.*, 1998). The output of the lowpass filter is mapped into $\tau(t)$ according to a nonlinearity that equals τ_{max} for zero input and equals τ_{min} for the maximum dc output of the OHC nonlinearity (Fig. A-1). The low-level time constant τ_{max} is specified according to psychophysical estimates of low-level human frequency selectivity (described below), and the high-level time constant τ_{min} is specified as a fraction of τ_{max} according to the cochlear-amplifier gain for the given CF (described below). Thus, the control path acts to decrease $\tau(t)$ (i.e., to broaden the nonlinear filter) as the level of the stimulus increases. This implementation produces the desired decrease in gain and level-dependent phase shifts near CF associated with the broadening of the nonlinear filter.

Broadened model tuning with increases in stimulus level is illustrated in Fig. A-2 for a 1-kHz CF. Normalized magnitude response at the output of the linear, narrowband filter is shown as a function of tone frequency for a range of stimulus levels. These responses are consistent with basilar-membrane magnitude responses to tones (Ruggero *et al.*, 1997). A compressive magnitude response occurs near CF, with the maximum compression occurring at CF and an essentially linear response well away from CF. Filter output at CF is shown as a function of tone level for a range of CFs in the top panel of Fig. A-3. The compressive response begins to occur at 20 dB SPL and is fully compressive by 40 dB SPL, consistent with physiological data (Ruggero *et al.*, 1997). The compression in the model response begins to decrease above 80 dB SPL; however, filter responses are still slightly compressive up to 120 dB SPL. The amount of compression in filter responses is largest for high CFs. The bottom panel of Fig. A-3 shows the cochlear amplifier gain in the model as a function of CF. The cochlear-amplifier gain was calculated as the reduction in filter response at 120 dB SPL relative to a linear response, consistent with the idea that full amplification occurs at low levels, and the gain of the cochlear amplifier is reduced as stimulus level increases. There is slightly more than 55 dB of gain at 8 kHz, and less than 20 dB of gain at 500 Hz, consistent

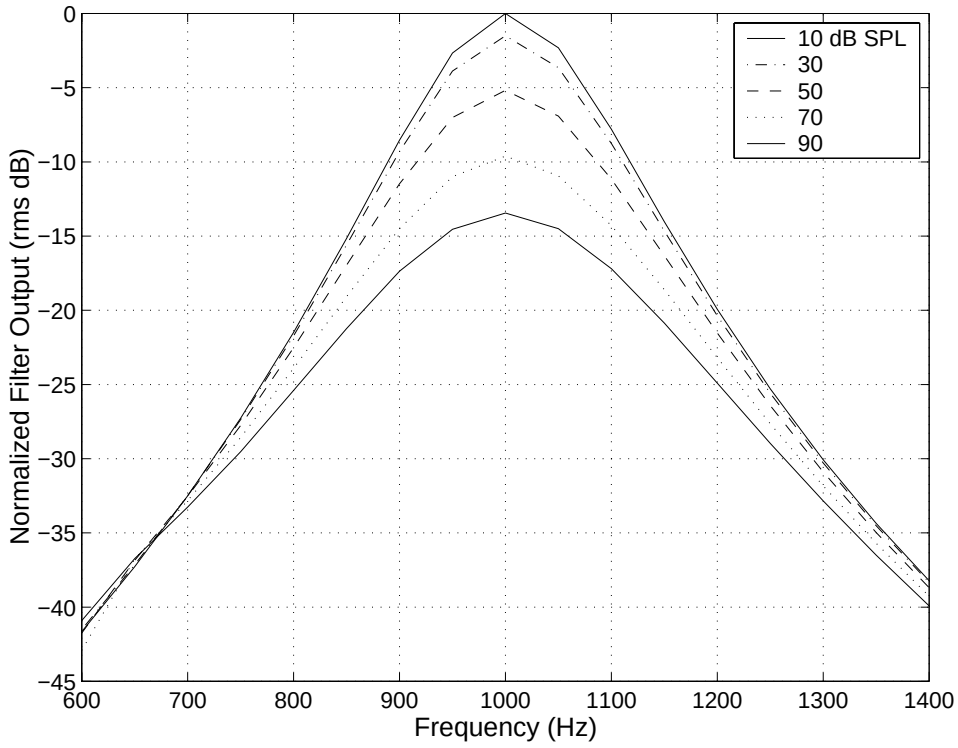

Figure A-2: Nonlinear-filter responses as a function of level. Normalized filter output is shown for a 1-kHz CF as a function of frequency for a range of stimulus levels. RMS Filter output was calculated over 10 cycles beginning at 40 ms of the tone burst, and normalized to the 10-dB SPL response.

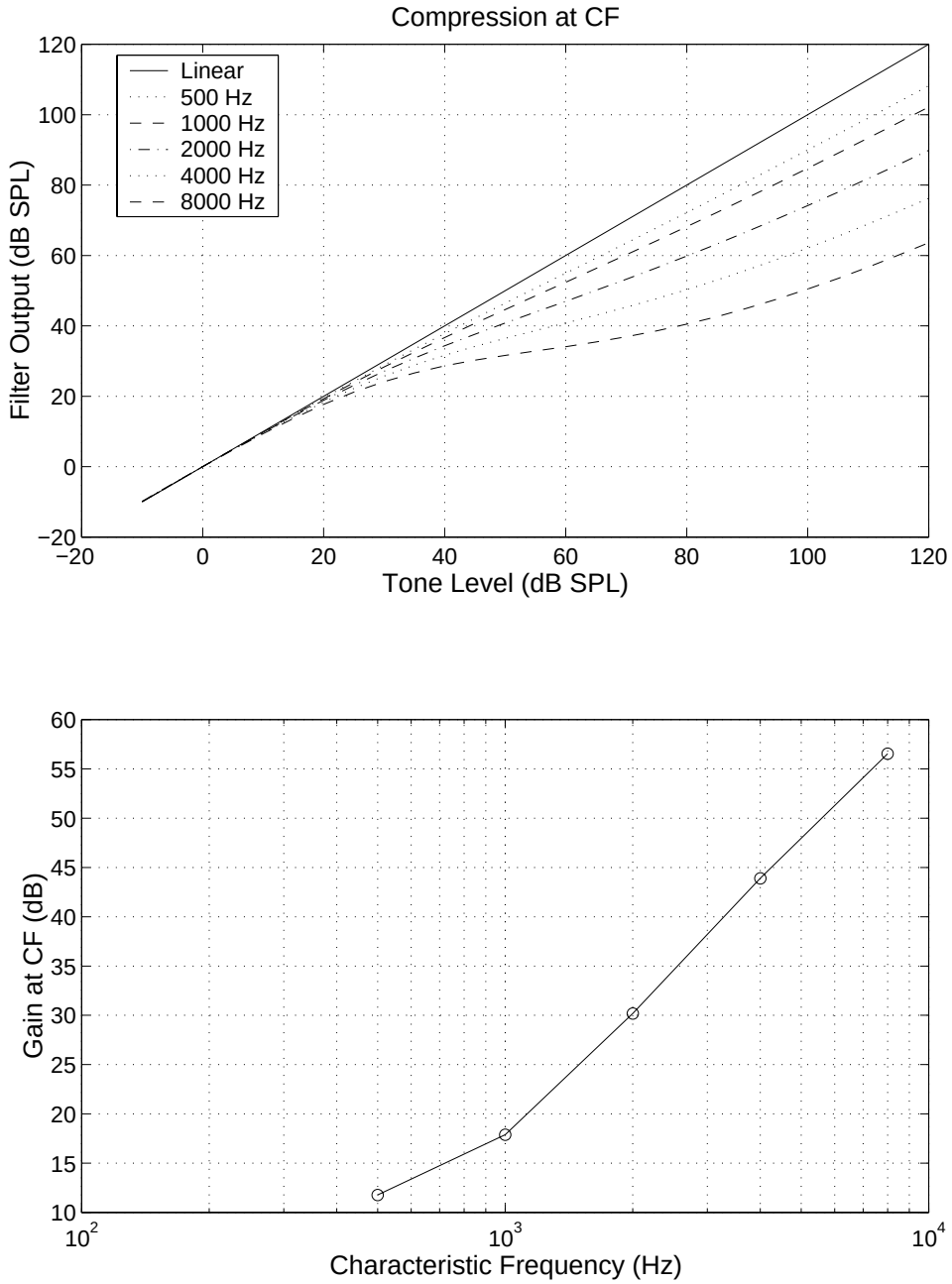

Figure A-3: Compressive magnitude responses of the nonlinear filter in the AN model for CF's ranging from 500 to 8000 Hz. The solid line represents a linear response, and the amount of compression can be seen to increase with CF. The compressive gain, defined as the difference between the linear and nonlinear responses in dB at 120 dB SPL, is shown in the bottom panel as a function of CF. Output of the filter represents the average ac response over 10 cycles of the stimulus beginning 40 ms after the onset of the tone stimulus.

with BM data from the chinchilla basal (Ruggero *et al.*, 1997) and apical turns (Cooper and Rhode, 1997). The smooth transition between low and high frequencies is consistent with several psychophysical measures of cochlear nonlinearity (Hicks and Bacon, 1999a), but has not been demonstrated physiologically.

Nonlinear-phase responses associated with broadened cochlear tuning as stimulus level increases are observed in AN responses (Anderson *et al.*, 1971). The time-varying AN discharge rate $r(t)$ in the model is calculated by passing the output of the linear, narrowband filter through an asymmetric (3:1) saturating nonlinearity, a lowpass filter, and a synapse model (discussed in more detail below). Nonlinear-phase responses of AN fibers are shown in Fig. A-4. The top row shows average discharge rate as a function of tone frequency for a range of stimulus levels. Phase normalized to the phase at 90 dB SPL is shown in the bottom panels as a function frequency for the same range of levels. The level-dependent phase changes below and above CF at both low (left column) and high (right column) frequencies are consistent with nonlinear phase responses observed in BM (Ruggero *et al.*, 1997), IHC (Cheatham and Dallos, 1998), and AN (Anderson *et al.*, 1971) responses, as discussed in Chapter 4).

The tuning of the model filters was specified according to psychophysical estimates of human frequency selectivity determined from the notched-noise method of measuring auditory filter shapes (see Chapter 6). The low-level bandwidths of the nonlinear filters were specified by

$$\tau_{max} = 1.2\{2\pi[1.019][24.7(4.37CF + 1)]\}^{-1}, \tag{A.1}$$

where the expression $24.7(4.37CF+1)$ describes mid-level estimates of equivalent-rectangular bandwidth (ERB) provided by Glasberg and Moore (1990), the factor of 1.019 represents a conversion between gamma-tone-filter bandwidth and ERB (Patterson *et al.*, 1987), and the increase in τ by a factor of 1.2 is consistent with the ratio of low- to mid-level ERBs (Glasberg and Moore, 1990). Threshold tuning-curves for the several versions of the AN model are shown in Fig. A-5. The nonlinear AN model represents a *healthy* cochlea; the *linear, broad, low threshold* version of the model corresponds to the AN model with $\tau(t) = \tau_{min}$ and

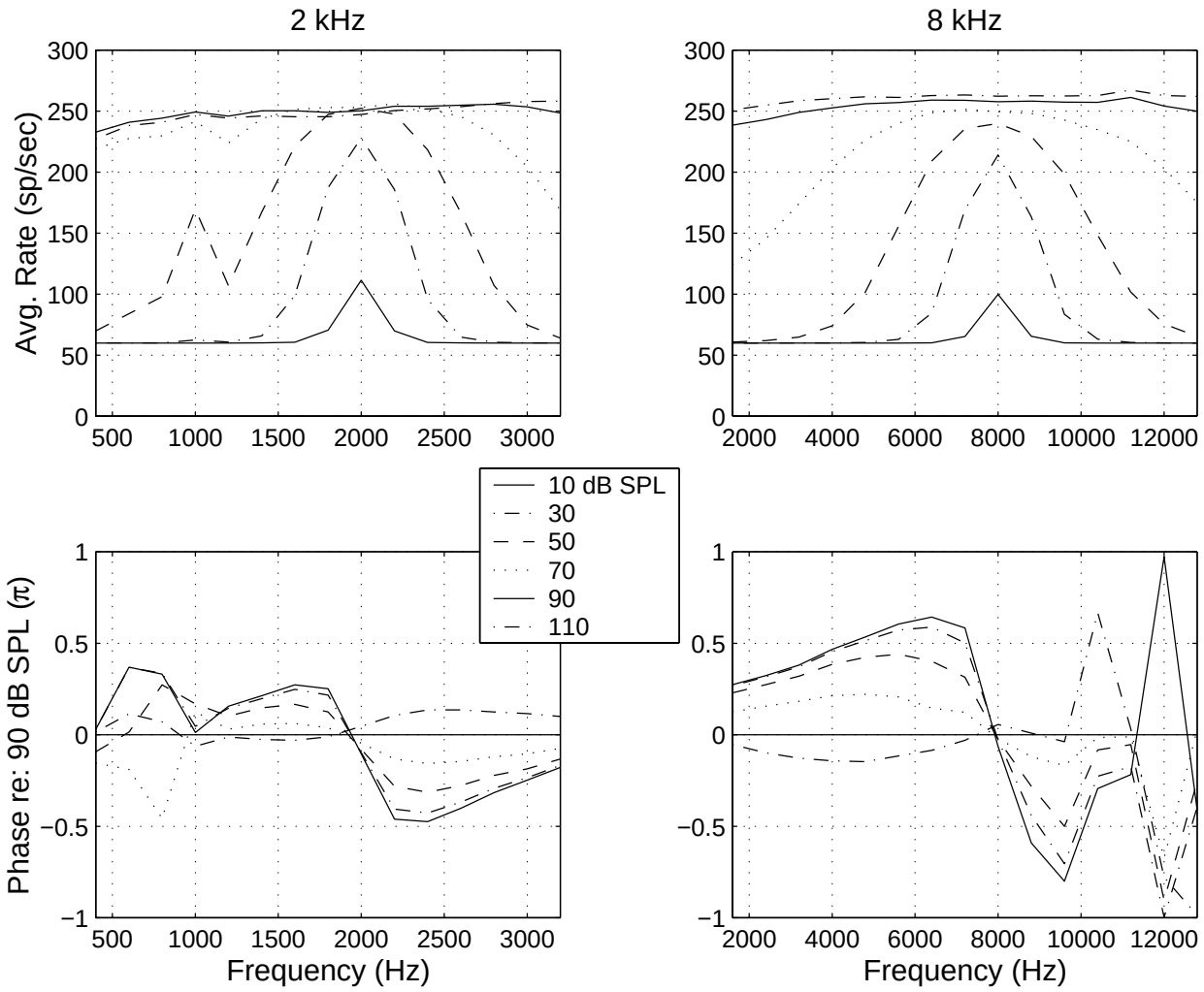

Figure A-4: Rate- and phase-response areas of 2- and 8-kHz CF fibers in the nonlinear AN model for a range of levels. The top row shows average rate as a function of frequency, where the rate was calculated over 10 cycles beginning at 20 ms of a 50-ms (5-ms rise/fall) tone burst. The bottom row shows phase as a function of frequency, where the phase has been normalized to the phase at 90 dB SPL. Thus any difference from zero represents a phase response that changes with level. Phase was calculated over one cycle of the tone burst beginning at 40 ms.

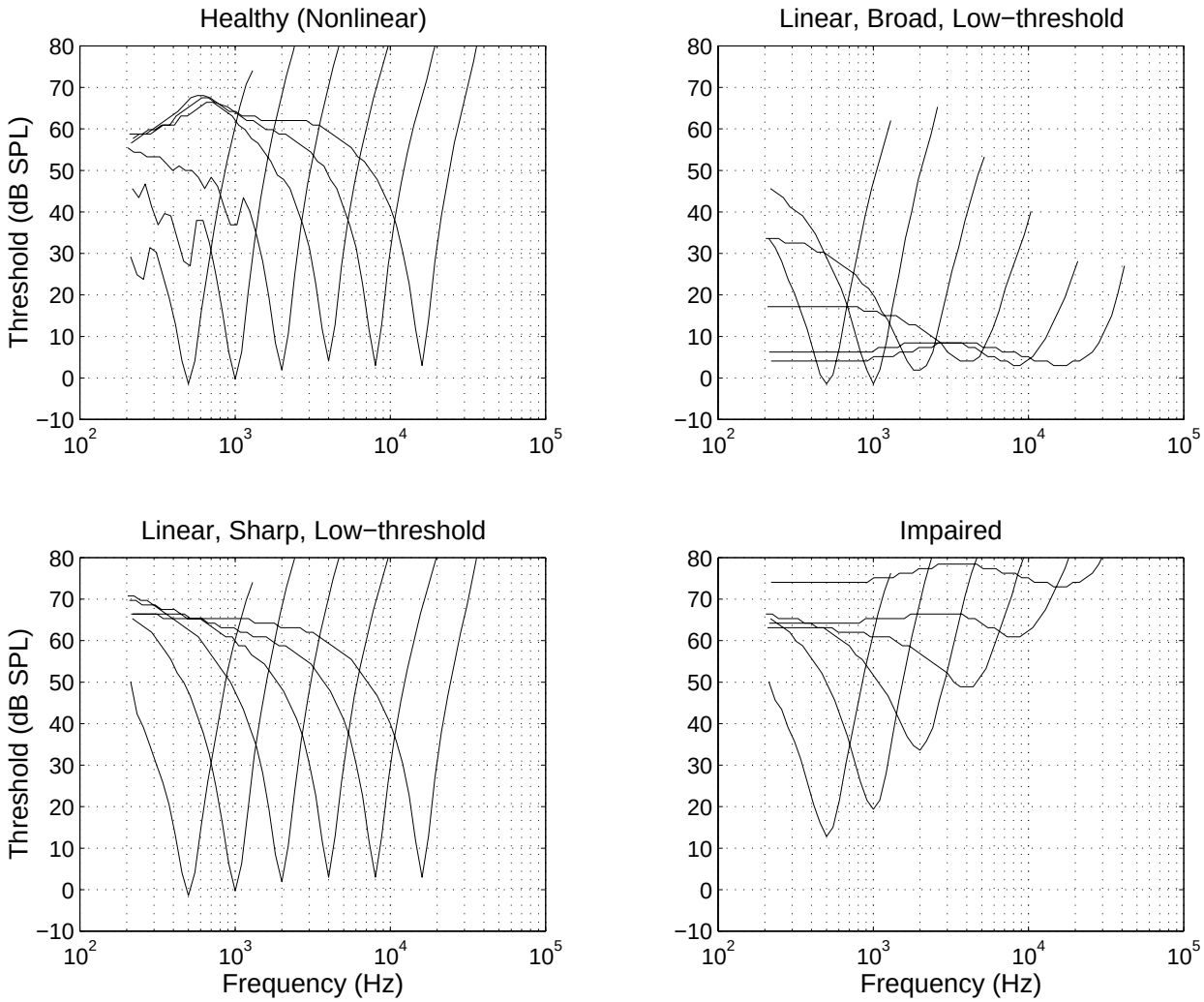

Figure A-5: Tuning Curves for the four versions of the AN model (upper-left: nonlinear AN model; upper-right: linear, broad tuning, low thresholds; lower-left: linear, sharp tuning, low thresholds; bottom-right: impaired). Stimuli were 50-ms (5-ms rise/fall) tone bursts, the average-rate was taken between 1.25-51.25 ms, and the criterion for threshold was 10 sp/sec above spontaneous rate.

full cochlear-amplifier gain, independent of stimulus level; the *linear, sharp, low threshold* model corresponds to $\tau(t) = \tau_{max}$ and full cochlear-amplifier gain; and the *impaired* model corresponds to $\tau(t) = \tau_{min}$ and no cochlear-amplifier gain. None of the effects of the external and middle ear are considered, and thus all versions of the model except the impaired version have thresholds for HSR fibers that are within a few dB of 0 dB SPL. The impaired AN model represents a cochlea that has completely damaged OHC's, but fully functioning IHC's. Note that loss of the gain provided by the cochlear amplifier results in the common sloping high-frequency hearing loss, with roughly 10 dB of loss at low frequencies and almost 60 dB of loss at high frequencies.

Two-tone suppression is another property of cochlear tuning that is associated with the cochlear amplifier (Sachs and Kiang, 1968; Arthur *et al.*, 1971; Delgutte, 1990b; Ruggero *et al.*, 1992). Tones of certain frequencies and levels that do not excite an AN fiber can reduce the discharge rate in response to a CF tone. It is believed that the non-excitatory stimulus acts to turn down the gain of the cochlear amplifier without exciting the AN fiber. In order to incorporate this property into the nonlinear AN model, the control-path filter (third-order) was made much wider than the excitatory filter (i.e., $\tau_{wb} = 0.2\tau_{max}$, as indicated by the dotted-lines shown with the control-path filter in Fig. A-1) (e.g., Geisler and Sinex, 1980). The active region of the cochlear amplifier is thought to be shifted basally on the BM from the excitatory CF (Yates, 1995). The CF of the control-path filter was shifted higher than the excitatory CF in the model by 1.2 mm on the BM, according to the human cochlear map (Greenwood, 1990; described in Chapter 2). Two-tone suppression tuning curves are often used to illustrate the frequency and level combinations for a suppressor tone that reduce the discharge rate in response to a CF-tone by a certain criterion (e.g., Sachs and Kiang, 1968; Arthur *et al.*, 1971; Delgutte, 1990b). The basal shift of the active region is consistent with the observation that the tip of two-tone suppression tuning curves often lies higher in frequency than the excitatory CF (Sachs and Kiang, 1968; Arthur *et al.*, 1971; Delgutte, 1990b). Figure A-6 shows two-tone suppression tuning curves for model CF's of 1- and 8-kHz, along with the excitatory tuning curves. Suppression is observed both below and above CF for both CF's, consistent with physiological data (Delgutte, 1990b). The suppression thresholds below the 1-kHz CF are above the excitatory thresholds, consistent

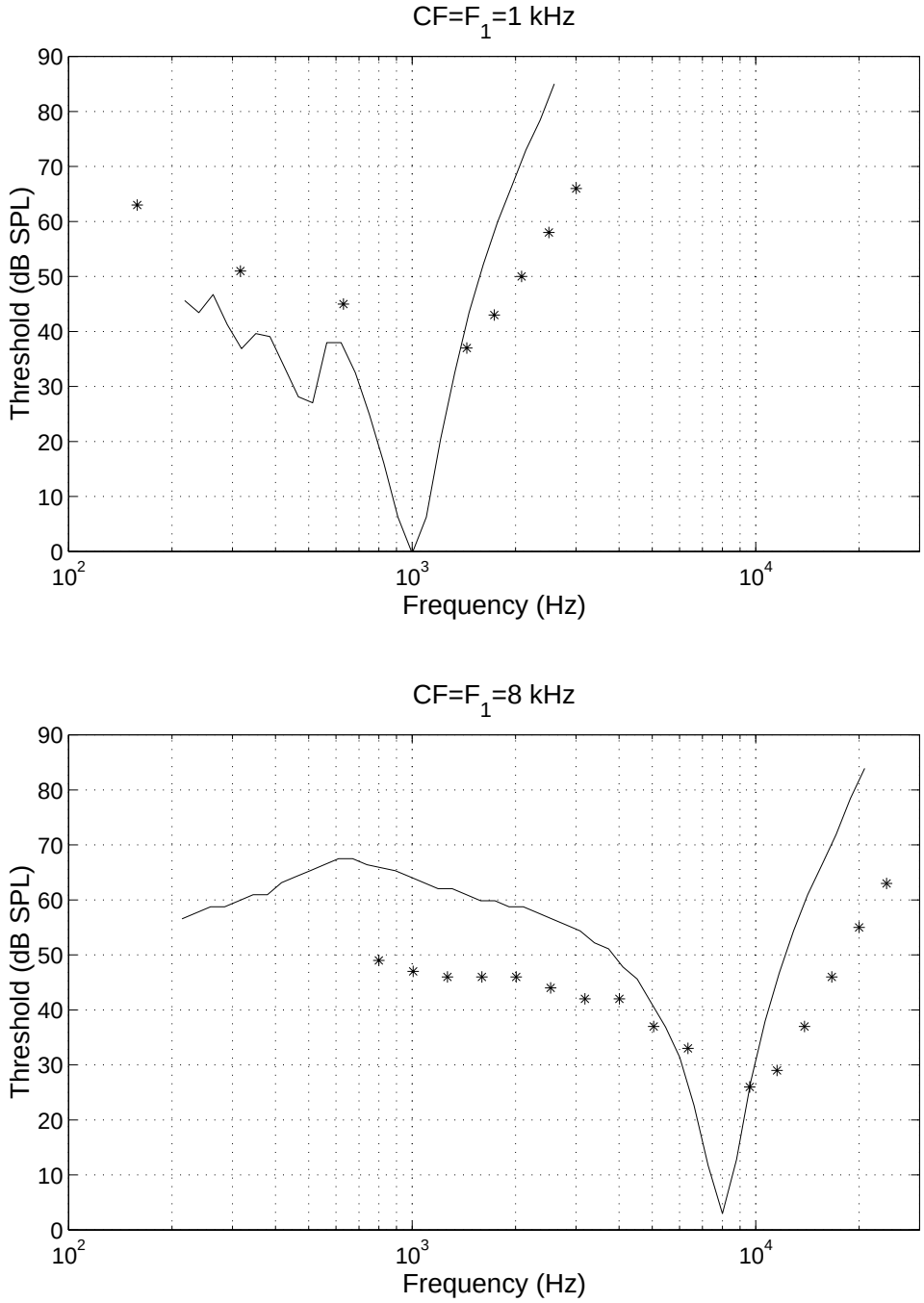

Figure A-6: Two-tone-suppression tuning curves for 1- and 8-kHz fibers in the nonlinear AN model. The solid curve represents the excitatory tuning curve for a 50-ms (5-ms rise/fall) tone burst, where average rate was calculated over the interval 1.25-51.25 ms, and the criterion for threshold was 10 sp/sec above spontaneous rate. The stars represent suppression thresholds (10 sp/sec criterion) when a second tone was presented in addition to a 24-dB SPL CF-tone [this level corresponds to roughly 67% of the driven rate, as in Delgutte (1990b)].

with the difficulty reported by Delgutte (1990b) in using a tuning-curve algorithm to measure two-tone suppression at frequencies below CF for low CF's. The growth of suppression is illustrated in Fig. A-7, similarly to Delgutte (1990b). The level of the CF-tone required to produce 67% of the driven discharge rate is plotted as a function of the suppressor-tone level for several suppressor frequencies above and below CF. The present version of the AN model shows similar growth rates above and below CF, inconsistent with physiological data for which the below-CF suppression grows up to 6 times steeper than above CF (Delgutte, 1990b). This discrepancy is not viewed to be critical for the present study because the conclusions in Chapters 5 and 6 only depend on the presence or absence of suppression (represented by setting $\tau_{wb} = 0.2\tau_{max}$ or $\tau_{wb} = \tau_{max}$, respectively). A more recent version of the AN model has demonstrated asymmetric growth rates, consistent with the physiological data, and will be useful for future studies (Zhang *et al.*, 2000).

Rate-level curves for CF-tones are shown in Figs. A-8 and A-9 for 1- and 8-kHz CF's, respectively, for the four versions of the AN model. Each panel shows rate-level curves for high- (60 sp/sec), medium- (5 sp/sec), and low-spontaneous-rate (1 sp/sec) AN fibers. A soft-rectifying function is used to transform the output of the IHC lowpass filter into the immediate permeability in the synapse model. The slope of this transformation was adjusted as a function of spontaneous rate (SR) so that the rate-level curves had shapes and relative thresholds that correspond well with those reported from physiological studies (e.g., Sachs and Abbas, 1974; Winter and Palmer, 1991). The rate-level curves for HSR fibers have thresholds near 0 dB SPL, and a dynamic range of about 30 dB SPL (Liberman, 1978; May and Sachs, 1992). The MSR fibers have slightly higher thresholds and larger dynamic ranges than the HSR fibers. The LSR fibers have the highest thresholds, and demonstrate much larger dynamic ranges. The LSR rate-level curves for the nonlinear version of the AN model have "sloping-saturation" shapes at 1 kHz, and "straight" shapes at 8 kHz, consistent with the rate-level curves reported by Winter and Palmer (1991) for guinea pig. The linear versions of the AN model have reduced dynamic ranges, especially at high CF's, consistent with physiological data from impaired animals (Evans, 1975a,b; Harrison and Evans, 1979; Harrison, 1981; Liberman and Kiang, 1984).

The synapse stage of the model is similar to that described in Chapter 2, except for

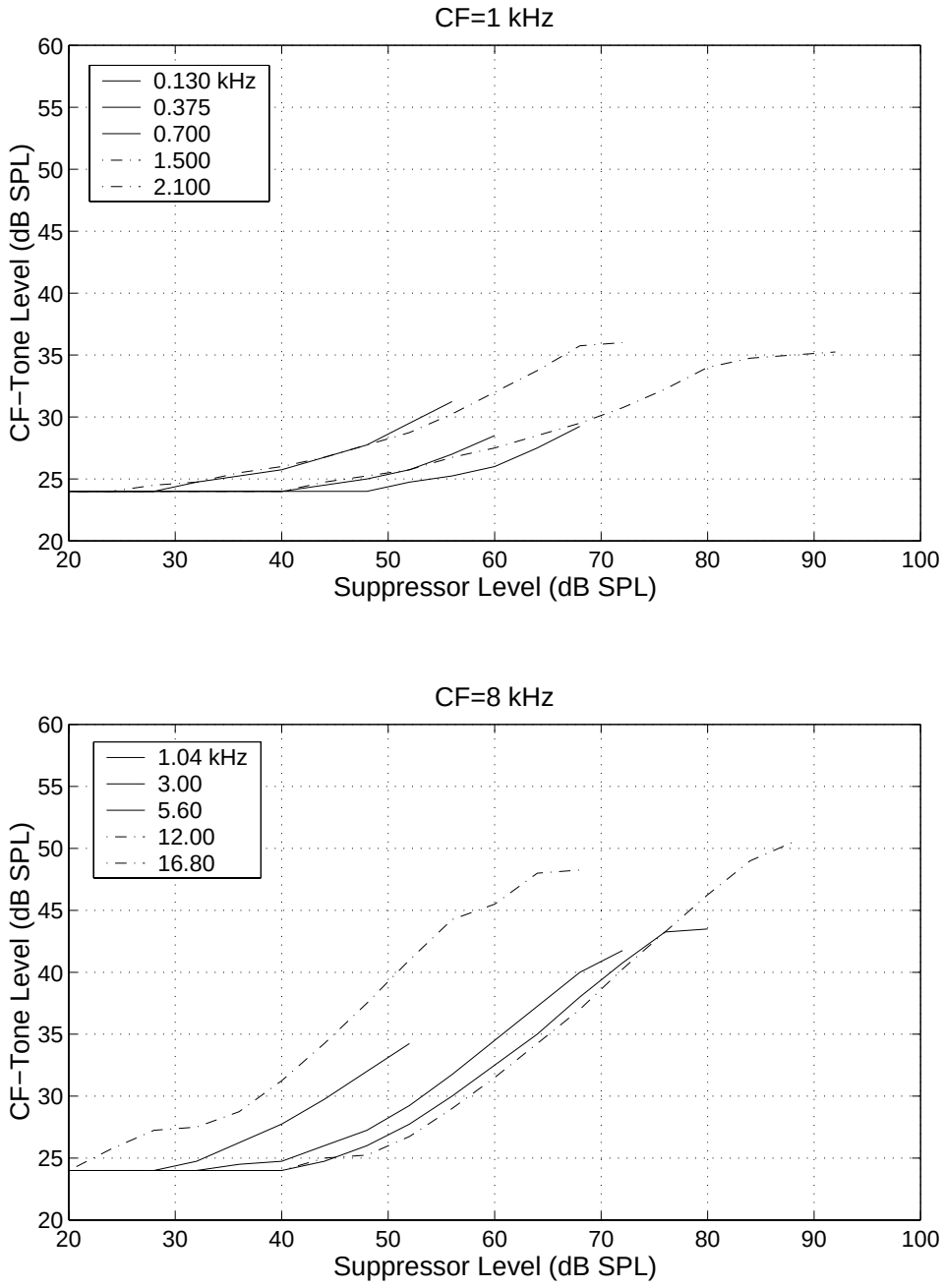

Figure A-7: The growth of two-tone-suppression for 1- and 8-kHz fibers in the nonlinear AN model. The level of a CF-tone needed to produce a criterion discharge rate in the presence of a suppressor tone is plotted as a function of suppressor-tone level for various suppressor-tone frequencies. The criterion discharge rate was 67% of the driven rate for the CF-tone in quiet (Delgutte, 1990b). The stimuli were 50-ms (5-ms rise/fall) tone bursts, where average rate was calculated over the interval 1.25-51.25 ms

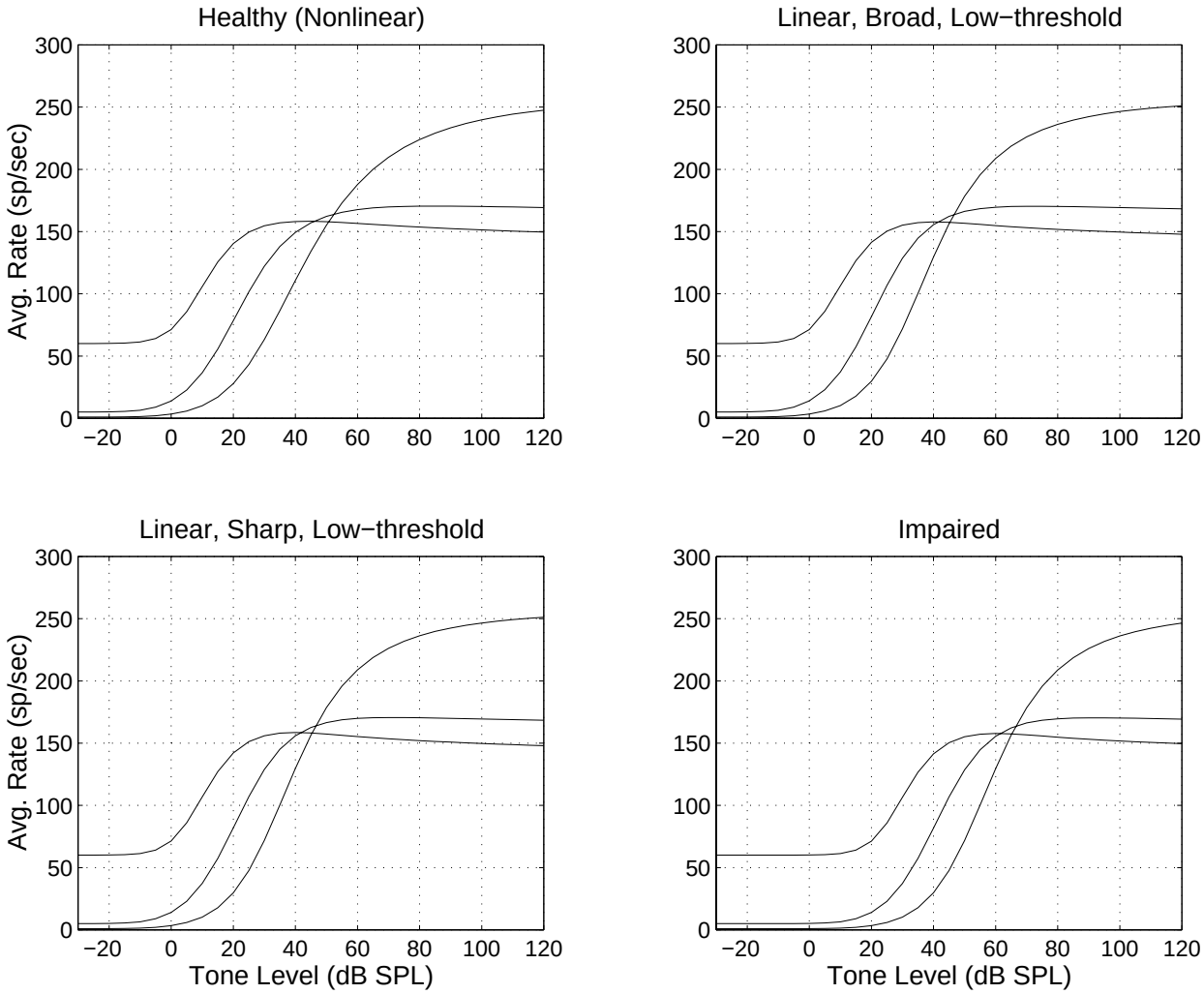

Figure A-8: Rate-level curves of 1-kHz fibers from the three spontaneous-rate (SR) groups (HSR: high; MSR: medium; LSR: low) in the four versions of the AN model (upper-left: nonlinear AN model; upper-right: linear, broad tuning, low thresholds; lower-left: linear, sharp tuning, low thresholds; bottom-right: impaired). Stimuli were 240-ms (10-ms rise/fall) tone bursts, and the average-rate was taken between 40-240 ms.

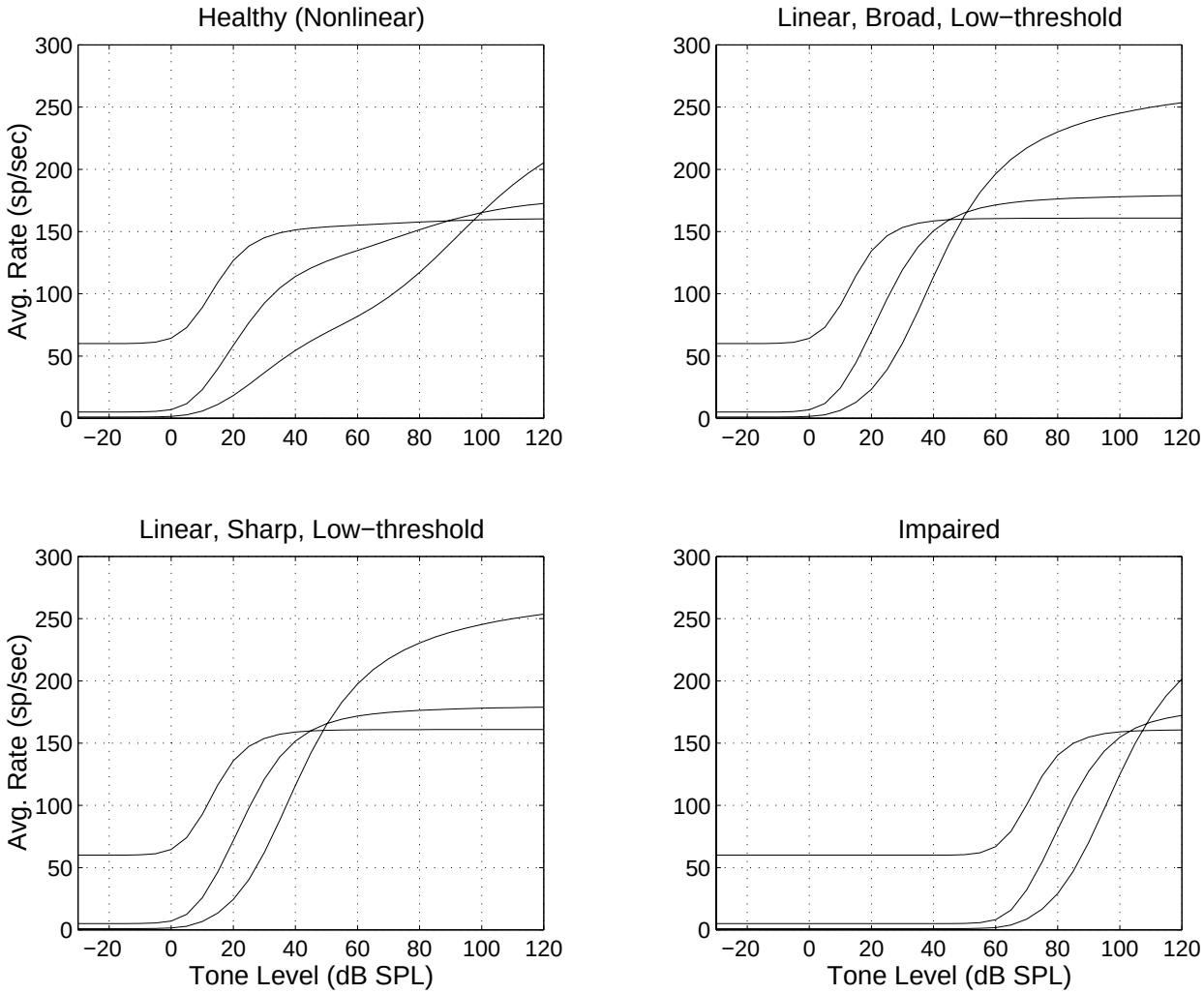

Figure A-9: Rate-level curves of 8-kHz fibers from the three spontaneous-rate (SR) groups (HSR: 60 sp/sec; MSR: 5 sp/sec; LSR: 1 sp/sec) in the four versions of the AN model (upper-left: nonlinear AN model; upper-right: linear, broad tuning, low thresholds; lower-left: linear, sharp tuning, low thresholds; bottom-right: impaired). Stimuli were 250-ms (10-ms rise/fall) tone bursts, and the average-rate was taken between 40-240 ms.

the adjustment for the three SR groups. Time-varying discharge rate waveforms for at-CF tones at three levels are shown in Fig. A-10 for a 1-kHz CF. The variation in the shape of the rate waveform as stimulus level increases is due to the adaptation properties associated with the neuro-transmitter release in the synapse. The onset response continues to increase with level, while the sustained rate is saturated at lower levels, as shown in Fig. A-11. The phase-locked response of AN fibers to low-frequency tones can be seen in Fig. A-10. The synchrony coefficient (or vector strength, which varies from 0 to 1) is a measure of the strength of phase-locking, and is shown in Fig. A-11 to increase with level over a narrow dynamic range. Synchrony threshold is roughly 20 dB lower than the rate threshold, while the peak synchrony coefficient occurs just above rate threshold. As level increases further, the synchrony coefficient decreases slightly. The ability of AN fibers to phase-lock decreases significantly above 2-3 kHz (Johnson, 1980), as shown in Fig. A-12. In the model, the rolloff in phase-locking occurs due to the IHC lowpass filter, which is 7th order, and has a cutoff frequency of 4500 Hz.

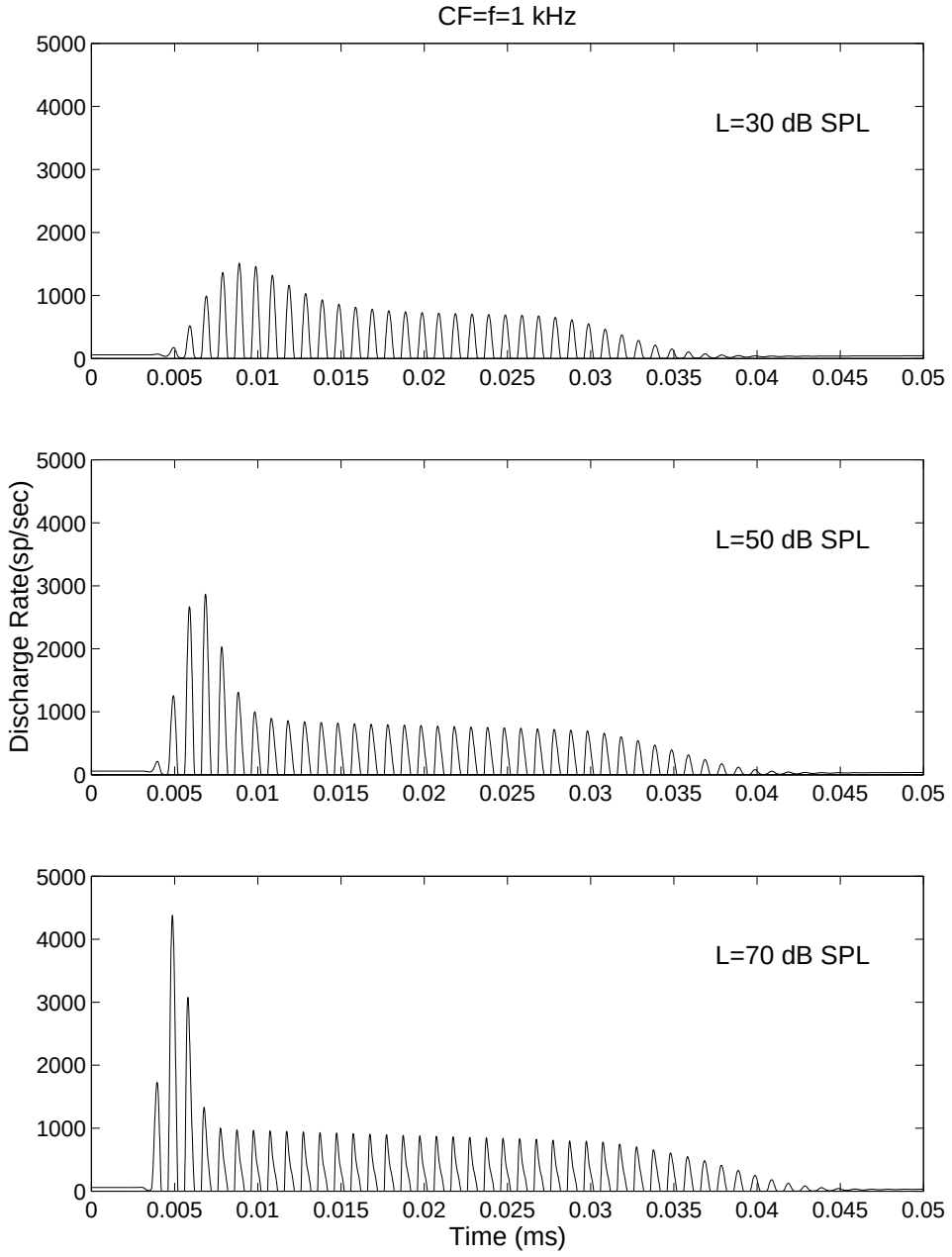

Figure A-10: Instantaneous discharge rate for a 1-kHz fiber from the nonlinear AN model in response to 25-ms (2.5-ms rise/fall) at-CF tone bursts for several levels.

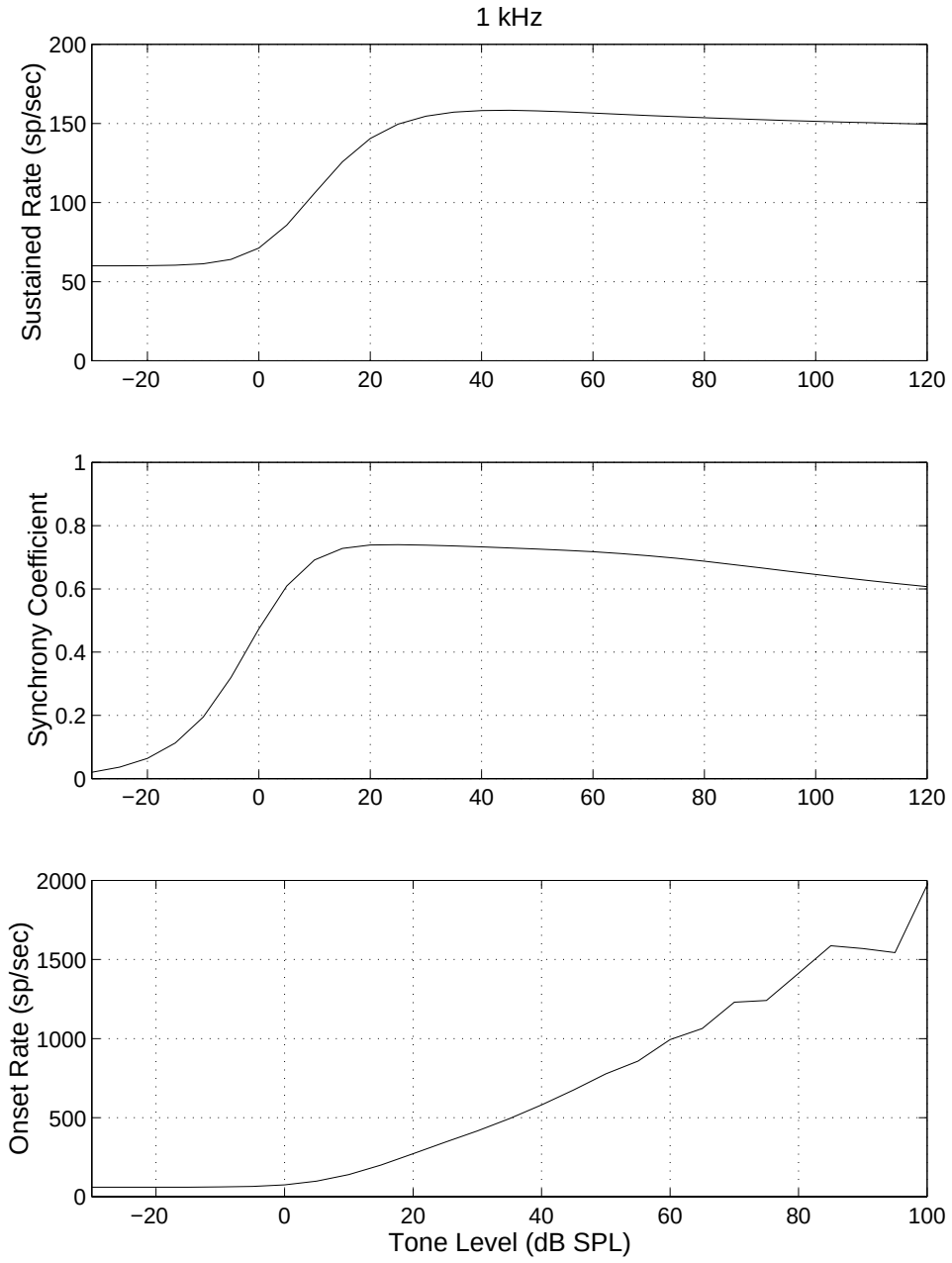

Figure A-11: Average rate, synchrony coefficient, and onset-rate as a function of level for a 1-kHz HSR (60 sp/sec) fiber for the nonlinear AN model. Stimuli were 250-ms (10-ms rise/fall) tones at CF. Average rate was calculated over the interval from 40 to 240 ms. Synchrony coefficient was calculated from one cycle beginning at 40 ms. Onset rate was calculated as the maximum average rate over one cycle within the first 20 ms of the stimulus.

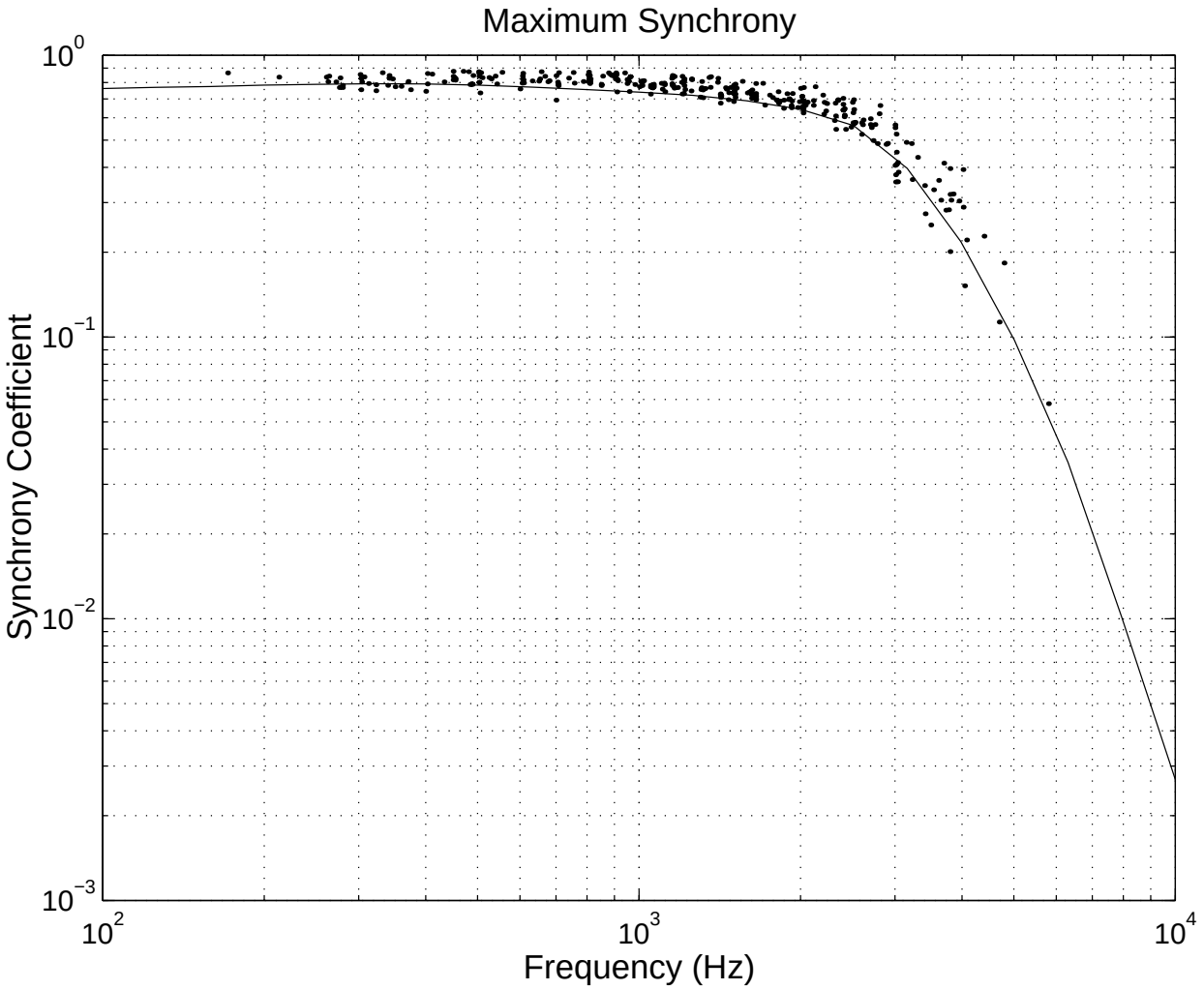

Figure A-12: The rolloff of phase-locking in the AN model as a function of frequency. The solid line shown the maximum synchrony coefficient from the AN model, and the dots represent physiological data from cat (Johnson, 1980). Synchrony coefficient was calculated over 1 cycle beginning at 40 ms of a 55-ms (5-ms rise/fall) tone burst at CF.

References

Allen, J.B. (**1997**). "Outer hair cells shift the excitation pattern via basilar membrane tension," in *Diversity in auditory mechanics*, edited by E.R. Lewis, G.R. Long, R.F. Lyon, P.M. Narins, C.R. Steel, and E. Hecht-Poinar (World Scientific Press, Singapore), pp. 167-175.

Allen, J.B., and Fahey, P.F. (**1992**). "Using acoustic distortion products to measure the cochlear amplifier gain on the basilar membrane," J. Acoust. Soc. Am. **92**, 178-188.

Allen, J.B., and Lonsbury-Martin, B.L. (**1993**). "Otoacoustic emissions," J. Acoust. Soc. Am. **93**, 568-569.

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

Arthur, R.M., Pfeiffer, R.R., and Suga, N. (**1971**). "Properties of 'two-tone inhibition' in primary auditory neurons," J. Physiol. **212**, 593-609.

Bacon, S.P., Boden L.N., Lee, J, and Repovsch, J.L. (**1999**). "Growth of simultaneous masking for $f_m < f_s$: Effects of overall frequency and level," J. Acoust. Soc. Am. **106**, 341-350.

Blackburn, C.C., and Sachs, M.B. (**1990**). "The representations of the steady-state vowel sound $/\mathcal{E}/$ in the discharge patterns of cat anteroventral cochlear nucleus neurons," J. Neurophysiol. **63**, 1191-1212.

Bos, C.E., and de Boer, E. (**1966**). "Masking and discrimination," J. Acoust. Soc. Am. **39**, 708-715.

Braida, L.D., and Durlach, N.I. (**1988**). "Peripheral and central factors in intensity perception," in *Auditory Function: Neurobiological Bases of Hearing*, edited by G.M. Edelman, W.E. Gall, and W.M. Cowan (Wiley, New York), pp. 559-583.

Bruce, I.C., Young, E.D., and Sachs, M.B. (**1999**). "Modifications of an auditory-periphery model to describe the effects of acoustic trauma on auditory nerve response," Abstracts of the 22rd Midwinter Meeting of the Association for Research in Otolaryngology, p. 209.

Bruce, I.C., Young, E.D., and Sachs, M.B. (**2000**). "Improvements of an auditory-periphery model to describe auditory nerve responses to speech stimuli," Abstracts of the 23rd Midwinter Meeting of the Association for Research in Otolaryngology, p. 277.

Buus, S. (**1990**). "Level discrimination of frozen and random noise," J. Acoust. Soc. Am. **87**, 2643-2654.

Buus, S., and Florentine, M. (**1991**). "Psychometric functions for level discrimination," J. Acoust. Soc. Am. **90**, 1371-1380.

Buus, S., and Florentine, M. (**1992**). "Possible relation of auditory-nerve adaptation to slow improvement in level discrimination with increasing duration," in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, New York), pp. 279-288.

Buus, S., Florentine, M., and Zwicker, T. (**1995**). "Psychometric functions for level discrimination in cochlearly impaired and normal listeners with equivalent-threshold masking," J. Acoust. Soc. Am. **98**, 853-861.

Cariani, P.A., and Delgutte, B. (**1996a**). "Neural correlates of the pitch of complex tones: I. Pitch and pitch salience," J. Neurophysiol. **76**, 1698-1716.

Cariani, P.A., and Delgutte, B. (**1996b**). "Neural correlates of the pitch of complex tones: II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch," J. Neurophysiol. **76**, 1717-1734.

Carlyon, R.P., and Moore, B.C.J. (**1984**). "Intensity discrimination: A severe departure from Weber's law," J. Acoust. Soc. Am. **76**, 1369-1376.

Carney, L.H. (**1990**). "Sensitivities of cells in the anteroventral cochlear nucleus of cat to spatiotemporal discharge patterns across primary afferents," J. Neurophysiol. **64**, 437-456.

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. (**1994**). "Spatiotemporal encoding of sound level: Models for normal encoding and recruitment of loudness," Hear. Res. **76**, 31-44.

Carney, L.H., Heinz, M.G., and Colburn, H.S. (**1999**). "Spatiotemporal coding of sound level: Quantifying the information provided by level-dependent phase cues," Abstracts of the 22nd Midwinter Meeting of the Association for Research in Otolaryngology, pp. 212-213.

Cheatham, M.A., and Dallos, P. (**1998**). "The level dependence of response phase: Observations from cochlear hair cells," J. Acoust. Soc. Am. **104**, 356-369.

Colburn, H.S. (**1969**). "Some physiological limitations on binaural performance," Ph.D. Dissertation, 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. (**1977a**). "Theory of binaural interaction based on auditory-nerve data. II.

Detection of tones in noise," J. Acoust. Soc. Am. **61**, 525-533.

Colburn, H.S. (**1977b**). "Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise. Supplementary material," AIP Document No. PAPS JASMA-**61**-525-98.

Colburn, H.S. (**1981**). "Intensity perception: relation of intensity discrimination to auditory-nerve firing patterns," Internal Memorandum, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA

Colburn, H.S. (**1996**). "Computational models of binaural processing," in *Auditory Computation*, edited by H.L. Hawkins, T.A. McMullen, A.N. Popper, and R.R. Fay (Springer-Verlag, New York), pp. 332-400.

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

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

Danaher, E.M., Osberger, M.J., and Pickett, J.M. (**1973**). "Discrimination of formant frequency transitions in synthetic vowels," J. Speech Hear. Res. **16**, 439-451.

Danaher, E.M., and Pickett, J.M. (**1975**). "Some masking effects produced by low-frequency vowel formants in persons with sensorineural hearing loss," J Speech Hear Res.18, 261-271.

Dau, T., Püschel, and Kohlrausch, A. (**1996a**). "A quantitative model of the 'effective' signal processing in the auditory system I. Model structure," J. Acoust. Soc. Am. **99**, 3615-3622.

Dau, T., Püschel, and Kohlrausch, A. (**1996b**). "A quantitative model of the 'effective' signal processing in the auditory system II. Simulations and measurements," J. Acoust. Soc. Am. **99**, 3623-3631.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997a**). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers," J. Acoust. Soc. Am. **102**, 2892-2905.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997b**). "Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration," J. Acoust. Soc. Am. **102**, 2906-2919.

de Boer, E. (**1966**). "Intensity discrimination of fluctuating signals," J. Acoust. Soc. Am. **40**, 552-560.

Delgutte, B. (**1987**). "Peripheral auditory processing of speech information: implications from a physiological study of intensity discrimination," in *The Psychophysics of Speech Perception*, edited by M.E.H. Schouten (Nijhoff, Dordrecht, the Netherlands), pp. 333-353.

Delgutte, B. (**1990a**). "Physiological mechanisms of psychophysical masking: Observations from auditory-nerve fibers," J. Acoust. Soc. Am. **87**, 791-809.

Delgutte, B. (**1990b**). "Two-tone rate suppression in auditory-nerve fibers: Dependence on suppressor frequency and level," Hear. Res. **49**, 225-246.

Delgutte, B. (**1995**). "Neural correlates of basic auditory psychophysical phenomena," Abstracts of the 18th Midwinter Meeting of the Association for Research in Otolaryngology, p. 146.

Delgutte, B. (**1996**). "Physiological models for basic auditory percepts," in *Auditory Computation*, edited by H.L. Hawkins, T.A. McMullen, A.N. Popper, and R.R. Fay (Springer-Verlag, New York), pp. 157-220.

Derleth, R.P., and Dau, T. (**2000**). "On the role of envelope fluctuation processing in spectral masking," J. Acoust. Soc. Am., in press.

Duifhuis, H. (**1973**). "Consequences of peripheral frequency selectivity for nonsimultaneous masking," J. Acoust. Soc. Am. **54**, 1471-1488.

Durlach, N.I., and Braida, L.D. (**1969**). "Intensity perception I: preliminary theory of intensity resolution," J. Acoust. Soc. Am. **46**, 372-383.

Durlach, N.I., Braida, L.D., and Ito, Y. (**1986**). "Towards a model for discrimination of broadband signals," J. Acoust. Soc. Am. **80**, 63-72.

Dye, R.H., and Hafter, E.R. (**1980**). "Just-noticeable differences of frequency for masked tones," J. Acoust. Soc. Am. **67**, 1746-1753.

Emmerich, D.S., Ellermeier, W., and Butensky, B. (**1989**). "A reexamination of the frequency discrimination of random-amplitude tones, and a test of Henning's modified energy-detector model," J. Acoust. Soc. Am. **85**, 1653-1659.

Erell, A. (**1988**). "Rate coding model for discrimination of simple tones in the presence of noise," J. Acoust. Soc. Am. **84**, 204-214.

Evans, E.F. (**1975a**). "Normal and abnormal functioning of the cochlear nerve," in *Sound Reception in Mammals*: Symp. Zool. Soc. London. 197, No. 37, edited by R.J. Bench, A. Pye, and J.D. Pye (Academic, London), 133-165.

Evans, E.F. (**1975b**). "The sharpening of cochlear frequency selectivity in the normal and abnormal cochlea," Audiol. **14**, 419-442.

Evans, E.F. (**1981**). "The dynamic range problem: Place and time coding at the level of cochlear nerve and nucleus," in *Neuronal Mechanisms of Hearing*, edited by J. Syka and L. Aitkin (Plenumm Press, New York), pp. 69-85.

Evans, E.F., Pratt, S.R., Spenner, H., and Cooper, N.P. (**1992**). "Comparisons of physiological and behavioral properties: Auditory frequency selectivity," in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, New York), pp. 159-169.

Fitzhugh, R. (**1958**). "A statistical analyzer for optic nerve messages," J. Gen. Physiol. **41**, 675-692.

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

Fletcher, H. (**1953**). *Speech and Hearing in Communication* (Van Nostrand, New York).

Florentine, M. (**1986**). "Level discrimination of tones as a function of duration," J. Acoust. Soc. Am. **79**, 792-798.

Florentine, M. (**1992**). "Effects of cochlear impairments and equivalent-threshold masking on psychoacoustic tuning curves," Audiology **31**, 241-253.

Florentine, M. and Buus, S. (**1981**). "An excitation pattern model for intensity discrimination," J. Acoust. Soc. Am. **70** 1646-1654.

Florentine, M. and Buus, S. (**1984**). "Temporal gap detection in sensorineural and simulated hearing impairments," J. Speech Hear. Res. **27**, 449-455.

Florentine, M., Buus, S., and Mason, C.R. (**1987**). "Level discrimination as a function of level for tones from 0.25 to 16 kHz," J. Acoust. Soc. Am. **81**, 1528-1541.

Florentine, M., Reed, C.M., Rabinowitz, W.M., Braida, L.D., Durlach, N.I., and Buus, S. (**1993**). "Intensity perception. XIV. Intensity discrimination in listeners with sensorineural hearing loss," J. Acoust. Soc. Am. **94**, 2575-2586.

Freyman, R.L., and Nelson, D.A. (**1986**). "Frequency discrimination as a function of tonal duration and excitation-pattern slopes in normal and hearing-impaired listeners," J. Acoust. Soc. Am. **79**, 1034-1044.

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

Geisler, C.D., and Rhode, W.S. (**1982**). "The phases of basilar-membrane vibrations," J. Acoust. Soc. Am. **71**, 1201-1203.

Gifford, R.H., and Bacon, S.P. (**2000**). "Contributions of suppression and excitation to simultaneous masking: Effects of signal frequency and masker-signal frequency relation," J. Acoust. Soc. Am. **106**, 1436-1451.

Gilbert, A.G., and Pickles, J.O. (**1980**). "Responses of auditory nerve fibres in the guinea pig to noise bands of different widths," Hear. Res. **2**, 327-333.

Gilkey, R.H., Robinson, D.E., and Hanna, T.E. (**1985**). "Effects of masker waveform and signal-to-masker phase relation on diotic and dichotic masking by reproducible noise," J. Acoust. Soc. Am. **78**, 1207-1219.

Glasberg, B.R., and Moore, B.C.J. (**1986**). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," J. Acoust. Soc. Am. **79**, 1020-1033.

Glasberg, B.R., Moore, B.C.J., and Bacon, S.P. (**1987**). "Gap detection and masking in hearing-impaired and normal-hearing subjects," J. Acoust. Soc. Am. **81**, 1546-1556.

Glasberg, B.R., and Moore, B.C.J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103-138.

Gold, T. (**1948**). "Hearing: The physical basis of the action of the cochlea," Proc. R. Soc. London (Biol. Sci.) **135**, 492-498.

Goldberg, J.M., and Brown, P.B. (**1969**). "Response properties of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: Some physiological mechanisms of sound localization," J. Neurophysiol. **32**, 613-636.

Goldstein, J.L. (**1990**). "Modeling rapid waveform compression on the basilar membrane as multiple-bandpass-nonlinearity filtering," Hear. Res. **49**, 39-60.

Goldstein, J.L. (**1991**). "Modeling the nonlinear cochlear mechanical basis of psychophysical tuning," J. Acoust. Soc. Am. **90**, 2267-2268.

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.

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

Grantham, D.W., and Yost, W.A. (**1982**). "Measures of intensity discrimination," J. Acoust. Soc. Am. **72**, 406-410.

Green, D.M. (**1988**). *Profile Analysis: Auditory Intensity Discrimination* (Oxford University Press, New York), Chapter 3.

Green, D.M., Kidd, G., Jr., and Picardi, M.C. (**1983**). "Successive versus simultaneous comparison in auditory intensity discrimination," J. Acoust. Soc. Am. **73**, 639-643.

Green, D.M., and Swets, J.A. (**1966**). *Signal Detection Theory and Psychophysics* [Wiley, New York, reprinted (1988) by Peninsula, Los Altos, CA].

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

Guinan, Jr., J.J. (**1996**). "Physiology of olivocochlear efferents," in *The Cochlea*, edited by P.J. Dallos, A.N. Popper, and R.R. Fay (Springer-Verlag, New York), pp. 435-502.

Hannley M., and Dorman M.F. (**1983**). "Susceptibility to intraspeech spread of masking in listeners with sensorineural hearing loss," J. Acoust. Soc. Am. **74**, 40-51.

Harrison, R.V. (**1981**). "Rate-versus-intensity function and related AP responses in normal and pathological guinea pig and human cochleas," J. Acoust. Soc. Am. **70**, 1036-1044.

Harrison, R.V., and Evans, E.F. (**1979**). "Cochlear fibre responses in guinea pigs with well defined cochlear lesions," in *Models of the Auditory System and Related Signal Processing Techniques*, edited by M. Hoke and E. de Boer, Scand. Audiol. Suppl, **9** 83-92.

Heinz, M.G., Carney, L.H., and Colburn, H.S. (**1999**). "Monaural, cross-frequency coincidence detection as a mechanism for decoding perceptual cues provided by the cochlear amplifier," J. Acoust. Soc. Am. **105**, 1023(A).

Heinz, M.G., Colburn, H.S., and Carney, L.H. (**2000a**). "Evaluating auditory performance limits: I. One-parameter discrimination based on computational auditory-nerve models," submitted to Neural Computation.

Heinz, M.G., Colburn, H.S., and Carney, L.H. (**2000b**). "Evaluating auditory performance limits: II. One-parameter discrimination with random level variation," submitted to Neural Computation.

von Helmholtz, H.L.F. (**1863**). *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik* (F. Vieweg und Sohn, Braunschweig, Germany). Translated as: *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, by A.J. Ellis from the 4th German edition, 1877, Leymans, London, 1885 (reprinted by Dover, New York, 1954).

Henning, G.B. (**1966**). "Frequency discrimination of random amplitude tones," J. Acoust. Soc. Am. **39**, 336-339.

Henning, G.B. (**1967**). "A model for auditory discrimination and detection," J. Acoust. Soc. Am. **42**, 1325-1334.

Hicks, M.L., and Bacon, S.P. (**1999a**). "Psychophysical measures of auditory nonlinearities as a function of frequency in individuals with normal hearing," J. Acoust. Soc. Am. **105**, 326-338.

Hicks, M.L., and Bacon, S.P. (**1999b**). "Effects of aspirin on psychophysical measures of frequency selectivity, two-tone suppression, and growth of masking," J. Acoust. Soc. Am. **106**, 1436-1451.

Hienz R.D., Aleszczyk C.M., and May B.J. (**1996**). "Vowel discrimination in cats: acquisition, effects of stimulus level, and performance in noise," J. Acoust. Soc. Am. **99**, 3656-3668.

Horst J.W., Javel E., and Farley G.R. (**1990**). "Coding of spectral fine structure in the auditory nerve. II: Level-dependent nonlinear responses," J. Acoust. Soc. Am. **88**, 2656-2681.

Houtgast, T. (**1977**). "Auditory-filter characteristics derived from direct-masking data and pulsation-threshold data with a rippled-noise masker," J. Acoust. Soc. Am. **62**, 409-415.

Houtsma, A.J.M. (**1995**). "Pitch perception," in *Hearing*, edited by B.C.J. Moore (Academic Press, New York), Chapter 8.

Isabelle, S.K. and Colburn, H.S. (**1991**). "Detection of tones in reproducible narrow-band noise," J. Acoust. Soc. Am. **89**, 352-359.

Javel, E., and Mott, J.B. (**1988**). "Physiological and psychophysical correlates of temporal

processing in hearing," Hear. Res. **34**, 275-294.

Jeffress, L.A. (**1948**). "A place theory of sound localization," J. Comp. Physiol. Psychol. **41**, 35-39.

Jenison, R.L. (**2000**). "Correlated cortical populations can enhance sound localization performance," J. Acoust. Soc. Am. **107**, 414-421.

Jesteadt, W., Wier, C.C., and Green, D.M. (**1977**). "Intensity discrimination as a function of frequency and sensation level," J. Acoust. Soc. Am. **61**, 169-177.

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.

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.

Joris, P.X., Carney, L.H., Smith, P.H., and Yin, T.C.T. (**1994**). "Enhancement of neural synchrony in the anteroventral cochlear nucleus. I. Responses to tones at the characteristic frequency," J. Neurophysiol. **71**, 1022-1036.

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

Kaernbach, C., and Demany, L. (**1998**). "Psychophysical evidence against the autocorrelation theory of auditory temporal processing," J. Acoust. Soc. Am. **104**, 2298-2306.

Keithley, E.M., and Schreiber, R.C. (**1987**). "Frequency map of the spiral ganglion in the cat," J. Acoust. Soc. Am. **81**, 1036-1042.

Kiang, N.Y.S. (**1984**). "Peripheral neural processing of auditory information," in *Handbook of Physiology, Section I: The Nervous System, Vol. III, Pt. 2*, edited by J.M. Brookhart and V.B. Mountcastle (American Physiological Society, Bethesda, MD), pp. 639-674.

Kiang, N.Y.S. (**1990**). "Curious oddments of auditory-nerve studies," Hear. Res. **49**, 1-16.

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

Kiang, N.Y.S., Baer, T., Marr, E.M., and Demont, D. (**1969**). "Discharge patterns of single auditory-nerve fibers as functions of tone level," J. Acoust. Soc. Am. **46**, 106.

Kiang, N.Y.S., Moxon, E.C., and Levine, R.A. (**1970**). "Auditory-nerve activity in cats with normal and abnormal cochleas," in *Sensorineural Hearing Loss, Ciba Symposium*, edited by G.E.W. Wolstenholme, and J. Knight (J. and A. Churchill, London), pp. 241-273.

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

Kidd, G., Jr., Mason, C.R., Brantley, M.A., and Owen, G.A. (**1989**). "Roving-level tone-in-noise detection," J. Acoust. Soc. Am. **86**, 1310-1317.

von Klitzing, R., and Kohlrausch, A. (**1994**). "Effect of masker level on overshoot in running- and frozen-noise maskers," J. Acoust. Soc. Am. **95**, 2192-2201.

Leek M.R., Dorman M.F., and Summerfield Q. (**1987**). "Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **81**, 148-154.

Leek M.R., and Summers V. (**1996**). "Reduced frequency selectivity and the preservation of spectral contrast in noise," J. Acoust. Soc. Am. **100**, 1796-1806.

Lentz, J.J., Richards, V.M., and Matiasek, M.R. (**1999**). "Different auditory filter bandwidth estimates based on profile analysis, notched noise, and hybrid tasks," J. Acoust. Soc. Am. **106**, 2779-2792.

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

Liberman, M.C., and Dodds, L.W. (**1984**). "Single-neuron labeling and chronic cochlear pathology. III. Stereocilia damage and alterations of threshold tuning curves," Hear. Res. **16**, 55-74.

Liberman, M.C., and Kiang, N.Y.S. (**1984**). "Single-neuron labeling and chronic cochlear pathology. IV. Stereocilia damage and alterations in rate- and phase-level functions," Hear. Res. **16**, 75-90.

Liberman, M.C., Dodds, L.W., and Learson, D.A. (**1986**). "Structure-function correlation in noise-damaged ears: A light and electron-microscopic study," in *Basic and Applied Aspects of Noise-Induced Hearing Loss*, edited by R.J. Salvi, D. Henderson, and Hamernik, R.P. (Plenum Press, New York), pp. 163-177.

Liberman, M.C., Dodds, L.W., and Pierce, S. (**1990**). "Afferent and efferent innervation of the cat cochlea: Quantitative analysis with light and electron microscopy," J. Comp. Neurol. **301**, 443-460.

Lin, T., and Goldstein, J.L. (**1995**). "Quantifying 2-factor phase relations in non-linear responses from low characteristic-frequency auditory-nerve fibers," Hear. Res. **90**, 126-138.

Lin, T., and Guinan, Jr., 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.

Lutfi, R.A., and Patterson, R.D. (**1984**). "On the growth of masking asymmetry with stimulus intensity," J. Acoust. Soc. Am. **76**, 739-745.

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

May, B.J., Le Prell, G.S., and Sachs, M.B. (**1998**). "Vowel representations in the ventral cochlear nucleus of the cat: Effects of level, background noise, and behavioral state," J.

Neurophysiol. **79**, 1755-1767.

McGill, W.J., and Goldberg, J.P. (**1968**). "A study of the near-miss involving Weber's law and pure-tone intensity discrimination," Perception Psychophysics **4**, 105-109.

McKee, S.P., Silverman, G.H., and Nakayama, K. (**1986**). "Precise velocity discrimination despite random variations in temporal frequency and contrast," Vision Res. **26**, 609-619.

Miller, G.A. (**1947**). "Sensitivity to changes in the intensity of white noise and its relation to masking and loudness," J. Acoust. Soc. Am. **19**, 606-619.

Miller, M.I., Barta, P.E., and Sachs, M.B. (**1987**). "Strategies for the representation of a tone in background noise in the temporal aspects of the discharge patterns of auditory-nerve fibers," J. Acoust. Soc. Am. **81**, 665-679.

Miller, R.L., Schilling, J.R., Franck, K.R., and Young, E.D. (**1997**). "Effects of acoustic trauma on the representation of the vowel $/\mathcal{E}/$ in cat auditory nerve fibers," J. Acoust. Soc. Am. **101**, 3602-3616.

Miller, R.L., Calhoun, B.M., and Young, E.D. (**1999**). "Discriminability of vowel representation in cat auditory nerve fibers after acoustic trauma," J. Acoust. Soc. Am. **105**, 311-325.

Moore, B.C.J. (**1973**). "Frequency difference limens for short-duration tones," J. Acoust. Soc. Am. **54**, 610-619.

Moore, B.C.J. (**1978**). "Psychophysical tuning curves measured in simultaneous and forward masking," J. Acoust. Soc. Am. **63**, 524-532.

Moore, B.C.J. (**1989**). *An Introduction to the Psychology of Hearing* (Academic Press, New York).

Moore, B.C.J. (**1995a**). "Frequency analysis and masking," in Hearing, edited by B.C.J. Moore (Academic Press, New York), Chapter 5.

Moore, B.C.J. (**1995b**). *Perceptual Consequences of Cochlear Damage* (Oxford University Press, New York).

Moore, B.C.J., and Glasberg, B.R. (**1981**). "Auditory filter shapes derived in simultaneous and forward masking," J. Acoust. Soc. Am. **70**, 1003-1014.

Moore, B.C.J., and Glasberg, B.R. (**1986**). "Comparisons of frequency selectivity in simultaneous and forward masking for subjects with unilateral cochlear impairments," J. Acoust. Soc. Am. **80**, 93-107.

Moore, B.C.J., and Glasberg, B.R. (**1987**). "Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns," Hear. Res. **28**, 209-225.

Moore, B.C.J., Poon, P.W.F., Bacon, S.P., and Glasberg, B.R. (**1987**). "The temporal course of masking and the auditory filter shape," J. Acoust. Soc. Am. **81**, 1873-1880.

Moore, B.C.J., and Glasberg, B.R. (**1989**). "Mechanisms underlying the frequency discrimination of pulsed tones and the detection of frequency modulation," J. Acoust. Soc. Am. **86**, 1722-1732.

Moore, B.C.J., and Vickers, D.A. (**1997**). "The role of spread excitation and suppression in simultaneous masking," J. Acoust. Soc. Am. **106**, 1436-1451.

Moore, B.C.J., and Oxenham, A.J. (**1998**). "Psychoacoustic consequences of compression in the peripheral auditory system," Psychol. Rev. **105**, 108-124.

Moore, B.C.J., Peters, R.W., and Stone, M.A. (**1999a**). "Benefits of linear amplification and multichannel compression for speech comprehension in backgrounds with spectral and temporal dips," J. Acoust. Soc. Am. **105**, 400-411.

Moore, B.C.J., Vickers, D.A., Plack, C.J., and Oxenham, A.J. (**1999b**). "Inter-relationship between different psychoacoustic measures assumed to be related to the cochlear active mechanism," J. Acoust. Soc. Am. **106**, 2761-2778.

Mountain, D.C. (**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.

Murugasu, E., and Russell, I.J. (**1996**). "The effect of efferent stimulation on basilar membrane displacement in the basal turn of the guinea pig cochlea," J. Neurosci. **16**, 325-332.

Nuttall, A.L., and Dolan, D.F. (**1996**). "Steady-state sinusoidal velocity responses of the basilar-membrane in guinea pig," J. Acoust. Soc. Am. **99**, 1556-1565.

Oxenham, A.J., and Moore, B.C.J. (**1994**). "Modeling the additivity of nonsimultaneous masking," Hear. Res. **80**, 105-118.

Oxenham, A.J., and Moore, B.C.J. (**1995**). "Overshoot and the "severe departure" from Weber's Law," J. Acoust. Soc. Am. **97**, 2442-2453.

Oxenham, A.J., Moore, B.C.J., and Vickers, D.A. (**1997**). "Short-term temporal integration: Evidence for the influence of peripheral compression," J. Acoust. Soc. Am. **101**, 3676-3687.

Oxenham, A.J. and Plack, C.J. (**1997**). "A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing," J. Acoust. Soc. Am. **101**, 3666-3675.

Oxenham, A.J., and Plack, C.J. (**1998**). "Suppression and the upward spread of masking," J. Acoust. Soc. Am. **104**, 3500-3510.

Parker, A.J., and Newsome, W.T. (**1998**). "Sense and the single neuron: Probing the physiology of perception," Annu. Rev. Neurosci. **21**, 227-277.

Parzen, E. (**1962**). *Stochastic Processes* (Holden-Day, San Francisco), Chapter 4.

Patterson, R.D. (**1976**). "Auditory filter shapes derived with noise stimuli," J. Acoust. Soc.

Am. **59**, 640-654.

Patterson, R.D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (**1987**). "An efficient auditory filterbank based on the gammatone function," paper presented at a meeting of the IOC Speech Group on Auditory Modeling at RSRE, December 14-15.

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

Patuzzi, R., and Robertson, D. (**1988**). "Tuning in the mammalian cochlea," Physiol. Rev. **68**, 1009-1082.

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

Payton, K.L. (**1988**). "Vowel processing by a model of the auditory periphery: A comparison to eighth-nerve responses," J. Acoust. Soc. Am. **83**, 145-162.

Peters, R.W., Moore, B.C.J., and Baer, T. (**1998**). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," J. Acoust. Soc. Am. **103**, 577-587.

Pickles, J.O. (**1988**). *An Introduction to the Physiology of Hearing* (Academic Press, New York).

Plack, C.J. (**1998**). "Beneficial effects of notched noise on intensity discrimination in the region of the severe departure," J. Acoust. Soc. Am. **103**, 2530-2538.

Plack, C.J. and Oxenham, A.J. (**1998**), "Basilar-membrane nonlinearity and the growth of forward masking," J. Acoust. Soc. Am. **103**, 1598-1608.

Plack, C.J. and Oxenham, A.J. (**2000**), "Basilar-membrane nonlinearity estimated by pulsation threshold," J. Acoust. Soc. Am. **107**, 501-507.

Plomp, R. (**1978**). "Auditory handicap of hearing impairment and the limited benefit of hearing aids," J. Acoust. Soc. Am. **63** 533-549.

Probst, R., Lonsbury-Martin, B.L., and Martin, G.K. (**1991**). "A review of otoacoustic emissions," J. Acoust. Soc. Am. **89**, 2027-2067.

Rabinowitz, W.M., Lim, J.S., Braida, L.D., and Durlach, N.I. (**1976**). "Intensity perception VI: Summary of deviations from Weber's law for 1-KHz tones," J. Acoust. Soc. Am. **59**, 1506-1509.

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

Recio, A., Rich, N.C., 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 Mössbauer technique," J. Acoust. Soc. Am. **49**, 1218-1231.

Rhode, W.S., and Greenberg, S. (**1992**). "Physiology of the cochlear nuclei," in *The Mammalian Auditory Pathway: Neurophysiology*, edited by A.N. Popper and R.R. Fay (Springer-Verlag, New York), pp. 94-152.

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

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

Rose, J.E., Gross, N.B., Geisler, C.D., and Hind, J.E. (**1966**). "Some neural mechanisms in the inferior colliculus of the cat which may be relevant to localization of a sound source," J. Neurophysiol. **29**, 288-314.

Rosen, S., and Baker, R.J. (**1994**). "Characterising auditory filter nonlinearity," Hear. Res. **73**, 231-243.

Rosen, S., Baker, R.J., and Darling, A. (**1998**). "Auditory filter nonlinearity at 2 kHz in normal hearing listeners," J. Acoust. Soc. Am. **103**, 2539-2550.

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. (**1973**). "Responses to noise of auditory-nerve fibers in the squirrel monkey ," J. Neurophysiol. **36**, 569-587.

Ruggero, M.A. (**1992**). "Physiology and coding of sound in the auditory nerve," in *The Mammalian Auditory Pathway: Neurophysiology*, edited by A.N. Popper, and R.R. Fay (Springer-Verlag, New York), pp. 34-93.

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.

Ruggero, M.A., Robles, L., and Rich, N.C. (**1992**). "Two-tone suppression in the basilar membrane of the cochlea: Mechanical basis of auditory-nerve rate suppression," J. Neurophysiol. **68**, 1087-1099.

Ruggero, M.A., Rich, N.C., Shivapuja, B.G., and Temchin, A.N. (**1996**). "Auditory-nerve responses to low-frequency tones: Intensity dependence," Aud. Neurosci. **2**, 159-185.

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.

Ryugo, D.K. (**1992**). "The auditory nerve: Peripheral innervation, cell body morphology, and central projections," in *The Mammalian Auditory Pathway: Neuroanatomy*, edited by D.B. Webster, A.N. Popper, and R.R. Fay (Springer-Verlag, New York), pp. 23-65.

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

Sachs, M.B., and Abbas, P.J. (**1974**). "Rate versus level functions for auditory nerve fibers in cats: Tone burst stimuli," J. Acoust. Soc. Am. **81**, 680-691.

Sachs, M.B., and Young, E.D. (**1979**). "Encoding of steady-state vowels in the auditory-nerve: Representation in terms of discharge rate," J. Acoust. Soc. Am. **66**, 470-479.

Sachs, M.B., and Young, E.D. (**1980**). "Effects of nonlinearities on speech encoding in the auditory nerve," J. Acoust. Soc. Am. **68**, 858-875.

Schalk, T.B., and Sachs, M.B. (**1980**). "Nonlinearities in auditory-nerve fiber responses to bandlimited noise," J. Acoust. Soc. Am. **67**, 903-913.

Scharf, B., Magnan, J., Collet, L., Ulmer, E., and Chays, A. (**1994**). "On the role of the olivocochlear bundle in hearing: A case study," Hear. Res. **75**, 11-26.

Schneider, B.A., and Parker, S. (**1987**). "Intensity discrimination and loudness for tones in notched noise," Perception & Psychophysics **41**, 253-261.

Schouten, J.F. (**1940**). "The residue and mechanism of hearing," Proc. Kon. Nedel. Akad. Wetenschap **43**, 991-999.

Schuknecht, H.F. (**1994**). "Auditory and cytocochlear correlates of inner ear disorders," Otolaryngol. Head. Neck. Surg. **110**, 530-538.

Sewell, W.F. (**1984**). "The effects of furosemide on the endocochlear potential and auditory-nerve fiber tuning curves in cats," Hear. Res. **14**, 305-314.

Shailer, M.J., Moore, B.C.J., Glasberg, B.R., Watson, N., and Harris, S. (**1990**). "Auditory filter shapes at 8 and 10 kHz," J. Acoust. Soc. Am. **88**, 141-148.

Shera, C.A., and Guinan, J.J., Jr. (**1999a**). "Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs," J. Acoust. Soc. Am. **105**, 782-798.

Shera, C.A., and Guinan, J.J., Jr. (**1999b**). "Frequency dependence of stimulus-frequency-emission phase: Implications for cochlear mechanics," in *Recent Developments in Auditory Mechanics*, edited by H. Wada, T. Takasaka, K. Ikeda, and K. Ohyama (World Scientific Press, Singapore), in press.

Shera, C.A., and Guinan, J.J., Jr. (**2000**). "Reflection-emission phase: A test of coherent reflection filtering and a window on cochlear tuning," Abstracts of the 23rd Midwinter Meeting of the Association for Research in Otolaryngology, 157.

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

Siebert, W.M. (**1968**). "Stimulus transformation in the peripheral auditory system," in *Recognizing Patterns*, edited by P.A. Kolers and M. Eden (MIT Press, Cambridge, MA)

pp. 104-133.

Siebert, W.M. (**1970**). "Frequency discrimination in the auditory system: place or periodicity mechanisms?," Proc. IEEE, **58**, 723-730.

Siebert, W.M. (**1973**). "What limits auditory performance?," in *Proceedings of Fourth International Biophysics Congress*, Vol. III. IUPAB, Moscow, pp. 399-414.

Siegel, R.A., and Colburn, H.S. (**1983**). "Internal and external noise in binaural detection," Hear. Res. **11**, 117-123.

Siegel, R.A. and Colburn, H.S. (**1989**). "Binaural processing of noisy stimuli: Internal/external noise ratios for diotic and dichotic stimuli," J. Acoust. Soc. Am. **86**, 2122-2128.

Sinex, D.G., and Havey, D.C. (**1986**). "Neural mechanisms of tone-on-tone masking: patterns of discharge rate and discharge synchrony related to rates of spontaneous discharge in the chinchilla auditory nerve," J. Neurophysiol. **56**, 1763-1780.

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

Spiegel, M.F., and Green, D.M. (**1981**). "Two procedures for estimating internal noise," J. Acoust. Soc. Am. **70**, 69-73.

van Tasell, D.J. (**1980**), "Perception of sound-formant transitions by hearing-impaired persons," Ear Hear. **1**, 130-136.

van Tasell, D.J. (**1993**). "Hearing loss, speech, and hearing aids," J. Speech Hear. Res. **36**, 228-244.

Teich, M.C., and Lachs, G. (**1979**). "A neural-counting model incorporating refractoriness and spread of excitation. I. Application to intensity discrimination," J. Acoust. Soc. Am. **66**, 1738-1749.

Teich, M.C., and Khanna, S.M. (**1985**). "Pulse-number distribution for the neural spike train in the cat's auditory-nerve," J. Acoust. Soc. Am. **77**, 1110-1128.

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

Turner, C.W., and van Tasell, D.J. (**1984**). "Sensorineural hearing loss and the discrimination of vowel-like stimuli," J. Acoust. Soc. Am. **75**, 562-565.

Turner, C.W., and Holte, L.A. (**1987**). "Discrimination of spectral-peak amplitude by normal and hearing-impaired subjects," J. Acoust. Soc. Am. **81**, 445-451.

Verschuure, J. and van Meeteren, A.A. (**1975**). "The effect of intensity on pitch," Acustica **32**, 33-44.

Viemeister, N.F. (**1974**). "Intensity discrimination of noise in the presence of band-reject noise," J. Acoust. Soc. Am. **56**, 1594-1600.

Viemeister, N.F. (**1983**). "Auditory intensity discrimination at high frequencies in the presence of noise," Science **221**, 1206-1208.

Viemeister, N.F. (**1988a**). "Psychophysical aspects of intensity discrimination," in *Auditory Function: Neurobiological Bases of Hearing*, edited by G.M. Edelman, W.E. Gall, and W.M. Cowan (Wiley, New York), pp. 213-241.

Viemeister, N.F. (**1988b**). "Intensity coding and the dynamic range problem," Hear. Res. **34**, 267-274.

Viemeister, N.F., and Bacon, S.P. (**1988**). "Intensity discrimination, increment detection, and magnitude estimation for 1-kHz tones," J. Acoust. Soc. Am. **84**, 172-178.

Wakefield, G.H., and Nelson, D.A. (**1985**). "Extension of a temporal model of frequency discrimination: Intensity effects in normal and hearing-impaired listeners," J. Acoust. Soc. Am. **77**, 613-619.

Weiss, T.F., and Rose, C. (**1988**). "A comparison of synchronization filters in different auditory receptor organs," Hear. Res. **33**, 175-180.

Westerman, L.A., and 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.

Wever, E.G. (**1949**). *Theory of Hearing* (Wiley, New York).

Wiederhold, M.L. (**1986**). "Physiology of the olivocochlear system," in *Neurobiology of Hearing*, edited by R.A. Altschuler, D.W. Hoffman, R.P. Robbin (Raven, New York), pp. 349-371.

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.

Winslow, R.L., Barta, P.E., and Sachs, M.B. (**1987**). "Rate coding in the auditory nerve," in *Auditory Processing of Complex Sounds*, edited by W.A. Yost and C.S. Watson (Erlbaum, New York), pp. 212-224.

Winslow, R.L., and Sachs, M.B. (**1988**). "Single-tone intensity discrimination based on auditory-nerve rate responses in backgrounds of quiet, noise, and with stimulation of the crossed olivocochlear bundle," Hear. Res. **35**, 165-190.

Winter, I.M., and Palmer, A.R. (**1991**). "Intensity coding in low-frequency auditory-nerve fibers of the guinea pig," J. Acoust. Soc. Am. **90**, 1958-1967.

Wong, J.C., Miller, R.L., Calhoun, B.M., Sachs, M.B., and Young, E.D. (**1998**). "Effects of high sound levels on response to the vowel /$\mathcal{E}$/ in cat auditory nerve", Hear. Res. **123**, 61-77.

Yates, G.K. (**1995**). "Cochlear structure and function," in *Hearing*, edited by B.C.J. Moore (Academic Press, New York), pp. 41-74.

Yin, T.C.T., Chan, J.C.K., and Carney, L.H. (**1987**). "Effects of interaural time delays

of noise stimuli on low-frequency cells in the cat's inferior colliculus. III. Evidence for cross-correlation," J. Neurophysiol. **58**, 562-583.

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

Young, E.D. (**1984**). "Response characteristics of neurons of the cochlear nuclei," in *Hearing Science*, edited by C.I. Berlin (College Hill Press, San Diego), pp. 423-460.

Young, E.D., and Sachs, M.B. (**1979**). "Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers," J. Acoust. Soc. Am. **66**, 1381-1403.

Young, E.D., and Barta, P.E. (**1986**). "Rate responses of auditory-nerve fibers in noise near masked threshold," J. Acoust. Soc. Am. **79**, 426-442.

Zhang, X., Heinz, M.G., Bruce, I.C., and Carney, L.H. (**2000**). "A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression," in prep.

Zwicker, E. (**1956**). "Die elementaren Grundlagen zur Bestimmung der Informationskapazität des Gehörs," Acustica **6**, 365-381.

Zwicker, E. (**1970**). "Masking and psychological excitation as consequences of the ear's frequency analysis," in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G.F. Smoorenburg (Sijthoff, Leiden).

Biographical Note

Michael Gregory Heinz
Born: November 25, 1970, Baltimore, Maryland
Education:

Attended Brown University from September 1988 - May 1992
Received Sc.B. in Electrical Engineering, May 1992
Attended Johns Hopkins University from September 1992 - May 1995
Received M.S.E. in Electrical and Computer Engineering, May 1994

Teaching:

At M.I.T. as a Teaching Assistant from September 1998 - December 1998
Acoustics of Speech and Hearing (HST.714/6.551)

Honors:

Elected to Tau Beta Pi in May 1991, Sigma Xi in May 1992

Publications:

Formby, C., Heinz, M.G., Luna, C.E., and Shaheen, M.K. (**1994**). "Masked detection thresholds and temporal integration for noise band signals," J. Acoust. Soc. Am. **96**, 102-114.

Heinz, M.G., Goldstein, Jr., M.H., and Formby, C. (**1996**). "Temporal gap detection thresholds in sinusoidal markers simulated with a multi-channel, multi-resolution model of the auditory periphery," Aud. Neurosci., **3**, 35-56.

Heinz, M.G. and Formby, C. (**1999**). "Detection of time- and band-limited increments and decrements in a random-level noise," J. Acoust. Soc. Am. **106**, 313-326.

Professional Memberships:

Acoustical Society of America
Association for Research in Otolaryngology
Institute of Electrical and Electronics Engineers