

**Physiological Correlates of
Temporal Envelope Perception**

Paul C. Nelson

ABSTRACT

The perception of speech and other behaviorally relevant sounds is strongly influenced by the amplitude-modulated (AM) properties of the signal (i.e., its temporal envelope). To gain a better understanding of the representation of AM stimulus features in the normal auditory system, experiments were carried out in psychophysics, physiology, and computational modeling, using comparable parametric spaces. Existing AM psychophysical data were supplemented with results from new translations of basic audio-frequency psychoacoustic paradigms into the AM-frequency domain to provide a wide range of stimulus conditions to test with the physiology and modeling. A reasonable working hypothesis that has emerged because of the qualitative parallels between AM perception and physiology is that neurons in the inferior colliculus (IC) are physiological implementations of the “effective” signal-processing style modulation filters derived from psychophysics. This testable hypothesis drove much of the experimental design and analysis, and several fundamental findings emerged as a result. To summarize the most basic results, we found that (1) the system appears to use a temporal code for AM at low modulation depths and an averaged or integrated response quantification at high depths, (2) temporal adaptation to AM is weak, as measured both perceptually and physiologically, and (3) some aspects of AM perception are context-dependent, a feature that is not present in either the IC responses or the simulated model responses. These findings have direct implications for models of AM processing and interpretation of proposed neural coding strategies. Other potential applications include the refinement of devices designed to assist the hearing impaired and the development of signal-processing strategies for speech recognition and audio coding systems.

Physiological Correlates of Temporal Envelope Perception

By

Paul Christian Nelson
B.S. Boston University, 2001
M.S. Syracuse University, 2003

DISSERTATION

Submitted in partial fulfillment of the requirements for the
degree of Doctor of Philosophy in Bioengineering
in the Graduate School of Syracuse University

May 2006

Approved _____

Dr. Laurel H. Carney

Date _____

Copyright 2006 Paul C. Nelson

All rights reserved

Table of Contents

Abstract	i
List of Figures	ix
Acknowledgments	xi
1 General introduction	1
1.1 Background and approach.....	1
1.2 Perceptual AM processing.....	1
1.3 Physiological AM processing.....	2
1.4 Connecting psychophysics and physiology.....	3
1.5 Summary.....	4
2 Cues for masked amplitude-modulation detection	5
Abstract.....	5
2.1 Introduction.....	5
2.2 Psychophysical experiment.....	9
2.2.1 Methods.....	9
2.2.1.1 Subjects and procedure.....	9
2.2.1.2 Apparatus and Stimuli.....	9
2.2.1.3 Conditions.....	10
2.2.2 Results & Discussion.....	10
2.2.2.1 Fixed-level modulation masker.....	10
2.2.2.2 Roving-level modulation masker.....	13
2.3 Modeling.....	15
2.3.1 Methods.....	15
2.3.1.1 Simulating threshold runs.....	15
2.3.1.2 Ruling out some potential cues.....	16
2.3.1.3 Trial-by-trial response analysis: decision-variable-reconstructed psychometric functions.....	17
2.3.2 Results and Discussion.....	18
2.3.2.1 Fixed-level modulation masker.....	18
2.3.2.2 Roving-level modulation masker.....	20
2.3.2.3 Equalized-envelope-energy modulation masker.....	22
2.3.2.4 Decision-variable-reconstructed psychometric functions.....	23
2.4 General discussion.....	26
2.4.1 Negative masking.....	26
2.4.2 Future directions.....	27
2.4.3 Summary.....	27

3 Comparison of intensity discrimination, increment detection, and comodulation masking release in the audio- and envelope-frequency domains.....	29
Abstract.....	29
3.1 Introduction.....	29
3.2 Experiment I. Intensity discrimination and increment detection in the envelope-frequency domain.....	32
3.2.1 Methods.....	32
3.2.1.1 Listeners.....	32
3.2.1.2 Apparatus and stimuli	32
3.2.1.3 Procedure	33
3.2.2 Results.....	34
3.2.2.1 Discrimination thresholds with gated and fringe presentation modes	34
3.2.2.2 Gated and fringe AM detection thresholds and comparison with ‘static’ intensity discrimination performance	35
3.2.3 Discussion.....	37
3.2.3.1 Adaptation and change detection	37
3.2.3.2 Relation to previous work.....	38
3.3. Experiment II. Tone-in-noise detection with a modulated masker in the envelope-frequency domain.....	39
3.3.1 Methods.....	39
3.3.2 Results.....	41
3.3.2.1 No release from masking in the envelope-frequency domain for 500-ms stimuli	41
3.3.2.2 Extending the time course of the slow masker fluctuations	42
3.3.3 Discussion.....	44
3.3.3.1 Relation to previous work.....	44
3.3.3.2 Interpreting time courses.....	45
3.4 Modeling.....	46
3.4.1 Methods.....	46
3.4.2 Results and Discussion	46
3.4.2.1 Audio-frequency level discrimination with gated and continuous carriers	46
3.4.2.2 Envelope-frequency detection and discrimination with gated and continuous carriers.....	51
3.4.2.3 CMR experiment.....	52
3.5 Summary and conclusions	55
4 Rate and timing cues for neural detection and discrimination of amplitude-modulated tones in the awake rabbit inferior colliculus	56
Abstract.....	56
4.1 Introduction.....	56
4.2 Methods.....	59
4.2.1 Animal preparation	59
4.2.2 Acoustic stimuli	60
4.2.3 Recording methods	60
4.2.4 Response classification and analysis	61

4.2.5 Implementation of the computational model	64
4.3 Results	64
4.3.1 Population pure tone responses and correlations with 100% SAM responses	64
4.3.2 Example MTFs and MDFs.....	70
4.3.2.1 Representative onset pure-tone response	70
4.3.2.2 Representative on+sustained pure-tone response	73
4.3.2.3 Representative sustained pure-tone response.....	75
4.3.2.4 Representative inhibited pure-tone response	77
4.3.3 Single-unit rate- and synchrony-based AM thresholds.....	79
4.3.3.1 Pure SAM detection.....	79
4.3.3.2 Masked SAM detection.....	82
4.3.3.3 SAM depth discrimination.....	84
4.3.4 Extending a computational model	86
4.4 Discussion	89
4.4.1 Temporal versus rate information	89
4.4.1.1 Temporal information: synchrony, phase, and intervals.....	89
4.4.1.2 Rate information: dynamic ranges and pooling.....	91
4.4.2 Comparing animal physiology and human psychophysics.....	92
4.4.3 Level effects.....	92
4.4.4 Temporal adaptation to AM.....	93
4.4.5 Implications for models of AM processing	94
4.4.6 Future directions	94
5 General discussion and summary.....	96
5.1 Applications	96
5.2 Perceptual versus physiological envelope processing	97
5.2.1. Similarities	97
5.2.2. Differences.....	98
5.2.2.1 General survey	98
5.2.2.2 Time constants: long in psychophysics, short in IC physiology.....	98
5.3 Overall summary.....	100

Appendix: A phenomenological model of peripheral and central neural responses to amplitude-modulated tones	101
Abstract.....	101
A.1 Introduction.....	101
A.1.1 Extracellular physiological responses to AM stimuli.....	102
A.1.2 Intracellular responses and studies of inhibition.....	103
A.1.3 Modeling.....	104
A.2 Methods.....	105
A.2.1 Auditory-nerve model.....	105
A.2.1.1 New signal path filters.....	105
A.2.1.2 Modified synapse model.....	106
A.2.1.3 Input and output signals and other details.....	107
A.2.2 Model cochlear nucleus cells.....	108
A.2.3 Model inferior colliculus cells.....	108
A.3 Results.....	109
A.3.1 AN model responses.....	109
A.3.2 Model CN cell responses.....	113
A.3.3 Model IC cell responses.....	115
A.3.3.1 MTFs & effect of varying time constants and delays.....	115
A.3.3.2 Effect of varying stimulus modulation depth.....	116
A.3.3.3 Effect of varying stimulus SPL.....	117
A.3.3.4 Model mechanisms and PSTs.....	118
A.3.4 Summary.....	120
A.4 Discussion.....	120
A.4.1 Limitations of the AN model.....	121
A.4.2 Model VCN cells: alternative mechanisms.....	121
A.4.3 Model IC cells.....	122
A.4.4 Future directions.....	124
Bibliography	126

List of Figures

2-1	Psychophysical masked AM detection thresholds at high masker depths.....	11
2-2	Masked AM detection thresholds over a wide range of masker depths.....	12
2-3	Roving-level masked AM detection thresholds.....	13
2-4	Equalized-envelope-energy (EEE) masked detection thresholds.....	15
2-5	Predicted model thresholds and comparison to fixed-level data.....	19
2-6	Predicted model thresholds and comparison to roving-level data.....	21
2-7	Predicted model thresholds and comparison to EEE data.....	22
2-8	Decision-variable-reconstructed psychometric functions.....	25
3-1	Gated-continuous differences in the audio- and envelope-frequency domains.....	34
3-2	Comparison of level discrimination and AM detection thresholds.....	36
3-3	Temporal envelopes of the stimuli used to test for AM-domain CMR.....	40
3-4	Thresholds in audio- and envelope-frequency CMR paradigms.....	42
3-5	Extending the time course of the masker fluctuations to observe AM CMR.....	43
3-6	Simulated responses to gated and continuous tone increments.....	47
3-7	Across-frequency rate-difference profiles for gated and continuous increments.....	49
3-8	Across-frequency information profiles for gated and continuous increments.....	50
3-9	Model responses to gated and fringe AM depth discrimination stimuli.....	52
3-10	Model responses to stimuli used in envelope-domain CMR paradigm.....	54
4-1	Pure-tone responses characteristics across the population of IC neurons.....	65
4-2	Weak inverse correlation between rate BMF and first-spike latency.....	67
4-3	Comparison of BMFs based on rate, synchrony, and synchronized rate.....	68
4-4	Sharpness of tuning in band-pass rate MTFs.....	69
4-5	Examples and prevalence of different rate MTF shapes.....	70
4-6	AM response properties of a neuron with an onset pure-tone PSTH.....	72
4-7	AM response properties of a neuron with an on + sustained pure-tone PSTH.....	74
4-8	AM response properties of a neuron with a sustained pure-tone PSTH.....	76
4-9	AM response properties of a neuron inhibited by pure tones.....	78
4-10	Neural AM detection thresholds based on rate and synchrony.....	80
4-11	SPL effects on neural AM detection thresholds.....	82
4-12	Neural AM detection thresholds in the presence of a competing masker.....	84
4-13	Comparison of psychophysical and rate-based neural AM depth discrimination.....	86
4-14	Comparison of SFIE model and the responses of two representative neurons.....	88
4-15	Response phase does not change systematically with AM depth.....	91
5-1	Responses of a typical IC neuron to AM forward masking stimuli.....	100

A-1	AN model responses and comparison to pure-tone PSTH data.....	107
A-2	Schematic diagram of the SFIE model.....	109
A-3	Comparison of model and actual AN MTFs and period histograms.....	110
A-4	Model and AN data: responses to variation in stimulus AM depth.....	111
A-5	Model and AN data: responses to variation in stimulus SPL.....	112
A-6	Model and AN data: responses to variation in stimulus f_m	113
A-7	Model and VCN data: responses to variation in stimulus f_m	114
A-8	IC model rate-MTFs with different assumed time constants.....	116
A-9	Model and IC data: responses to variation in stimulus m and f_m	117
A-10	Model and IC data: responses to variation in stimulus SPL and f_m	118
A-11	Model and IC data: PST histograms for a range of stimulus f_m	119
A-12	SFIE model summary: AN, VCN, and IC rate and synchrony MTFs.....	120

ACKNOWLEDGMENTS

I have been incredibly fortunate to have received encouragement, motivation, and technical assistance from a diverse support system that has continually developed over the last five-plus years. My advisor and friend, Laurel Carney, has always been the foundation and common thread among this group of people, starting with my time as her undergraduate advisee and senior thesis student at Boston University. In addition to providing the ultimate role model in terms of her enthusiasm, dedication, attention to detail, ethics, intuition, and compassion, Laurel also allowed me to constantly evolve and follow through on the questions that intrigued me. I am grateful to her for gradually treating me as a collaborator and for her patience as I slowly grew into that role.

For me, one of the most rewarding benefits of life as one of Laurel's graduate students was having the opportunity for collaboration with some of the leading research groups in the field of auditory temporal processing. I would like to thank Neal Viemeister at the University of Minnesota for giving me the chance to work in his lab for a summer as an oblivious undergrad; Mark Stellmack made that experience a positive one. Magdalena Wojtczak has provided valuable input regarding the design of the physiological experiments so as to maximize their potential interest to the psychophysical community. Neal, Mark, and Magda have had a steady and significant impact on the direction of the experiments and analyses described here.

I was also lucky to spend five months in 2004 in Torsten Dau's newly established Center for Applied Hearing Research at the Technical University of Denmark; the empirical data described in Chapter 3 of this thesis were collected while I was there. I am grateful to Torsten and to Stephan Ewert not only for their insightful guidance and participation in the experiments, but also for their friendship and generosity. Tobias Piechowiak and Gilles Pigasse made social life a lot easier for a ridiculous American in Europe, and Thomas Ulrich Christiansen helped immensely with working out the logistical details associated with a prolonged stay in a foreign country.

There is no doubt that the physiological data collection included in this project would not have happened without the help of Shigeyuki Kuwada at the University of Connecticut. One reason that this is true is because Shig was the first to describe the unanesthetized rabbit preparation for auditory physiology and openly share the details and encourage others (including Laurel) to adopt it. Shig also generously allowed me to visit his lab on two separate occasions for several days for the sole purpose of making electrodes to use in Syracuse for our project. Blagoje Filipovic was also enormously helpful with the dirty details of efficiently manufacturing big batches of electrodes.

My peers have provided valuable scientific and personal support along the way. At Boston University, Xuedong (Frank) Zhang, Qing Tan, and Neil Letendre helped me get started with a modeling study that paved the way for the remainder of the project. Fellow students at Syracuse University with a definite impact on the development of the work described here include Sean Davidson, Yan (Felicia) Gai, Mike Anzalone, B. Scott Jackson, and Satish Iyengar. I would also like to thank Lauren Calandruccio for always being willing to listen and communicate on my (strange) wavelength, and Deb Burnett for letting me consistently beat her in tennis.

The staff at the Institute for Sensory Research (ISR) helped with much of the data collection and provided valuable technical support. I am particularly grateful to Anita

Sterns for her daily assistance with the physiological recordings and animal handling, and to Lorraine Pawson for sharing her expertise and time in carrying out the histological protocols. Thanks, also, to Bill Dossert for constructing the electrode puller, Kristina Abrams for animal care support, and Annette Statum for keeping me in line in general.

I appreciate the feedback and encouragement provided by the rest of my thesis committee, including Robert L. Smith, Lisa Osadciw, and especially Karen Doherty. I also feel fortunate to have had thought-provoking interactions with the founder of the ISR, Josef Zwislocki, who amazed me with his wide-ranging knowledge and ability to explain and discuss concepts on many levels.

On a personal level, I can't imagine anything my family could have done to be more supportive over the last 5 (or 27) years. The unwavering confidence of my mom and dad has been at times baffling, but I am certainly thankful for it nonetheless. Finally, I'd like to thank Roni Pietrazak for playfully giving me that little nudge and extra incentive to finish what I started.

CHAPTER 1

General introduction

1.1 BACKGROUND AND APPROACH

Naturally occurring sounds vary in magnitude along two dimensions: frequency and time. It is common for the auditory system to first be described as a frequency-spectrum analyzer, but it is the temporal aspects of hearing that are strikingly different from those of the other senses. And while many of the foundations of psychoacoustics and auditory physiology have been built on studies of responses to static tonal stimulation, recent work in both fields has turned toward more behaviorally relevant sounds, dynamic in frequency, time, or both.

Two approaches can be used to gain a better understanding of responses to real-world stimuli in the auditory system. One tactic is to analyze responses to the sounds that are actually encountered in real-world situations, such as human speech or animal vocalizations. The other style is to build up to such complex sounds, by starting with more synthetic stimuli which can be described with a manageable parameter space. Such a gradual approach improves the chances of identifying the specific features of sounds that underlie the responses of listeners (in psychoacoustics) or neurons (in physiology). The experiments described here are examples of the second method: the stimuli used have time-varying amplitudes but essentially static frequency spectra. These amplitude-modulated (AM) signals represent a middle ground in complexity and practicality between pure tones and speech or other natural sounds.

When discussing temporal cues in sounds, it is necessary to differentiate features that change on different time scales. On a relatively short time scale, a sound's pressure waveform fluctuates about zero; these variations are referred to as *fine structure*, and are determined by the instantaneous (audio, or spectral) frequency of the stimulus. In contrast, a signal's temporal *envelope* changes on a longer time scale and is always positive; the envelope is a description of the slow variations in overall level that define the instantaneous amplitude. It is useful to describe complex temporal envelopes in the modulation- (or envelope-) frequency domain. Virtually all natural sounds have complex audio-frequency domain spectra and complex modulation-frequency domain spectra.

1.2 PERCEPTUAL AM PROCESSING

Certain parallels have emerged by comparing investigations of processing in the audio- and modulation-frequency domains. One clear example is the tuning, or frequency selectivity, that has been observed in both domains. Perceptual tuning is most easily demonstrated in masking experiments, where a noise masker makes detection of a signal most difficult when the masker and signal are close in frequency. Psychophysical masked-threshold patterns that clearly indicate tuning have been measured in both the audio-frequency and AM-frequency domains (e.g. Wegel and Lane, 1924; Houtgast, 1989).

Although perceptual frequency selectivity to both envelope and fine structure has become widely accepted in recent years, there is less agreement on the appropriate

response quantification of the putative modulation filters. Perhaps the most fundamental question that has been entertained in the literature is whether the perceptually relevant decision variable (DV) is averaged or integrated across time, or is computed from local features of a limited portion of the temporal envelope representation. Previous psychophysical work has addressed this problem by using different aspects of stimulus envelope (or signal-processing “black-box” model responses) to predict behavioral thresholds. Chapter 2 describes another contribution to this field by comparing psychophysical masked AM detection thresholds to predicted performance based on a battery of DVs. Some of the tested DVs are based on average (“long-term”) representations of the stimulus, and other DVs depend on local temporal features of the stimulus envelope. The AM detection psychophysical paradigms were adapted from audio-frequency tasks that were designed to minimize or disrupt the availability of long-term cues; this strategy allowed for a direct examination of the relative impact of such manipulations. In addition to previously proposed signal-based DVs, the responses of a physiological model designed to predict neural responses to envelope fluctuations (described in the Appendix) were also quantified and tested in their ability to predict behavioral performance. The neural model’s success in reproducing a specific aspect of the psychophysical data (where the signal-based DVs failed) led to a conclusion regarding the structure of the effective perceptual envelope-processor.

The experiments of Ch. 3 are related to those of Ch. 2 in that they were both direct translations of basic audio-frequency psychoacoustics into the envelope-frequency domain, and computational modeling was used in both chapters to analyze and interpret the empirical results. The psychophysical investigations described in Ch. 3 were geared towards understanding the perceptual salience of second-order envelope fluctuations (referred to as the “venelope” in the literature) in modulation depth discrimination and masked detection tasks. The results implied a relatively weak contribution of the venelope, in that the system did not appear to emphasize transient changes in the venelope as it did with the envelope. Another significant finding from Ch. 3 was the observed perceptual “sluggishness” in monaural AM processing, of interest because of its potential link between basic AM psychophysics and higher-order perceptual processes such as auditory grouping and streaming. In contrast to the quantitative threshold predictions found in Ch. 2, the simulations of Ch. 3 were more qualitative in nature, providing intuitive explanations for performance trends measured in the listeners. At the base of all of the modeling was a common assumed model structure, which was physiologically motivated but relatively straightforward in terms of its underlying assumptions and implementation details. Details of the model structure and comparisons to physiological data are included as an Appendix to this document.

1.3 PHYSIOLOGICAL AM PROCESSING

A natural question to ask before considering specific features of neural responses to envelope fluctuations in the inferior colliculus (IC) is, “Why should behavioral AM sensitivity be compared to physiological data from the IC?” Frankly, one reason that the recordings were made in the rabbit IC is purely practical: the physical organization of the Dutch-belted rabbit’s skull and brain allows for a straightforward approach to the IC, with minimal disruption of surrounding brain areas. Fortunately, there are many more scientifically interesting answers to the question posed above. The IC is an almost

obligatory synapse for ascending auditory information. As such, the information required to explain perception must be present, even in its sub-cortical processing position. However, the form that this information takes in the context of AM coding at the level of the IC is neither immediately clear nor widely agreed upon.

Previous work in the area has suggested that the qualitative response properties of IC neurons to sinusoidal AM are in some ways drastically different from those of more peripheral stations such as the auditory nerve (AN) and cochlear nucleus (CN). Specifically, many IC cells exhibit rate-tuning to modulation frequency: these neurons only respond strongly after stimulus onset if the envelope patterns occur within a limited range of fluctuation rates. This behavior is not typically seen in AN fibers or CN cells – often, the average discharge rate does not vary systematically with modulation frequency in most cell types found in these more peripheral locations. Instead, the temporal discharge pattern (synchronization to the envelope period, for instance) changes when the modulation frequency is varied. Central to the IC (i.e. thalamus and primary auditory cortex), rate-tuning is also observed, although the best modulation frequencies are usually lower than those observed in IC recordings.

Because of this apparent transition from rate responses that do not depend on changes in perceptually-relevant parameters (such as modulation depth and frequency) in the periphery to rate responses that are heavily dependent on the same variations in the central auditory system, there is a natural tendency to assume that information is being transformed into a “rate code,” and that this translation is perhaps complete by the level of the IC. Chapter 4 of this document provides a reminder that these tendencies to relate AM perception and average-rate-based neurophysiology are built on foundations of speculation. The missing link in the physiological literature is the fact that a key parameter space (modulation depth) has not been explored in a way that allows direct comparison to behavioral AM detection and discrimination thresholds. That crucial gap is filled in by the results in Chapter 4.

If envelope fluctuations are not represented by a pure rate code in the IC, then what is a more appropriate response quantification that might allow for a better prediction of perceptual results? Of course, one can imagine any number of *post-hoc* calculations; physiologists have historically preferred several alternatives. Most of them can be defined as representing some aspect of a Fourier analysis of the response. Perhaps the simplest or most intuitive such computation is commonly referred to as synchrony or synchronization to the envelope: this metric can be thought of as the relative level of the Fourier response component at the signal modulation frequency with respect to the DC (average rate) component level. Synchrony can also (equivalently) be conceptualized using circular statistics; if calculated in this context, the reported measure is often called vector strength. In Chapter 4, predicted neural detection thresholds based on changes in average rate and synchrony are compared to determine which features of the response are correlated with perceptual limitations and abilities.

1.4 CONNECTING PSYCHOPHYSICS AND PHYSIOLOGY

Relating envelope-processing physiology and psychophysics is the central theme of this dissertation. There are several issues that make direct quantitative comparison of existing physiological and psychoacoustical data inherently difficult. (1) Perceptual detection and discrimination abilities and limitations are based on the net response of the

entire system; single-cell physiology only provides information about isolated elements within the system. (2) The use of anesthesia in most physiological preparations raises questions of differences in attention- and alertness-related processes between behavioral and neural responses. (3) Different parameter variations are often used in the two fields. Psychophysicists obtain measures of responses to small changes in a few parameters of the signal, while physiologists usually attempt to map out a neuron's response properties over as wide a parameter space as time allows. (4) Human listeners have provided the majority of perceptual data; animal models are used in physiology because of the invasiveness of the procedures necessary to collect single-cell data.

The experiments described here directly address the first three of the four issues described above. (1) Neural responses were collected from a large population of single cells in a division of the brain that has been called a "hub" of auditory signal processing. Because of the large sample size, we were able to predict the modulation-processing performance of the system as a whole by considering the most sensitive responses to stimuli that were optimized for each individual cell. (2) Unwanted effects of anesthesia in the physiology were avoided by using an unanesthetized rabbit preparation that allows for daily recording sessions for several months. (3) Parameter variations in the physiology were chosen based on the psychophysics. In both approaches, the key parameter was the signal modulation depth, which was varied over comparable ranges in the two experimental techniques.

The fourth concern (human psychophysics versus animal physiology) is not addressed here. The most practical way to deal with the fact that the data from the two approaches are from different species would be to obtain behavioral thresholds in the animal model. This is a clear direction for future work. Another issue worth mentioning is that the rabbits were awake but not behaving during the recording sessions, while the human listeners were awake and behaving during the listening tasks. The effects of attention on neural responses are likely to increase as the auditory pathway is ascended; the data needed to determine the magnitude of the effect in the rabbit IC are lacking, mainly due to the practical difficulties associated with recording neural responses in a behaving animal. Despite these limitations, the empirical data presented in this thesis allow for the most direct comparisons of envelope perception and neural processing that have been provided to date.

1.5 SUMMARY

This thesis fuses perceptual and physiological experiments with computational modeling in an integrated effort to better understand the mechanisms of auditory temporal envelope perception. It is the interdisciplinary nature of the approach that leads to the key results, which suggest a new set of constraints for models of modulation processing and provide insight into the fundamental coding strategies used by the auditory system to represent AM.

CHAPTER 2

Cues for masked amplitude-modulation detection

ABSTRACT

The ability of psychoacoustic models to predict listeners' performance depends on two key stages: pre-processing and the generation of a decision variable. The goal of the current study was to determine the perceptually relevant decision variables in masked amplitude-modulation detection tasks in which the modulation depth of the masker was systematically varied. Potential cues were made unreliable by roving the overall modulation depth from trial to trial or were reduced in salience by equalizing the envelope energy of the standard and target after the signal was added. Listeners' performance was significantly degraded in both paradigms compared to the baseline (fixed-level modulation masker) condition, which was similar to those used in previous studies of masking in the envelope-frequency domain. Although this observation was broadly consistent with a simple long-term envelope power-spectrum model, there were several aspects of the data that were not. For example, the steep rate of change in threshold with masker depth and the fact that an optimal amount of envelope noise could enhance performance were not predicted by decision variables calculated directly from the stimulus envelope. A physiologically based processing model suggested a realistic nonlinear mechanism that could give rise to these second-order features of the data.

Note: This chapter was submitted to the J. Acoust. Soc. Am as a paper with the same title, by P. C. Nelson and L. H. Carney.

2.1 INTRODUCTION

Behaviorally relevant acoustic stimuli such as speech cannot be defined solely by their long-term audio-frequency composition. Temporal variations in a signal's spectrum and interactions between individual spectral components result in amplitude-modulated (AM) sounds. Viemeister (1979) used concepts from linear systems analysis as a framework to determine the effective temporal modulation transfer function (MTF) of the auditory system by measuring the just-noticeable modulation depth of a sinusoidally amplitude-modulated (SAM) noise for a range of modulation frequencies (f_m). Viemeister's approach has proven highly valuable as a first-order approximation of the system's (low-pass) properties and as a starting point for many other studies. For example, the modulation filter-bank model structure (e.g. Dau *et al.*, 1997a), which assumes that the envelope of the output of each audio-frequency channel passes through a bank of band-pass filters (broadly) tuned to f_m , is able to account for several perceptual findings that a low-pass pre-processor cannot explain (i.e. Dau *et al.* 1997a,b; Dau *et al.*, 1999; Ewert and Dau, 2000).

To predict psychophysical thresholds, the output of any model must be concisely quantified with some decision variable (DV). And while the pre-processing model structures are fundamentally different for the two models mentioned above, both the Viemeister (1979) model and the most recent implementations of the Dau model (the envelope power-spectrum model (EPSM), Ewert and Dau, 2000; Ewert *et al.*, 2002; Ewert and Dau, 2004) assume an average root-mean-square (rms) DV at the output of their envelope-filtering process. This

assumption has been shown to be reasonable for many types of AM-detection tasks, but it is not clear whether decision statistics that rely on local temporal envelope features (instead of average or long-term features) would be equally successful as quantifications of the model outputs.

The broad goal of the current set of experiments was to further elucidate which features of AM stimuli are perceptually salient and used by listeners in modulation detection tasks. To accomplish this, empirical data are presented that provide critical tests for various DVs. We have borrowed paradigms from the audio-frequency tone-in-noise (TIN) detection literature that highlight shortcomings of long-term decision statistics in the spectral-frequency domain (roving-level and energy-equalized TIN detection), and translated them into the modulation-frequency domain. Because the stimuli had envelope-frequency bandwidths smaller than the presumed modulation filter widths, the internal representation of the stimulus envelope was similar for the low-pass (Viemeister) model and the band-pass (Dau) model. An alternative model, developed to predict responses of inferior colliculus (IC) neurons to AM signals (Nelson and Carney, 2004), was tested alongside the previously proposed psychophysical (signal-processing) models. The working hypothesis was that a physiologically motivated model structure would shape the internal representation of the stimulus more like the real system than “effective” signal-processing models.

There are two reasons to consider a *masked* AM-detection task (instead of pure, or unmasked, AM detection) to test our hypotheses. First, several reasonable techniques can be used to adjust a given model’s unmasked detection abilities, which makes it difficult to dismiss one competing decision statistic over another. A more interesting reason is that real-world sounds have complex modulation spectra, so it is useful to consider envelope detection abilities and limitations for stimuli other than pure sinusoidal AM. Previous studies of masked-AM detection have focused on the effects of varying the frequencies of the signal and/or masker modulation (Houtgast, 1989; Bacon and Grantham, 1989; Strickland and Viemeister, 1996; Dau *et al.* 1997a; Ewert and Dau, 2000; Ewert *et al.*, 2002). Here, masker level (or masker modulation depth) was the only systematically manipulated stimulus dimension. Predicted signal-detection thresholds based on a battery of potentially relevant DVs were compared to the masked thresholds measured psychophysically. Because several decision devices predicted statistically similar thresholds, a more detailed analysis of the relationships between DVs and listener responses on a trial-by-trial basis was also carried out.

A subset of the potential perceptually relevant decision devices investigated in the present study can be introduced in the context of previous work. Perhaps the most influential and straightforward DV assumed in previous AM-coding work is the long-term rms energy measured at the output of some envelope-filtering process. Such a statistic can explain the shape of the temporal modulation transfer function (with low-pass pre-processing: Viemeister, 1979; Strickland and Viemeister, 1996) and the envelope-frequency selectivity observed in experiments measuring sinusoidal AM-detection thresholds in the presence of a narrowband-noise masker modulation applied to the same carrier (with band-pass pre-processing: Ewert and Dau, 2000; Ewert *et al.*, 2002). Moore and Sek (2000) measured detection thresholds for stimuli with three AM-frequency components for three different phase configurations, and found no dependence of thresholds on the components’ relative phases. This finding is also consistent with predictions of an average (rms) envelope statistic. Note that any local temporal structure

present in the stimulus (or its internal representation) is discarded with an average (rms) metric.

Strickland and Viemeister (1996) concluded that the ratio of the maximum value to the minimum value of the envelope (max/min) was the best predictor of listeners' thresholds in a tone-on-tone modulation masking experiment. In contrast to the rms statistic, which averages over the entire temporal waveform, max/min makes decisions based on only two points in the envelope representation. Crest factor (ratio of maximum envelope value to the envelope rms) represents a compromise in some sense: a single value of the waveform is normalized by an averaged value. Lorenzi *et al.* (1999) accounted for performance in a (supra-threshold) modulation component phase discrimination task by basing decisions on the crest factor of a low-pass filtered version of the envelope of their stimuli. DVs based on the higher-order moments of envelope amplitude distributions have also been tested in various envelope-processing tasks (i.e. skewness: Lorenzi *et al.*, 1999; kurtosis: Strickland and Viemeister, 1996).

Another aspect of a signal with a complex modulation spectrum is its envelope, or 2nd-order envelope (Shofner *et al.*, 1996; Ewert *et al.*, 2002; Lorenzi *et al.* 2001a, b). Envelope cues could potentially be used in modulation masking experiments, especially in conditions with tonal maskers and noise signals (Ewert *et al.*, 2002). This line of reasoning parallels results from audio-frequency tone and noise masking experiments in which envelope cues have been shown to have various effects on detection performance, depending on the masker-signal configuration (i.e. the asymmetry of masking; see Derleth and Dau, 2000). It is reasonable to hypothesize that envelope fluctuations may also provide a detection cue for conditions with sinusoidal signals and random maskers (as measured in the present study), especially when first-order envelope cues are made unreliable or completely removed.

As an alternative to signal-processing-based DVs, we also made threshold predictions based on a physiologically motivated model for neural responses to AM tones (Nelson and Carney, 2004). The average discharge rate of model inferior colliculus cells was tested as a physiologically realistic DV, alongside several of the signal statistics described above. In the model cells, discharge rate increases monotonically with signal modulation depth. Interactions between strong inhibitory and weaker excitatory inputs result in a 'hard' threshold modulation depth that limits the model's detection performance even in the absence of internal or external (stimulus-induced) noise sources. Model-cell rate MTFs are band-pass, with Q-values (measured at the half-maximal-rate points) of approximately 1. This broad tuning is realized in the physiological model by assuming different time courses in the effective low-pass filtering properties of inhibition and excitation. The Q-values are consistent with the signal-processing modulation filters derived recently by Dau and co-workers to predict several aspects of psychophysical envelope coding (Dau *et al.*, 1997a; Ewert and Dau, 2000; Ewert *et al.*, 2002). For the band-limited stimuli used in the present study, the filtering properties of the IC model cells have little effect on shaping the internal representation of the envelope. Again, the focus is on understanding the perceptually salient quantifications of the internally represented envelope (as opposed to testing the validity of a band-pass modulation filter vs. a 'smoothing' or low-pass modulation filter).

Independent of the chosen DV, simulations of psychophysical experiments must include some mechanism to limit model performance in the detection and discrimination

of deterministic stimuli (without external noise). The most common way to do this is to add some amount of internal noise, either to the internal representation, or to the final value of the decision statistic in each interval. Ewert and Dau (2004) have provided some insight into the appropriate statistical description of the internal noise relevant to envelope-processing tasks. They measured AM depth-discrimination thresholds for a wide range of standard depths, and found the Weber fraction for sinusoidal carriers to be independent of standard depth, as long as the standard was well above threshold. This can be accounted for in a model by assuming a constant *ratio* between the DVs in the target and standard interval at threshold, or by including an internal noise whose variance is proportional to the value of the assumed decision statistic. For low standard depths (i.e. -28 and -23 dB in $20 \log m$, where m is linear modulation depth), the situation was different. In this range, a constant increase in modulation depth was required to reach discrimination threshold (independent of the standard depth). This can be thought of as arising from a second type of internal noise process – one with a *fixed variance*, which dominates threshold measurements at low modulation depths. We will address Ewert and Dau’s (2004) findings, but we will also consider model predictions with a fixed-variance noise only, as a “best-case scenario” for the various decision statistics (i.e. if a decision statistic predicts higher thresholds than the listeners’ performance with the fixed-variance noise alone, it would certainly not be able to account for thresholds if the constant-ratio noise, or Weber-fraction noise, were also included).

Two specific paradigms that have been used in the audio-frequency domain to test the power spectrum model of masking were translated into the envelope-frequency domain in the current study: roving-level and equal-energy TIN detection. A within-trial rove in overall energy renders long-term rms cues unreliable, and models based on energy cues predict higher thresholds in a roving-level situation. The absolute amount of increase over fixed-level conditions depends on the rove range (Green, 1983). Kidd *et al.* (1989) found that roving the overall level by 32 dB in an audio-frequency TIN detection task did *not* have a significant effect on thresholds (for noise bandwidths greater than one-third of the psychophysically-measured auditory-filter bandwidth). In another paradigm that challenges energy-based audio-frequency models of masking, Richards and Nekrich (1993) measured the detectability of tones in narrow bands of masking noise after the energy in the two observation intervals was equalized. Pure long-term energy models predict that such a task would be impossible (for sub-critical bandwidths), but listeners performed the task reliably. Richards and Nekrich (1993) attributed their results to differences in the envelopes of the noise-alone and tone-plus-noise stimuli.

With this body of previous work in mind, we present here psychophysical masked-AM detection data and predicted thresholds based on a diverse set of decision statistics. Measured and simulated thresholds in roving-level and equal-energy conditions are compared to those from a baseline fixed-level masker condition, over a wide range of masker modulation depths.

2.2 PSYCHOPHYSICAL EXPERIMENT

2.2.1 Methods

2.2.1.1 Subjects and procedure

Four listeners with normal hearing participated in the experiment. Pure-tone thresholds for all of the subjects were less than 15 dB HL at octave frequencies between 500 Hz and 8 kHz. The authors served as two of the subjects (S2 and S3) and had experience in psychoacoustic measurements. The remaining two listeners had no previous experience. A training period, typically lasting three or more 1.5-hour sessions, was provided in which masked and absolute modulation thresholds were estimated using procedures similar to those described below. Further training was provided for the roving-level and equal-envelope-energy (EEE) conditions (see below). Data collection began when thresholds for a subject stabilized; there were typically no learning effects observed after 4-5 tracks on a given condition. The listeners became familiar with the different stimulus conditions, and were aware of the particular condition prior to the start of a track.

Masked SAM detection thresholds were obtained using an adaptive two-interval, two-alternative forced-choice (2I, 2AFC) procedure with a two-down, one-up stepping rule that estimated the modulation depth necessary for 70.7% correct detection (Levitt, 1971). This combination of parameters resulted in a threshold estimate that corresponded to a d' of about 0.8. In the randomly chosen target interval, the signal modulation was imposed along with a masker modulation on the tone carrier. The standard interval contained only the masker modulation. The signal modulation depth m at the beginning of a track was set well above threshold, and was varied initially by 3-dB steps (in $20 \log m$), and in steps of 1.5 dB after the first two reversals. The tracking procedure was run until 16 reversals were obtained; threshold for a given track was taken as the mean modulation depth of the last ten reversals. For each stimulus condition, thresholds presented here are the mean of four such estimates. Only tracks in which the standard deviation of the last 10 reversals was less than 3 dB were included in further analysis. Across-subject average data are presented as the mean and standard deviation of the 16 threshold estimates (4 listeners x 4 tracks per condition).

2.2.1.2 Apparatus and Stimuli

Subjects listened diotically through calibrated Sennheiser HD 580 headphones while seated in a sound-treated booth. Stimuli were digitally generated at a sampling rate of 48.828 kHz and converted to analog signals via the TDT System III two-channel real-time processor (RP2.1) digital-to-analog converter and the TDT System III headphone buffer (HB7), with its gain set to -27 dB (to eliminate background noise). Signals were generated and presented with visual feedback using MATLAB. Noise waveforms were saved for both intervals on every trial (by recording random-number-generator seeds) so that the exact stimuli could be reconstructed for *post hoc* analysis (see Modeling Methods).

The two intervals were each 600 ms in duration including 50-ms \cos^2 ramps, and were presented with a 500-ms inter-stimulus interval. Both the sinusoidal signal (always in sine phase) and the narrow-band Gaussian-noise masker modulation were applied to the envelope of a 2800-Hz tone carrier for the entire duration of the stimulus. The signal frequency was 64 Hz; the masker was centered on the signal frequency and had a

bandwidth of 32 Hz. These parameters were chosen to satisfy several specific constraints. First, the modulation frequencies were low enough to avoid the introduction of audio-frequency spectral resolution cues that arise when the sidebands generated by modulation are remote from the carrier frequency component. In addition, the bandwidth of the masker was wide enough to allow for the slower 2nd-order (envelope) fluctuations to fall within a range that could potentially be detected in a 600-ms duration signal (the envelope energy was concentrated around 10 Hz). The AM signal and masker parameters were also influenced by modeling considerations, as described below.

Two statistically independent realizations of the masker were generated for the standard and target intervals. An additive approach, as opposed to the multiplicative one used in several related studies (Ewert and Dau, 2000; Ewert *et al.*, 2002; Houtgast, 1989), was used to combine the signal and masker. This allowed for more careful control of the envelope-frequency domain magnitude spectrum (i.e. addition of time-domain waveforms results in the addition of their frequency-domain spectra, whereas multiplication of time waveforms is equivalent to a convolution of their frequency spectra). The equation for the stimuli in both intervals is:

$$s(t) = c\{\sin(2\pi f_c t)[1 + m \sin(2\pi f_m t) + M(t)]\} ,$$

where f_c is the carrier frequency, m is the stimulus modulation depth (zero in the standard interval), f_m is the signal modulation frequency, and $M(t)$ is the masker waveform (zero when measuring absolute thresholds). Masker level was defined in terms of the root-mean-square (rms) of $M(t)$. The compensation factor c was included so the overall power in both intervals was equivalent to that of a 65-dB SPL pure tone. Every stimulus was checked for over-modulation caused by the stochastic nature of the narrowband maskers; no envelope with a modulation index greater than one was presented to the listeners.

2.2.1.3 Conditions

The acoustic stimuli used in this experiment were similar to those described in Ewert *et al.* (2002). Different parameter variations, as well as minor procedural modifications distinguish the two studies. Ewert *et al.* (2002) focused on frequency effects (of both signal and masker). Here, we explicitly considered the effect of masker level (i.e. the masker rms modulation depth) and the consequences of systematically controlling the availability of envelope-detection cues. Thresholds for three conditions were measured: (1) SAM detection with a fixed-level modulation masker, (2) SAM detection with a random 10-dB within-trial rove in masker level, and (3) SAM detection with EEE in the standard and target intervals (after the signal was added). The roving-level condition effectively made envelope energy an unreliable cue; the EEE condition strongly attenuated first-order envelope energy differences as a cue for detection. Thresholds from the fixed-level condition provided a baseline for evaluating the consequences of these two manipulations. Note that the fixed-level condition was comparable to those of previous studies (i.e. Ewert *et al.*, 2002).

2.2.2 Results & Discussion

2.2.2.1 Fixed-level modulation masker

General trends in the results were similar across the four listeners, but individual sensitivity varied considerably in the masked-AM detection task. Both individual (upper panel) and mean thresholds (lower panel) are shown in Fig. 2-1 for the detection of a 64-Hz sinusoidal modulation in the presence of an additional masker modulation. The

masker had a bandwidth of 32 Hz, and was always centered on the signal frequency. Signal thresholds are shown for a 10-dB range of masker modulation depths.

Thresholds increased monotonically as the masker level increased over this range of masker depths. Listener S4 was less sensitive than the other three subjects, while the thresholds of subject S3 increased at a rate less than 1 dB/dB. Mean thresholds were 1-2 dB ($20 \log m$) lower than the masker modulation depth (dB rms), and increased with a slope of 1 dB/dB. These results are consistent with those of Houtgast (1989), who measured detection thresholds for an 8-Hz sinusoidal signal modulation in the presence of a 2.8-Hz bandwidth masker modulation. In contrast with the present study, Houtgast (1989) combined the signal and masker multiplicatively and imposed them on a noise carrier.

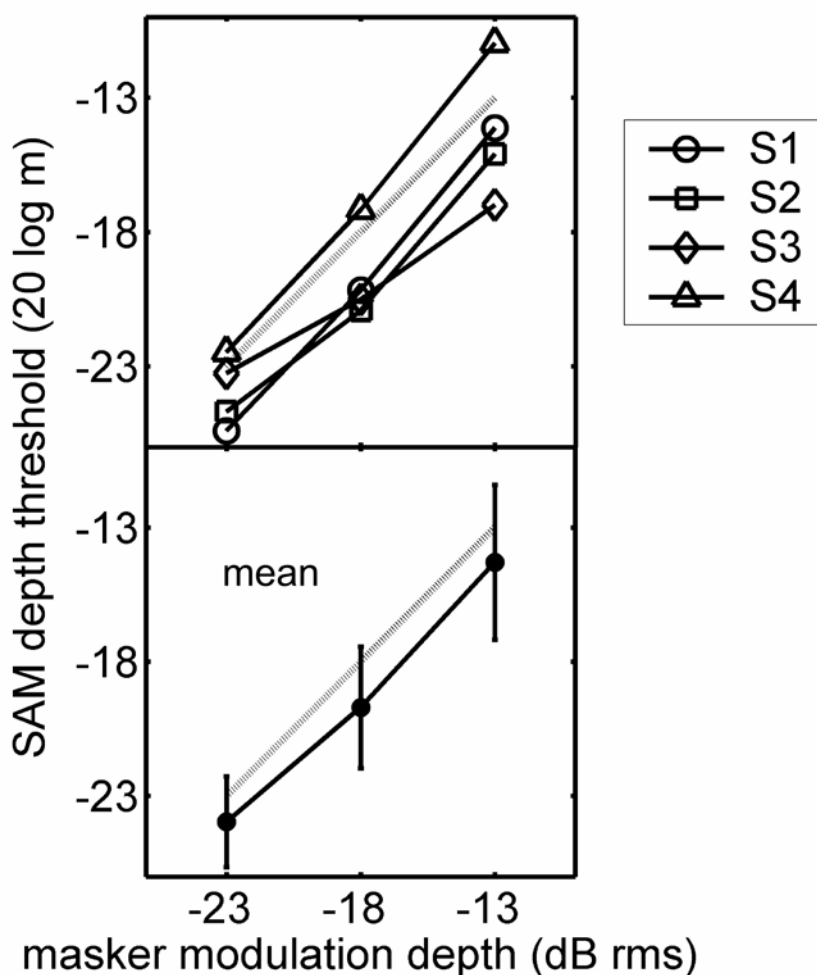


FIG 2-1. Individual (top) and mean (bottom) masked-SAM detection sensitivity. Thresholds at these supra-threshold masker depths increased at a rate of about 1 dB ($20 \log m$) per 1 dB (masker rms); the dashed lines in the two panels serve as a reference with a 1 dB / dB slope. Signal $f_m = 64$ Hz; masker bandwidth = 32 Hz, centered on signal frequency; SPL = 65 dB; carrier $f_c = 2800$ Hz; duration = 600 ms. Standard deviations of individual listener threshold estimates were between 2 and 4 dB (error bars omitted for clarity).

Somewhat less intuitive are the patterns of thresholds measured for lower-level maskers. In efforts to map out the entire range of masker modulation depths that produced masking while still avoiding over-modulation, for the purpose of the roving-level experiment (below), it became clear that some of the listeners' masked thresholds were *lower* than their pure AM-detection thresholds. This 'facilitation' is illustrated in Fig. 2-2 in the form of non-monotonic threshold vs. masker level functions for two of the four listeners (S2 and S3). The thresholds for the three right-most points in each function are re-plotted from Fig. 2-1. Unmasked detection thresholds ranged from -25 dB to -30 dB (masker level = -99 dB rms; left-most point on each plot), and were consistent with previously-reported pure-tone SAM detection thresholds for comparable f_c , f_m , and SPL (i.e. Kohlrausch *et al.*, 2000). The external variability of the noise maskers began to influence thresholds between -40 dB and -30 dB rms. The presence of the region of facilitation was not related to absolute sensitivity to AM; the two subjects that exhibited the clearest facilitation had the lowest (S2) and highest (S3) thresholds in unmasked AM detection. In addition, the masker level that resulted in the most facilitation was the same for both listeners (-28 dB rms).

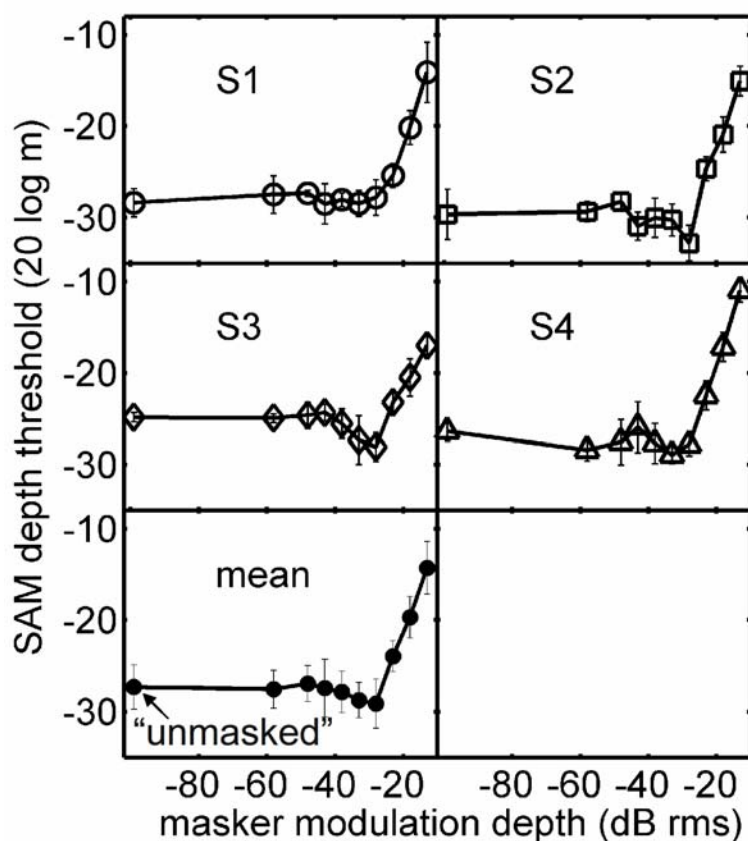


FIG 2-2. Masked-detection thresholds for a wide range of masker modulation depths. Two of the listeners (S2 and S3) exhibited a non-monotonic dependence of sensitivity on masker level; their thresholds were lower for a masker level of -28 dB than in the unmasked condition. The three right-most points in each panel are re-plotted from Fig. 2-1; these masker levels consistently caused 'positive' masking without causing over-modulation.

Strickland and Viemeister (1996) and Bacon and Grantham (1989) reported facilitation in some of their tone-on-tone modulation masking conditions, when the frequency of the masker was well below that of the signal. They accounted for this type of negative masking by assuming that their listeners were able to attend to the valleys of the masker when its fluctuations were slow enough, resulting in a temporally localized larger effective modulation depth. The facilitation illustrated in Fig. 2-2 is fundamentally different: the masker and signal occupy the same frequency region, and inherent fluctuations in the narrowband masker made the timing of its valleys unpredictable to our listeners. Also, the negative masking effects in previous studies increased as the masker modulation depth increased; the effect observed in the current study is only measurable at very low masker depths (near or even below detection thresholds). Potential mechanisms underlying on-frequency, low-level noise-masker facilitation will be evaluated in the Modeling section below.

2.2.2.2 Roving-level modulation masker

The effect of introducing a random 10-dB within-trial rove in masker level on listeners' thresholds is shown in Fig. 2-3. Because the masker modulation depth was different in every interval, it was necessary to track on the level of the signal with respect to the level of the noise (i.e. the difference between the two in dB). Detection thresholds are plotted for a fixed-level (−18 dB rms) noise masker (filled bars) and for the roving level (uniformly distributed from −23 to −13 dB rms) noise masker (open bars). Individual and across-subject average thresholds are included in the figure.

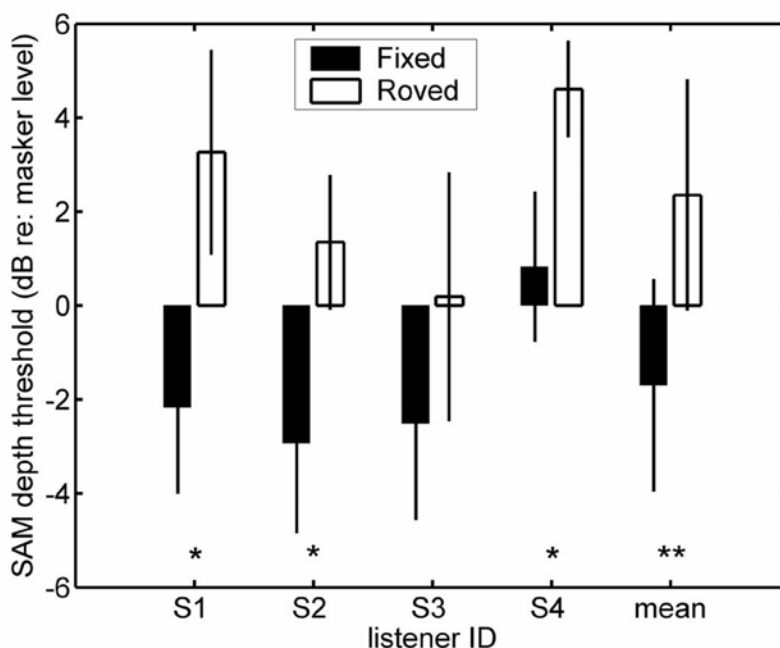


FIG 2-3. Comparison of fixed-level thresholds at a masker depth of −18 dB (closed bars) and roving-level thresholds, where the masker depth was randomly chosen from a uniformly-distributed 10-dB range centered on −18 dB (open bars). Asterisks indicate cases where the difference between the fixed- and roving-level thresholds was significant (* $p < 0.02$; ** $p < 0.0001$).

In general, thresholds in the roving-level condition were 3-5 dB higher than those in the fixed-level case. The effect was significant (t-test, $p < 0.02$) for three of the four individual listeners, and highly significant ($p < 0.0001$) when the across-subject mean and variance was considered. The 10-dB rove in masker level increased the mean thresholds by 4 dB. Unfortunately, the small dynamic range of AM maskers precluded the use of larger rove ranges in the present study (i.e. the masker must be intense enough to cause masking, but not so strong as to result in over-modulation, especially in the signal interval). Despite the limitations, the significant effect of this relatively small rove range contrasts with results from audio-frequency TIN detection experiments, where even a 32-dB rove in masker level did not significantly affect listeners' thresholds (except at the narrowest bandwidth tested, Kidd *et al.*, 1989). The convincing results of Kidd *et al.* (1989) provide a critical test that challenges the power spectrum model of masking in the audio-frequency domain. Qualitatively, models which assume the long-term energy of the (AC-coupled) envelope as the perceptually relevant quantity (e.g. Viemeister, 1979; Ewert and Dau 2002) are not seriously challenged by the current results obtained with the roving-level modulation masker. A more careful analysis of this general statement is provided in the Modeling section of this paper.

2.2.2.3 Equalized-envelope-energy modulation masker

As an alternative approach to test energy-based models, (long-term) first-order envelope cues were *removed* by forcing the rms modulation depth of the standard and target intervals to be the same, regardless of the level of the signal (in $20 \log m$). The task was the same as in the fixed-level and roving-level condition: listeners chose the interval containing the sinusoidal signal modulation. Pure long-term energy decision statistics did not provide any cues for detection in this paradigm (as long as the masker bandwidth was within the pass-band of the envelope-filtering process). Over-modulation was not an issue in the EEE condition: the average depth in both standard and target intervals was determined by the depth of the masker-alone modulation. Qualitatively, the signal-interval envelope fluctuations became more sinusoidal as m increased (but the overall depth was the same in both the standard and target envelopes).

Example waveforms for a -13 dB rms standard depth are illustrated in Fig. 2-4(a) along with individual listener and mean thresholds for a 10-dB range of masker-alone modulation depths [Fig. 2-4(b)]. Note that absolute thresholds are not plotted in Fig. 2-4; instead, increases in threshold over the corresponding fixed-level masker condition are shown. The key result illustrated in Fig. 2-4 is that the listeners were able to perform the task, although measured thresholds were about 10 dB worse on average than in the fixed-level condition (in which the overall rms modulation depth was allowed to naturally vary across intervals). Perhaps the most striking aspect of the individual thresholds is the high variability both within and across listeners (note the expanded scale of the y-axis). Anecdotally, the task became considerably more difficult in the EEE condition, and listeners reported the use of a very different strategy compared to that employed in the fixed-level case. The following sections quantitatively explore potential cues that could explain thresholds in all three masker configurations (fixed-level, roving-level, and EEE).

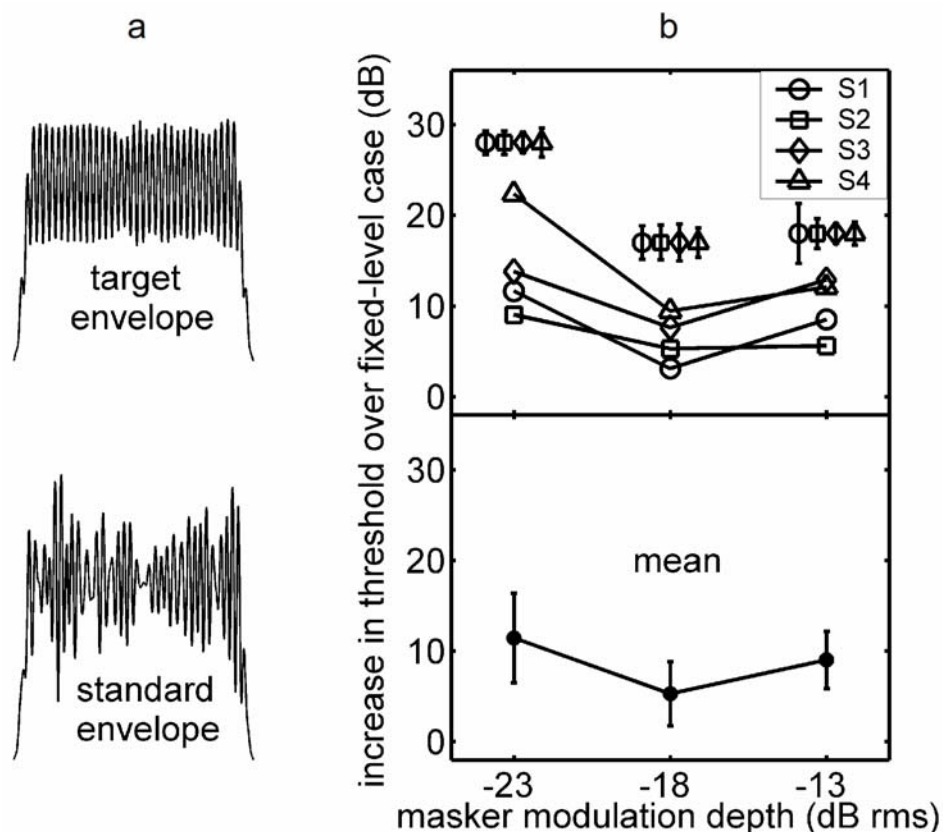


FIG 2-4. Effect of equalizing the overall modulation depth in the two observation intervals after the signal was added. a) Example waveforms: Masker depth = -13 dB rms; signal added to target interval masker at a $+20$ dB SNR. b) Increases in thresholds over comparable fixed-level conditions are plotted (absolute thresholds are shown along with model predictions in Fig. 2-7). Individual listener standard deviations are plotted above the corresponding means (b, top panel).

2.3 MODELING

2.3.1 Methods

2.3.1.1 Simulating threshold runs

Masked-detection thresholds were determined for each assumed DV using the same procedure, stimuli, and conditions as described in the psychophysical methods. The mean and standard deviation of 16 estimates were obtained (for comparison to the 4 subjects \times 4 repetitions measured psychophysically). Only the steady-state portion of the envelope (the central 500 ms) was used to compute decision statistics.

Several DVs were calculated from the stimuli's Hilbert envelopes and used in simulated tracking procedures: (1) root-mean-square (rms) AC-coupled envelope energy, (2) average local modulation depth, (3) average rate of model IC cell, (4) crest factor, (5) maximum local modulation depth, and (6) max/min ratio. The first three DVs can be considered 'long-term', as they are based on an integrated representation of the entire

steady-state envelope. The remaining three statistics assume that short-term fluctuations are salient perceptual cues in the masked modulation-detection task.

DVs based on local modulation depths (2 and 5 in the list above) were calculated from a running ratio of the ac to dc envelope energy in each cycle of the signal modulation. More specifically, the max/min ratio was computed for every cycle of the steady-state envelope, and divided by the mean value of the envelope for that same time period. From the resulting 32 points (500 ms of a 64-Hz signal), an average value was computed (in the case of the average local depth DV), or the maximum value was extracted (for the maximum local depth DV).

Because the envelope-frequency spectra of the stimuli were always within the pass-band of a 64-Hz modulation band-pass filter and a modulation low-pass filter with a cutoff frequency of 150 Hz, there was no filtering applied to the Hilbert envelopes before determining the signal-based decision statistics. In this respect, predictions based on the model IC cell average rate are different from the others: rate MTFs of simulated IC neurons are band-pass. Again, this difference has very limited consequence for the stimuli presented in this study (but see EEE predictions). Only the cell tuned to the signal frequency (64 Hz) was considered. Implementation details for the physiological model were the same as in Nelson and Carney (2004), except the convolution of alpha functions and instantaneous rate functions was carried out in the frequency domain for computational reasons (see website for code: web.syr.edu/~lacarney). Model parameters were matched to those describing the cell in Figs. 8(c), 9, 10, and 11 in Nelson and Carney (2004), except the auditory-nerve (AN) characteristic frequency (CF) was set to the stimulus carrier frequency ($f_c=2800$ Hz), and the strength of inhibition (re: excitation) at the level of the IC model cells ($S_{IC,INH}$) was set to 1.1. Also, the stimulus presentation level was set to 24 dB SPL for simulations with the model IC cells. This SPL resulted in near-maximal synchrony at the level of model AN fibers (compared to responses at other SPLs), which translates into higher rates in model IC cell responses. Similar responses would be expected from off-CF AN fibers for stimuli presented at higher SPLs [e.g. Joris and Yin (1992), their Fig. 8(c)]. The decision to use low-SPL stimuli as inputs to the physiological model carries with it an assumption that the central nervous system is able to weight responses from peripheral channels that are least affected by saturation and/or compression. Such assumptions are at least indirectly supported by psychophysical work that shows improvements in modulation detection performance as the overall SPL is increased (i.e. Kohlrausch et. al., 2000) despite the fact that saturation and compression are likely to be affecting near-CF responses.

Since the stimuli were deterministic, it was necessary to limit model performance in conditions without a masker (pure AM detection, or a masker level of -99 dB) by adding a Gaussian random variable to the final values of the decision statistics in each interval. The variance of this noise was adjusted to yield pure SAM detection thresholds of about -27 dB, and was held constant for all of the experimental conditions once it was determined.

2.3.1.2 Ruling out some potential cues

Several DVs were unable to predict trends or absolute thresholds comparable to the listeners in any of the masked-AM detection conditions (except at the lowest masker depths, where the task is essentially pure AM detection). Because of their poor performance in general, simulations based on skewness (the 3rd central moment of the

envelope amplitude distribution), kurtosis (the 4th moment), and envelope fluctuations are not included in the figures presented here. Predicted thresholds based on these three DVs were too high, often immeasurable, and also highly variable across the 16 estimates. The skewness of the point-by-point envelope distribution did not reliably change when the sinusoidal signal was added for any of the noise levels in this task. Values of envelope standard deviation and kurtosis consistently decreased when the tone was added, but only at modulation depths much higher than the listeners' thresholds.

Another decision statistic that was unable to predict performance in the current set of experiments was one based on a quantification of the instantaneous frequency (IF) of the envelope time waveform. A noise-alone modulated carrier would be expected to have higher variability in its envelope IF than that of a tone-plus-noise-modulated carrier. The bandwidth of the modulation maskers (32 Hz), along with external stimulus variability and the use of relatively low modulation depths, made envelope IF an unreliable cue for SAM detection in the present study. Tracking simulations based on a target-interval drop in envelope IF variance resulted in predicted thresholds that were highly variable across tracks and higher than the listeners' thresholds.

Several recent studies have used a cross-correlation calculation between a template response derived at some supra-threshold signal level and the "current" stimulus representation as a method to quantify model responses and predict psychophysical thresholds (i.e. Dau *et al.*, 1997a,b; Ewert and Dau, 2004). This technique is optimal in the sense that the signal is assumed to be known exactly, both in terms of its magnitude and phase, and has been shown to reasonably predict performance in a wide variety of psychoacoustical experiments, including modulation detection and modulation masking (Dau *et al.*, 1997a,b; Dau *et al.*, 1999). Accounting for performance with such a strategy comes at the expense of being able to identify the specific perceptually relevant features of the stimulus or response. As such, a correlation-based DV is fundamentally different from the other DVs considered in this study. It is interesting to point out that a template-based approach predicts no effect of masker-depth-roving or energy-equalization (simulations not shown). Furthermore, the templates for all three masker conditions (averaged over many noise tokens) are essentially identical: the sinusoidal signal is the only portion of the stimulus that remains after averaging many repetitions of the signal-plus-noise waveform. This aspect of the cross-correlation model contrasts with the tracking simulations and qualitative listener comments that all point to the use of a different cue in the EEE conditions as compared to the fixed-level and roving-level masker conditions.

2.3.1.3 Trial-by-trial response analysis: decision-variable-reconstructed psychometric functions

The ability of a given decision statistic to track on realistic detection thresholds is necessary but not sufficient as a requirement for concluding that listeners are using the cue on a trial-by-trial basis. To further test each potential cue, several key conditions were analyzed by comparing the values of different DVs that were derived from the exact stimuli presented to the listeners during the tracking procedure. The first step in this analysis (which required no additional time of the listeners) was to save the standard and target interval waveforms as the listeners performed the 2I,2AFC task (in practice, only the MATLAB random number generator seeds were saved). To re-create the stimuli presented during the track, the only other variables needed were the modulation depths of

the signal and masker in both intervals, and the masker-level configuration (i.e. fixed, roved, or equal-energy).

The logic behind the decision-variable-reconstructed psychometric (DVRP) functions was as follows: if listeners were using the assumed DV as a primary cue, then their performance should have systematically depended on the magnitude of the difference between the DVs calculated from the two intervals. If there was no difference, performance should have been at chance; for big differences, performance should have been near 100% correct. When percent correct was plotted as a function of the difference in DVs (target interval DV – standard interval DV), two shapes of the DVRP function were possible if the DV was salient and being consistently used by the listeners: monotonically increasing with high values of percent correct at large positive differences, and monotonically decreasing with the best performance when the difference between the target and standard DV was negative (i.e. the presence of the tone modulation was consistently signaled by a *lower* value of the DV). In general, the sign of the slope of the function indicated whether the signal interval corresponded to the higher or lower value of the DV.

Three representative masker conditions were analyzed with DVRP functions: (1) fixed-level masker (–18 dB rms), (2) roving-level masker (chosen from a uniform distribution from –23 dB to –13 dB rms), and (3) EEE masker (masker-alone depth of –18 dB rms). To generate each function, responses were combined across the four listeners and four tracks per condition. This resulted in approximately 80 observations per point in each function [16 tracks x ~50 trials per track / 10 points (bins) per function]. The spacing of consecutive points was not fixed; instead, a fixed number of responses were placed into unevenly spaced bins. There was no internal noise introduced to construct the DVRP functions; external stimulus variability was the only random factor included.

2.3.2 Results and Discussion

2.3.2.1 Fixed-level modulation masker

Predictions for each of the decision statistics along with the average listener thresholds for the fixed-level masker condition are shown in Fig. 2-5. The solid horizontal line in each panel indicates the listeners' mean threshold with no masker. Dashed lines without symbols in Fig. 2-5 show corresponding unmasked detection performance for each tested cue, which was determined in the simulations by a fixed-variance noise added to the DVs.

First, consider the simulated thresholds in the left column of Fig. 2-5. These DVs (envelope rms, average local modulation depth, and IC cell average rate) make up the subset of the 'long-term' statistics that predict performance broadly consistent with that of the listeners. However, each of them has shortcomings, even in this straightforward fixed-masker-level task (in which no detection cues have been manipulated). Average local depth and rms thresholds were almost identical (this is also true for the roving-level and equal-energy conditions), and suffered the same inconsistencies with the data. Specifically, the external variability of the stimulus caused increases in threshold at masker depths considerably lower than in the data. As a result, predicted thresholds based on envelope rms and average local depth were higher than the listeners' thresholds in the observed 'dip' around -30 dB rms. Because the slope of the function was lower in

the rms and average depth predictions than it was in the data, the curves re-converged at the highest masker levels tested.

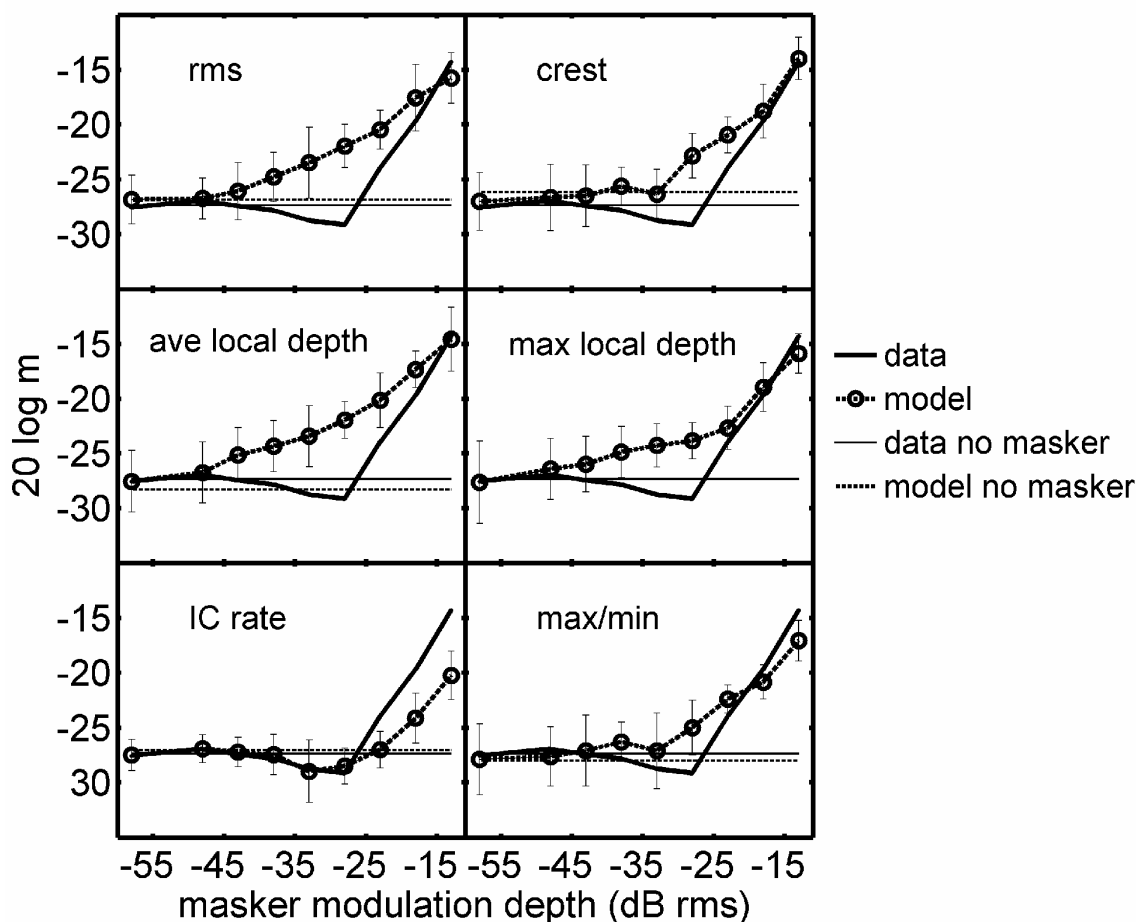


FIG 2-5. Model thresholds from simulated tracks based on six DVs (o), along with the listeners' fixed-level data from Fig. 2-2 (solid lines, no symbols). Unmasked (pure) SAM detection thresholds are shown with the solid horizontal lines. Unmasked model thresholds (thick dashed lines) were set by adding a fixed-variance internal noise to the DV in each interval. Predictions based on long-term DVs are shown in the left column; local temporal features were included in the simulations summarized in the right column.

Thresholds based on the average discharge rate of a model IC cell are shown in the bottom row of the left column in Fig. 2-5. This threshold-masker level function distinguishes itself from any of the signal-based DV predictions in two important respects. First, the IC model correctly predicted a non-monotonic dependence of signal sensitivity on masker modulation depth. Second, simulated thresholds were *lower* than the data at masker depths above -23 dB. The non-monotonic shape was a direct result of the 'hard' modulation-depth threshold that arises from the strong inhibitory inputs present in the model cell. In the absence of any internal noise or external signal variability, the model IC cell did not respond until the signal modulation depth was above ~ -32 dB. This value was set by the strength of inhibition relative to the excitation: stronger

inhibition results in higher thresholds. When an appropriate amount of noise was added to the sinusoidal signal (i.e. at a masker modulation depth of -28 dB rms), the instantaneous modulation depth rose above this hard threshold more often than it did with an equal-amplitude signal in ‘quiet’, resulting in lower detection thresholds and a pronounced ‘dip’ in the threshold-masker level function. At higher masker modulation depths (> -23 dB rms), the external variability of the masker swamped out this subtle effect. The fact that the absolute values of thresholds at these higher masker levels were lower in the simulations than in the data does not represent a fatal flaw. In fact, the inclusion of a Weber-fraction-type noise (i.e. one that is proportional to the stimulus or its response, see Ewert and Dau, 2004) would improve the match between model and data at these masker depths. The fit between other DVs and the data would not be improved by including this multiplicative type of internal noise because they predicted thresholds higher than the data, even with ‘fixed-variance’ internal noise alone.

Local temporal features were incorporated in the simulated thresholds shown in the right column of Fig. 2-5. In general, the fit to the data was slightly better for these DVs than the other signal-based statistics: the overestimation of thresholds in the data’s shallow notch was less severe, especially for the crest factor and max/min DVs. In a sense, it was surprising that predictions based on only two points of the modulating waveform (i.e. the max/min statistic) were more consistent with the listeners’ performance than traditional long-term (rms) measures. Ewert and Dau (2002) showed that a pure long-term cue could account for several trends in the frequency effects of modulation masking, using a masker modulation depth of -10 dB rms. Interestingly, at comparable masker levels in the present study, long-term and local feature cues were all reasonable predictors of modulation detection thresholds (i.e. model thresholds were either similar to or lower than the data). From the thresholds measured here across the range of low masker depths, it seems fair to conclude that predicted performance based on DVs that incorporated temporally-local features were as consistent with the listeners’ performance as rms (time-averaged) cues. Implementation of the physiological model suggested a mechanism (namely, a modulation-depth threshold nonlinearity) that could be incorporated into models for envelope processing to account for the non-monotonicity observed in the listeners’ threshold-masker level functions.

2.3.2.2 Roving-level modulation masker

When the masker modulation depth was randomly chosen from a 10-dB uniformly-distributed range of values centered at -18 dB rms on a trial-by-trial basis, mean psychophysical thresholds increased by about 4 dB over the -18 dB fixed-level condition (Fig. 2-3). We can ask the simple question: how much are thresholds based on these different DVs affected by the same manipulation? Figure 2-6 answers this question by comparing the fixed-level and roving-level thresholds for each DV (the listeners’ mean data are re-plotted from Fig. 2-3 at the far left). The simulated thresholds for fixed-level conditions in Fig. 2-6 are identical to those illustrated in Fig. 2-5 at a masker modulation depth of -18 dB rms, but re-plotted as a SNR for direct comparison to the roving-level condition, where the SNR was the tracking variable. Simulations using long-term DVs are represented in the first three columns to the right of the mean data. Thresholds with various combinations of local features included are shown in the last three pairs of bars.

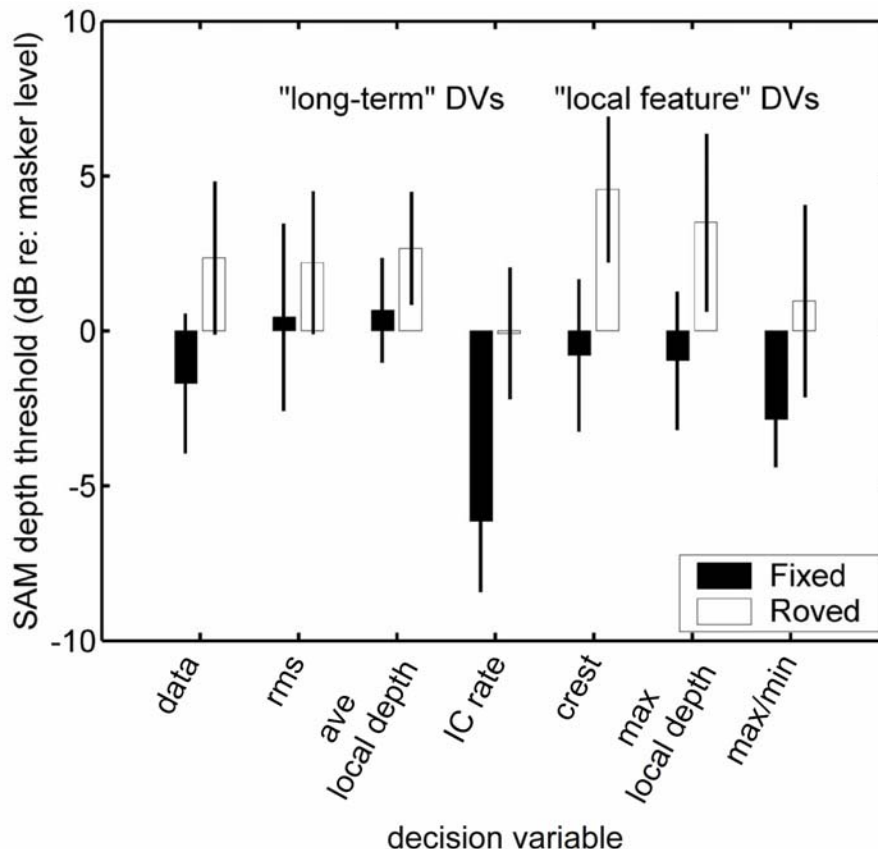


FIG 2-6. Roving-level masker simulations and comparison with listeners' mean data. Format is the same as Fig. 2-3, but the thresholds of different DVs are shown instead of the performance of different listeners.

Two aspects of the simulations are worth noting. First, all six of the tested DVs were affected by an amount that was consistent with the effect seen in the data: thresholds were increased by 2-6 dB when the masker level was roved. Also, thresholds based on the model IC cell's average discharge rate were lower than those obtained with the signal-based statistics. Again, this is not a serious failure of the physiological model. The inclusion of a second noise source, proportional to the magnitude of the stimulus or response fluctuations, would increase thresholds at these (relatively high) masker modulation depths, while maintaining the difference between the fixed-level and roving-level conditions. The other DVs would not benefit from such a modification (with the possible exception of the max/min statistic), as their predicted thresholds were at or above the actual data with a fixed-variance noise source alone.

Despite the subtle differences between the effects predicted by the different DVs, the variability of the threshold estimates with respect to the small observed threshold elevation does not allow for strong conclusions supporting or disputing the validity of a specific DV in the roving-level task. The use of a larger rove range would have potentially produced more pronounced effects that could have critically tested the different DVs; unfortunately, the limited dynamic range that was available in the amplitude-modulation domain for this manipulation did not allow for definitive answers

to these questions. In our paradigm, only masker modulation depths that clearly caused masking while also avoiding over-modulation were used.

2.3.2.3 Equalized-envelope-energy modulation masker

Predictions based on each of the tested DVs for the EEE conditions are compared to listeners' thresholds in Fig. 2-7 (dashed lines). The fixed-level thresholds are also replotted (solid lines) to provide a baseline for comparison to the EEE thresholds. Actual data are shown in each panel without symbols or error bars; DV predictions are shown with symbols and error bars (\diamond = EEE; \circ = fixed-level). By definition, rms and average-local-depth metrics were unable to track on thresholds in the EEE conditions (the signal-plus-noise modulation depth was adjusted to have the same long-term rms depth as the corresponding noise-alone interval).

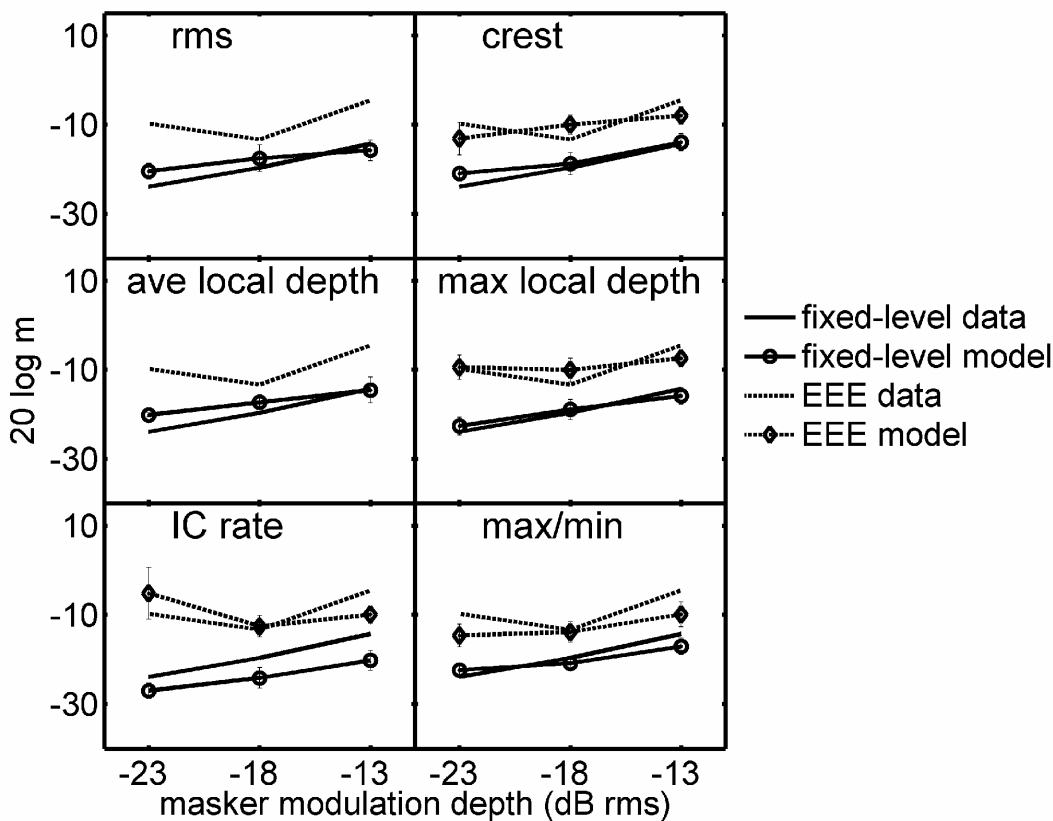


FIG 2-7. Comparison of measured thresholds (lines with no symbols) and predicted thresholds (connected symbols with error bars) for EEE (dashed lines) and fixed-level conditions (solid lines). DVs are arranged as in Fig. 2-5. If a given combination of DV and masker condition resulted in over-modulation or tracks that did not converge, simulated thresholds were not plotted (this occurred with the rms and average local depth DVs in EEE conditions).

The only long-term DV that was able to consistently track on a reasonable signal level at threshold was the discharge rate of a model IC cell. The predictions of the physiological IC model were comparable to the listeners' thresholds at masker depths of -18 and -13 dB rms. However, the IC model predicted that the difference between fixed-level and EEE thresholds should decrease with masker depth. This was not observed in

the data. The fact that the IC rate DV could predict EEE thresholds at all was a result of the effective band-pass envelope filtering that preceded the decision device. Although the bandwidth of the stimuli was within the pass-band of the filter (the half-rate Q-value is about 1, corresponding to a 64-Hz pass-band for the cell tuned to the signal modulation frequency), the resulting output spectra were nevertheless shaped by the cell's modulation-tuning properties. There was less attenuation of the energy concentrated near the peak of the modulation filter, so target-interval stimuli in the EEE conditions could elicit a larger response (higher discharge rate) than the standard-interval stimuli. A similar effect would be expected for an rms DV following any realistic envelope band-pass filtering process. Still, such a cue did not predict the appropriate variation in EEE thresholds with masker depth, and the absolute difference between fixed-level and EEE thresholds was higher than observed in the data. The use of an invariant cue across all conditions was also not consistent with listeners' anecdotal reports suggesting that their strategy was very different between the fixed-level and equal-energy conditions.

The DVs that were the best predictors of the listeners' EEE data were those based on local temporal envelope features (Fig. 2-7, right column). Max/min, crest factor, and max local depth all accounted reasonably well for the difference in performance between the fixed-level and EEE thresholds as well as the absolute values of the thresholds and the higher variability associated with the EEE data. Importantly, the decision rule had to be switched for these three statistics: simulations selected the interval with the larger value of the DV in fixed-level conditions and the interval with the smaller value in EEE conditions as the target interval. This sign-flipping was qualitatively consistent with subjective accounts from the listeners that they had developed a different strategy (based on feedback) in the EEE paradigm: often the 'smoother' or 'more regular' envelope was reportedly chosen as the signal interval. Given this, the success of the local feature statistics in predicting the data was somewhat surprising, since the calculations were all heavily weighted by a small temporal portion of the envelope waveform. Nevertheless, the match to the data was quite good, and it was clear that the most straightforward way to account for the listeners' performance in the EEE conditions was to incorporate information about local fluctuations into the decision device. The finding that local temporal features were crucial for explaining the EEE data was different than the conclusions drawn from analogous audio-frequency energy-equalized TIN detection tasks (i.e. Richards and Nekrich, 1993), in which the overall flattening of the (long-term) envelope when a tone was added could explain performance. The corollary cue in the current experiment would be a drop in the (long-term) envelope energy, which we determined to be incapable of predicting performance consistent with the listeners.

2.3.2.4 Decision-variable-reconstructed psychometric functions

As an alternative method to compare and contrast different DVs (beyond predicting thresholds), we analyzed trial-by-trial decisions made by the listeners and considered how those choices correlated with the magnitude and direction of variation in each DV between the two stimuli presented to the listener. Figure 2-8 shows DVRP functions for three masker conditions and six DVs. The masker modulation depth for all of the conditions shown was nominally -18 dB rms (note that this value was randomly chosen from a 10-dB range in the roving-level conditions, and was effectively attenuated in the target interval for the EEE conditions). Fixed-level results are contained in the first column; roving-level and EEE analyses are shown in the second and third columns,

respectively. The vertical dashed line in each panel indicates the point where the target and standard DVs were the same (i.e. their difference was zero). The ordinate limits are set at 0 and 100% correct in each panel. The x-axes are fixed for each DV; they were determined by the largest range of variation observed across the three conditions (usually the roving-level case; see figure caption for exact values).

DVRP functions for the fixed-level condition (left column) can be placed in one of two categories. The first three DVs (AC-coupled envelope rms, average local depth, and IC rate) all showed a consistent increase in percent correct as the target interval DV became larger than the DV measured in the corresponding standard interval. Also, these three cues were typically not ‘confused’ by the task: the vast majority of the trials resulted in a positive difference between target and standard DV.

The second group of statistics is made up of DVs based on local temporal features (crest factor, max local depth, and max/min ratio). Listeners’ percent correct increased for positive differences (target DV > standard DV) as they did for the long-term statistics in the top three rows, but a higher proportion of trials resulted in standard-interval DVs that were larger than the target-interval DVs (>20%; represented by points to the left of the vertical dashed line). This confusion suggested that long-term DVs may have been more reliable cues in the fixed-level masker conditions, because they were less susceptible to changes in local features caused entirely by the stochastic nature of the maskers.

Roving-level masker DVRP functions (Fig. 2-8, middle column) spanned a wider range of DV differences than fixed-level or EEE-masker conditions, as expected. Because the roving-level DVRP functions were qualitatively similar, none of the six DVs (Fig. 2-8, middle column) can be considered more or less consistent than any of the others in this paradigm. In about one-quarter of the trials, percent correct dropped below chance (50%) when the cue in the standard interval was bigger than that in the target interval for all six of the tested DVs. It would be possible for the functions to be symmetric about zero if the listeners recognized conditions where the level-roving caused such a reversal in cue direction. This was not seen in any of the DVRP functions; the listeners tended to choose the interval with the larger DV value, regardless of the particular random combination of standard and target masker level. The similarity of the DVRP functions for the roving-level results makes it difficult to point out one DV as being more consistent with the data. Similar conclusions were made with the roving-level threshold-tracking simulations (Fig. 2-6).

The target-interval rms modulation depth was normalized to match that of the standard interval in the EEE conditions. DVRP functions for a -18 dB EEE masker are shown in the right column of Fig. 2-8. Compared to the fixed- and roving-level cases, the spread of DV differences is highly compressed for rms, average local depth, and IC rate decision statistics (Fig. 2-8, top three panels, right column). This reflects the stimulus manipulation; the fact that there was any spread in the rms DVRP was because only the steady-state portion (the central 500 ms) of the stimulus was used to compute the DVs, while the entire duration of modulation was equalized in the stimuli presented to the listeners (including the onset and offset ramps). Still, there were no strong trends in the upper three EEE DVRP functions of Fig. 2-8: percent correct was nearly independent of the DV difference, which was always close to zero. The situation was different for the local-feature-based DVs (bottom-right three panels in Fig. 2-8). High signal modulation

depths resulted in low values of crest factor, max local depth, and max/min ratio. The tracking simulations (Fig. 2-7) and DVRP functions suggested that listeners were using a drop in the value of a DV that incorporated some local feature; max/min ratio predicted absolute thresholds that were slightly closer to the listeners' thresholds than crest factor or max local depth.

Analysis of DVRP functions provided a different angle on the same question that was addressed with the tracking simulations; the consistency between the two approaches is reassuring. The technique is promising for pulling apart decision statistics in other psychophysical tasks that include external stimulus variability, especially those with competing DVs that are weakly correlated with one another. The procedure requires no assumptions to be made about internal noise, and no additional time from the listeners. Analysis of adaptive tracking procedure responses has previously been validated as an efficient and accurate way to extract psychometric functions (Dai, 1995); the current implementation simply considered statistics based on the actual stimuli presented, instead of the signal level, or modulation depth, presented in each trial. In the context of the current simulations, two specific and important pieces of information are reinforced with the DVRP functions. First, they reiterate the notion that the sign or direction of the cue flips in EEE conditions for the local-feature DVs with respect to the direction of the cue in fixed- and roving-level masker conditions. Second, the proportion of DV calculations that elicit a larger value in the standard interval in the fixed-level conditions is higher for the short-term DVs than for the long-term DVs.

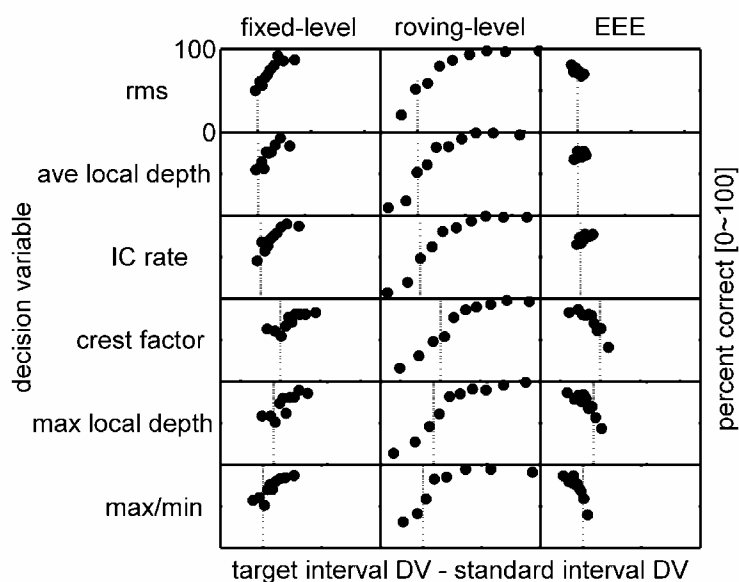


FIG 2-8. Decision-variable-reconstructed psychometric functions for six DVs and three key masker configurations: fixed-level (-18 dB masker depth; left column), roving-level (masker depths randomly chosen from a 10-dB range centered at -18 dB; middle column), and equal-envelope-energy (-18 dB standard-interval masker depth). The range of differences plotted for each DV was determined by the range covered in the roving-level conditions. The ranges were: envelope rms: $[-0.07 \ 0.23]$; average local depth: $[-0.004 \ 0.013]$; model IC cell rate: $[-6 \ 18]$; crest factor: $[-0.3 \ 0.5]$; maximum local depth: $[-0.01 \ 0.02]$; max/min ratio: $[-1.46 \ 4.00]$. Vertical dashed lines indicate the zero-difference point on the x-axis.

2.4 GENERAL DISCUSSION

The remainder of this article is divided into three parts. In the first section, potential mechanisms underlying specific features of the fixed-level results are further discussed. Directions for future work are detailed in the second section. Finally, the key psychophysical and modeling results are summarized.

2.4.1 Negative masking

The non-monotonic relationship between sensitivity and noise level apparent in two of the four listeners in the fixed-level masker condition (Fig. 2-2) can be interpreted as stochastic resonance (for a recent review, see Wiesenfeld and Jaramillo, 1998). There are (at least) two straightforward mechanisms that could underlie such an effect. One possible explanation is that the listeners used a non-optimal criterion that remained constant across noise level [see Tougaard (2000) for an analysis of such an assumption]. This interpretation is less than satisfying for several reasons. First, the presence of the non-monotonicity is not related to the listeners' pure AM-detection thresholds (with no masker). If the effect was simply an epiphenomenon of poor criterion placement, the two listeners whose data suggest stochastic resonance should have been less sensitive than the other listeners at low masker depths. Another problem with the poor-criterion explanation is related to the types of mistakes that such a mechanism would predict. In low-masker-level conditions, the fixed DV criterion is never reached, and as a result, the signal is never "perceived" as being present. The opposite is true for the high masker levels, where even the noise-alone DV distribution lies above the fixed criterion: the above explanation suggests that the signal should sound as though it is present on every trial. These bias-related observations are also inconsistent with subjective impressions given by the listeners, and they suggest that some other mechanism may underlie the stochastic resonance effects.

Another mechanism that can explain the non-monotonicity in our data is based on a combination of weak signals and a threshold nonlinearity (i.e. Ward *et al.*, 2002). If a system does not respond to a sub-threshold periodic stimulus, the addition of noise may push the input amplitude above threshold at a mean frequency related to the periodicity of the weak signal. An example of such a system with an envelope (modulation-depth) threshold is the physiological model tested here (Nelson and Carney, 2004). The ability of such a simple model to account for the effect highlights the potential advantages of using physiologically-motivated model front-ends when predicting psychophysics to gain insight into underlying mechanisms. In addition to modeling work, there is also direct physiological evidence suggesting that central auditory neurons respond in a way consistent with a modulation-depth threshold device. Adding a low-level noise modulation to a sinusoidal AM can both enhance neural synchronization to the tone and increase the average discharge rate over responses to the SAM tone alone in the frog auditory midbrain (Bibikov, 2002). The negative masking effects observed in the current study are also likely related to similar psychophysical measures in cochlear implant listeners of masked (electrically stimulated) modulation detection thresholds (Chatterjee and Robert, 2001).

2.4.2 Future directions

A main focus of future work will be to quantitatively relate actual (as opposed to modeled) midbrain physiological responses to psychophysical performance in AM detection tasks. To date, the relative roles of timing (i.e. synchronization to the envelope) and average rate information as neural substrates for AM perception at low modulation depths (near behavioral thresholds) remain unclear. The rate versus timing debate can be thought of as a discussion of underlying neural DVs, similar to the classifications of signal-based DVs as long-term or local-feature-containing. It is typically assumed that information about AM is largely transformed into an average-rate-based scheme by the level of the IC (which is one reason we only considered the rate responses of our model IC cell here), but the majority of data supporting that view comes from stimuli with high modulation depths (for a review, see Joris *et al.* 2004). The fact that our listeners could perform the EEE task suggests that the local temporal structure of AM stimuli is available as a cue under certain conditions. To reconcile these inconsistencies, we are currently recording responses in the awake rabbit IC to both pure SAM and noise-masked SAM across a wide range of modulation depths (from -35 dB to 0 dB in $20 \log m$).

Another issue that deserves further study is the effect of including a “Weber-fraction noise,” along with the fixed-variance internal noise that was used here to limit performance with deterministic stimuli. Existing data suggest that tone-carrier AM-depth discrimination sensitivities may be determined by a fixed-variance noise at low modulation depths and a noise that is proportional to the elicited response at high modulation depths (i.e. Ewert and Dau, 2004). Assuming that the listeners were using an overall depth-related cue, then the fixed-level masker SAM detection paradigm can be thought of as depth discrimination task, with both external and internal noise processes playing a role. At the highest masker depths tested (-13 dB rms), most of the DV-derived thresholds are at or below the listeners’ data (Fig. 2-5), suggesting the need for an additional source of noise at high modulation depths. This is consistent with the findings from the AM-depth discrimination literature. To better account for all of the data presented here, it seems necessary to implement a model with a modulation depth threshold, along with some form of local feature detection and two types of internal noise (fixed-variance and Weber-fraction).

2.4.3 Summary

- (i) SAM depth thresholds in an on-frequency masked AM-detection task were influenced by external stimulus variability at very low masker modulation depths (i.e. -40 to -30 dB rms). Negative masking, or stochastic resonance, was observed in two of the four listeners at masker levels around -30 dB rms (Fig. 2-2).
- (ii) Roving the overall modulation depth (Fig. 2-3) or equalizing the long-term envelope energies (Fig. 2-4) from trial to trial both resulted in significant increases in threshold. These findings contrast with observations in comparable TIN detection tasks in the audio-frequency domain.
- (iii) Tracking simulations showed that several competing DVs were able to qualitatively account for performance for the fixed-level (baseline) and roving-level masker conditions (Fig. 2-5).
- (iv) Reconstruction of psychometric functions based on a variety of DVs revealed that long-term statistics (averaged across the entire stimulus duration) may have been more

robust cues in the fixed-level condition than statistics based on local temporal features. This was inferred because of the larger proportion of trials that resulted in the standard interval DV being larger than the corresponding target interval DV when local features were assumed to be the primary detection cues (Fig. 2-8).

(v) Thresholds in the EEE conditions could only be accounted for with a “local feature” DV, as long-term cues were minimized by equalizing the overall energy of the standard and target envelopes, after the sinusoidal signal was added. Listeners apparently chose the interval with a lower max/min ratio, crest factor, maximum local depth, or some other local feature cue in these conditions (Fig. 2-7).

(vi) Implementing a physiologically motivated model structure and comparing predictions based on its rate responses to the fixed-level data showed that a hard modulation-depth threshold mechanism can predict negative masking at low masker depths. This suggests that such a nonlinearity could be included (along with an internal noise source) to limit performance in the absence of external variability in a more complete model of envelope processing (Fig. 2-5).

ACKNOWLEDGMENTS

This research was supported by NIH-NIDCD F31-7268 (PCN) and NIH-NIDCD R01-01641 (LHC, PCN). We thank Magdalena Wojtczak, Karen Doherty, Yan Gai, and the reviewers for helpful comments on earlier versions of the manuscript.

CHAPTER 3

Comparison of intensity discrimination, increment detection, and comodulation masking release in the audio- and envelope-frequency domains

ABSTRACT

In general, the temporal structure of stimuli must be considered to account for certain observations made in the audio-frequency domain. Two such phenomena are (1) a heightened sensitivity to amplitude increments with a temporal fringe compared to gated intensity discrimination performance, and (2) lower tone-in-noise detection thresholds using a modulated masker compared to those using an unmodulated masker. One qualitative interpretation of these findings is that temporal envelope cues provide an additional dimension which listeners can monitor, alongside cues in the spectral dimension. In the current study, translations of these two experiments were carried out to test the hypothesis that analogous cues might be used in the envelope-frequency domain. Pure-tone carrier amplitude-modulation (AM) depth-discrimination thresholds were measured using both traditional gated stimuli and using a temporally modulated fringe for a fixed standard depth ($m_s = 0.25$) and a range of AM frequencies (4-64 Hz). In the other experiment, masked sinusoidal AM detection thresholds were compared in conditions with and without slow and regular fluctuations imposed on the instantaneous masker depth. In contrast to the audio-frequency results, in the modulation-frequency domain thresholds in both tasks were largely independent of the stimulus temporal structure. A physiologically motivated model that effectively acts as a first-order envelope change detector accounted for several, but not all, of the key aspects of the data. *Note: A version of this chapter will be submitted to the J. Acoust. Soc. Am. as a paper with the same title, by P. C. Nelson, S. D. Ewert, L. H. Carney, and T. Dau.*

3.1 INTRODUCTION

A variety of fundamental experimental paradigms originally used in audio-frequency psychoacoustics have recently been translated into their envelope-frequency equivalents. In the process, certain parallels have emerged between the effective signal processing that is inferred to take place in the two domains. Masked tone-detection experiments that compare the amount of masking with different spectral relationships between the signal and masker indicate perceptual frequency tuning in both audio frequency (e.g. Wegel and Lane, 1924) and envelope frequency (e.g. Houtgast, 1989). Also, the “asymmetry of masking” has been observed in both domains [e.g. Moore *et al.*, 1998 (audio frequency); Ewert *et al.*, 2002 (envelope frequency)], with tones proving to be relatively ineffective maskers of noise signals. Wojtczak and Viemeister (2005) showed non-simultaneous (forward) masking in the envelope-frequency domain that has direct counterparts in audio-frequency psychophysics (e.g. Luscher and Zwislocki, 1949). Tonal intensity resolution performance in both domains is also broadly similar, with an approximately 1 - 2 dB increase in the standard level (SPL in audio frequency, 20 log m

in envelope frequency) required to reliably discriminate the levels of two supra-threshold sounds [i.e. Florentine *et al.*, 1987 (audio); Ewert and Dau, 2004 (envelope)]. These qualitative similarities suggest that the processes underlying perception in the two domains may be conceptualized in a single framework, despite the fact that the underlying neural mechanisms are presumably quite different.

In the current study, we tested the hypothesis that two other robust audio-frequency phenomena would be observed in the envelope-frequency domain. These phenomena are (1) a heightened sensitivity to increments with a continuous carrier (or a temporal fringe) relative to gated intensity discrimination performance, and (2) lower thresholds in a tone-in-noise detection task with a temporally amplitude-modulated (AM) masker than in conditions with a random (unmodulated) masker. The second observation has been termed comodulation masking release (CMR) because it is most robust when several audio-frequency channels are simultaneously and coherently modulated. Both spectral-frequency observations can be at least partially attributed to AM-related cues (e.g. Gallun and Hafter, 2006; Schooneveldt and Moore, 1989). Therefore, upon transposition into the envelope-frequency domain, the analogous cues in such tasks would be related to the 2nd-order envelope, or ‘venelope’ (Ewert *et al.* 2002). The results presented here point out some limitations of describing the fundamental processes in the two domains within a common framework.

Listeners are more sensitive to audio-frequency intensity differences presented as continuous-carrier level increments than presented as gated tones with different SPLs (e.g. Campbell and Lasky, 1967; Viemeister and Bacon, 1988; Bacon & Viemeister, 1994). An energy-based detection model cannot explain the difference in thresholds in the two conditions. Instead, the temporal structure of the standard, target, and inter-stimulus intervals must be taken into account. This finding can be considered from several perspectives. One possibility is that the memory requirements of the system are higher in gated-carrier intensity discrimination than for increment detection, where listeners could potentially perform the task without comparing across intervals (Harris, 1963). This explanation is less than satisfactory, because near threshold, there is certainly an element of comparison across intervals in the continuous-carrier task: the listener must decide which interval sounded *the most* like it contained an increment or ‘bump’. Also, the relatively short intervals between stimuli probably render memory noise negligible with respect to sensation noise in most two- or three-interval paradigms (Durlach and Braida, 1969).

Another related explanation holds that the improved sensitivity results because the system could be making decisions by detecting changes in the increment task (Macmillan, 1971; Hafter *et al.*, 1997). The phrase ‘change detector’ is rather nebulous; one specific difference that arises between gated and continuous paradigms is in the envelope spectra of the stimuli. Onsets and offsets of gated stimuli result in excitation of the putative modulation filterbank (i.e. Dau *et al.*, 1997) that depends on the shape and duration of the ramps applied to the carrier (Gallun and Hafter, 2006). The presence or absence of a small difference in the modulation properties between two gated stimuli might be masked by neural onset adaptation (Irwin & Purdy, 1982). Increments to a continuous pedestal, on the other hand, yield a more well-defined change in the envelope of the signal interval, without the complications that arise when the signal and pedestal are gated simultaneously. Physiologically, absolute firing-rate changes in single

auditory-nerve fiber (ANF) responses to increases in intensity do not depend on the temporal position of the increment (e.g. Smith and Zwislocki, 1975; Smith and Brachman, 1982; Smith *et al.*, 1985). Instead, the *relative* increase in instantaneous rate increases as the delay between pedestal onset and increment is lengthened, because the response to the pedestal decreases with time. Short-term peripheral neural adaptation is essentially complete after about 100 ms (e.g. Smith, 1979); psychophysical fringe effects in increment detection are delay-dependent over time courses of several hundred milliseconds (e.g. Green, 1969; Leshowitz and Cudahy, 1972; Bacon and Viemeister, 1985).

Tone-in-noise detection tasks that compare masking by modulated and unmodulated maskers have emerged more recently as challenges to pure long-term power-spectrum models of audio-frequency masking (e.g. Schooneveldt and Moore, 1989; Verhey *et al.*, 1999). For the purposes of this article, we will focus on a simple class of CMR experiments: those that use one band-limited noise, centered on the signal frequency. This class of CMR paradigms usually yields the most significant and robust release from masking. At least three cues could potentially underlie a release from masking (i.e. lower thresholds with a modulated masker compared to unmodulated masker conditions). A ‘dip-listening’ model suggests that the listeners are able to selectively attend to the periods of the masker with low modulation depths, thus improving the local signal-to-noise ratio (SNR; Buus, 1985). Another possible within-channel cue is the overall smoothing of the masker fluctuations upon addition of the signal (Schooneveldt and Moore, 1989; Verhey *et al.*, 1999). Across-channel comparisons of target-interval differences might also be used if the bandwidth of the masker is sufficiently broad (Hall *et al.*, 1984). All of these mechanisms have been used to understand CMR in the audio-frequency domain.

With these spectral-frequency empirical observations in mind, we present here envelope-frequency-domain versions of the experiments that led to them. The perceptual salience of envelope cues will determine whether differential effects will be observed in (1) continuous- and gated-carrier AM depth discrimination and (2) sinusoidal AM detection in the presence of a noise modulation masker with and without slow and regular fluctuations in overall depth. Our current understanding of envelope perception can be summarized as follows: Sinusoidal modulation of the depth of a first-order AM carrier is detectable (e.g. Lorenzi *et al.* 2001), but the perceptual salience of envelope components is generally found to be weaker than that of first-order envelope fluctuations (Ewert *et al.* 2002). Envelope fluctuations can interact with envelope detection and vice versa (e.g. Ewert *et al.* 2002). For example, the detection of a 5-Hz modulation can be affected by the presence of a masker modulation tone complex consisting of components with 5 Hz spacing (e.g., 45 and 50 Hz) in a phase-sensitive way; the envelope of the masker (at 5 Hz) thus interacts with the signal modulation (Moore *et al.* 1999). In terms of qualitative underlying mechanisms, envelope energy is not present in the envelope spectrum, but several physiologically realistic nonlinearities could transform 2nd-order energy into 1st-order envelope features (Shofner *et al.*, 1996). Alternatively, envelope cues could take the form of temporal variations in the output of envelope-frequency-tuned modulation filters (i.e. Ewert *et al.*, 2002; Füllgrabe *et al.*, 2005; Füllgrabe and Lorenzi, 2005).

The remainder of this article is divided into three main sections. Two lines of psychophysical experiments are described and discussed in the first two sections. The

third part focuses on interpretation of the findings with the help of a physiologically motivated computational model.

3.2 EXPERIMENT I. INTENSITY DISCRIMINATION AND INCREMENT DETECTION IN THE ENVELOPE-FREQUENCY DOMAIN

The goal of the first set of experiments was to determine whether continuous-carrier AM-depth-discrimination thresholds were lower than traditional gated-carrier thresholds. Two reasonable hypotheses lead to predictions of a difference in performance between the two paradigms. First, adaptation at the output of modulation-tuned channels could mask across-interval depth differences, resulting in poorer performance in the gated conditions. Alternatively, because energy increment detection in the audio-frequency domain is at least partly associated with modulation detection and coded along the modulation dimension (Gallun and Hafter, 2006), a corresponding task in the envelope domain may provide another cue along an additional dimension, the envelope dimension, which might lead to lower thresholds in the continuous-carrier condition relative to the gated case.

3.2.1 Methods

3.2.1.1 Listeners

The experiments were carried out at the Centre for Applied Hearing Research at the Technical University of Denmark (DTU). All of the listeners participated voluntarily and had pure-tone detection thresholds less than 20 dB HL at octave frequencies between 125 and 8000 Hz. Their ages ranged from 23 to 39 years. Three of the four subjects in the main experiments had significant experience in related psychoacoustic testing; two of the authors (PCN, TD) were part of this group. Four additional listeners were recruited to participate in ‘auxiliary experiments’ because of the relatively large across-subject variability.

3.2.1.2 Apparatus and stimuli

Subjects listened diotically via Sennheiser HD 580 circumaural headphones in a double-walled sound-attenuating booth. Stimulus generation and presentation were carried out in MATLAB using the AFC software package developed at the University of Oldenburg and at DTU. A 48-kHz sampling rate was used to digitally generate stimuli. The carrier was a 70-dB SPL, 5.5-kHz pure tone. Sinusoidal AM was applied for the entire 500-ms duration, and a 50-ms Hanning window was applied at the onset and offset of observation-interval stimuli. Inter-observation-interval durations (between possible target interval presentations) were also 500 ms in duration.

The modulating waveforms in the gated conditions were identical to those described in Ewert and Dau (2004). The observation-interval stimuli are described by the following equation:

$$s(t) = \sin(2\pi f_c t) [1 + m_s \sqrt{1 + m_{inc}} \sin(2\pi f_m t)],$$

where f_c is the carrier frequency (5500 Hz), m_s is the standard modulation depth, m_{inc} is the relative depth increment (zero in the standard intervals), and f_m is the modulation frequency. The comparison (target interval) depth can be related to the standard depth and depth increment simply as $m_c = m_s \sqrt{1 + m_{inc}}$. Using notation more in line with audio-frequency intensity discrimination literature, m_{inc} can also be thought of as the Weber

fraction, i.e. $m_{inc} = (m_c^2 - m_s^2) / m_s^2$. Whereas the earlier study (Ewert and Dau, 2004) focused on the effects of standard-interval modulation depth for a fixed-frequency (16-Hz) sinusoidal AM, the current experiments used a fixed standard depth (m_s) of -12 dB (in $20 \log m$; linear $m = 0.25$), and varied two other parameters. Here, the influences of modulation frequency ($f_m = 4, 8, 16, 32, \text{ and } 64 \text{ Hz}$) and gating choices were investigated. We will refer to the traditional AM depth-discrimination stimuli (i.e. Wakefield and Viemeister, 1990; Lee and Bacon, 1997; Ewert and Dau, 2004) as “gated” and the envelope-domain equivalent of increment detection as “quasi-continuous” or “fringe” conditions.

The critical difference between the gated and fringe conditions was confined to the stimulus presented between observation intervals in the three-interval paradigm. In gated conditions, a silent interval separated three modulated tones (the two standard intervals contained tones with a modulation depth of 0.25; the target interval was a tone with some AM depth higher than 0.25). In contrast, the quasi-continuous conditions were comprised of a 500-ms, 25% modulated tone that was present in the two inter-observation intervals and also before the first interval and after the third and final interval. Example stimulus waveforms for the gated and fringe conditions are shown in Fig. 3-1C. Stimulus amplitudes in all three intervals were gated with 50-ms ramps, regardless of the gating mode (this was also the case for the audio-frequency level discrimination stimuli, which are described next).

For comparison, thresholds for the audio-frequency versions of pure-tone gated intensity discrimination and quasi-continuous increment detection (i.e. with a temporal fringe), were also measured in the same subjects. Signal duration, inter-observation-interval, gating parameters, SPL, and carrier frequency were the same as in the AM-frequency experiments. The corresponding example stimulus waveforms are shown in Fig. 3-1A.

3.2.1.3 Procedure

Listeners were trained until four consecutive threshold estimates in each condition showed no evidence of learning. Two additional threshold estimates were obtained if the standard deviation of the four estimates was greater than 3 dB (this happened once in all of the experiments described here). Average data are presented as the mean and standard deviation of the subjects’ final depth-discrimination threshold estimates.

A three-interval, three-alternative forced-choice paradigm with visual correct-answer feedback was used along with a two-down, one-up adaptive tracking procedure (Levitt, 1971). This combination of parameters yields convergence on the 70.7% point of the psychometric function and a threshold estimate that corresponds to a d' of unity. The listeners’ task was to identify the observation interval containing the higher signal AM depth. Observation-interval timing was identified with visual cues presented synchronously with the standard and target interval stimuli on the computer monitor. The stimulus parameter that was varied in the tracking procedure was the fractional AM depth increment in dB ($10 \log m_{inc}$). The initial 4-dB signal-interval step size was halved after each of the first two track reversals occurring after consecutive correct responses. Six reversals were required after the final 1-dB step size was reached; threshold for a given track was taken as the mean signal level corresponding to the target-interval AM depth used at those six points. The order of stimulus presentation was randomized across parameters (gating mode and AM frequency) for each listener.

The audio-frequency intensity-discrimination experimental procedures were essentially identical to those used to measure AM depth-discrimination sensitivity. The tracking variable used was also similar. The Weber fraction in dB ($10 \log \Delta I/I$) was adjusted until the target interval was just noticeably different from the two standard observation intervals.

3.2.2 Results

3.2.2.1 Discrimination thresholds with gated and fringe presentation modes

The magnitude of the audio-frequency gated-continuous difference was measured first; mean intensity discrimination results are shown in Fig. 3-1B. Enhanced sensitivity to increment (fringe) conditions has been demonstrated in previous studies; the average difference for the listeners in the current study was 4-5 dB. The magnitude of the effect in our listeners is in line with the average 4.6 dB difference found at SPLs above 35 dB by Viemeister and Bacon (1988), who used a continuous 1000-Hz carrier and 200-ms observation intervals. Absolute discrimination thresholds in the gated conditions in the current study ($10 \log \Delta I/I = -6.5$ dB) are slightly better than the thresholds for matching carrier frequencies and standard levels reported by Florentine *et al.* (1987); this may be attributable to differences in presentation mode (monaural vs. diotic).

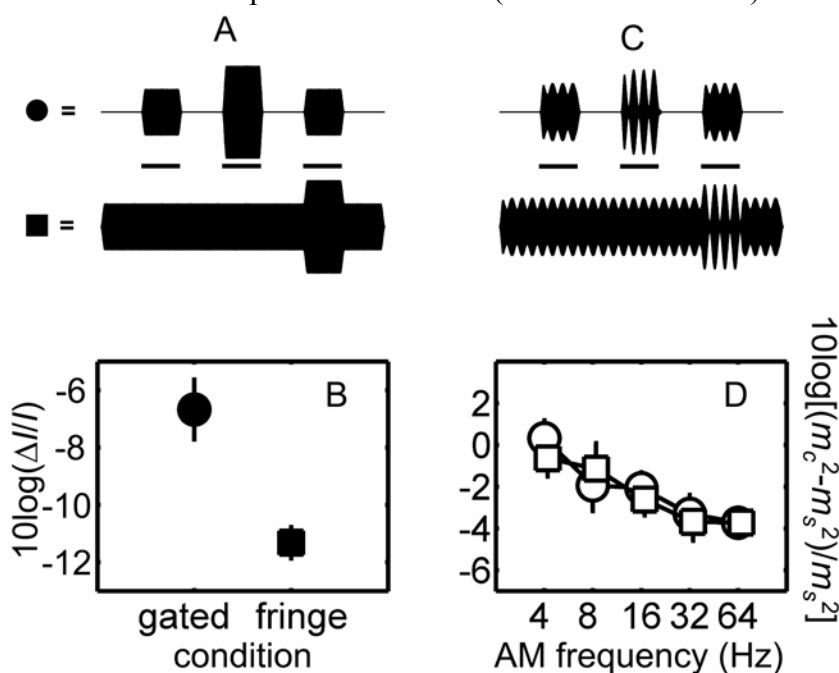


FIG 3-1. Comparison of the gated-continuous difference in the audio (left) and envelope (right) frequency domains. Schematic illustrations of stimulus waveforms in the two experiments are shown in A and C (horizontal bars between the stimuli indicate the timing of the 500-ms observation intervals). B: Audio-frequency intensity discrimination thresholds measured with gated (closed circles) and quasi-continuous pedestals (closed squares). D: AM depth discrimination thresholds for a 25% modulated standard obtained with traditional gated intervals (open circles), and with quasi-continuous modulation presented before, between, and after the observation intervals (open squares). For both B and D: $f_c = 5500$ Hz; standard SPL = 70 dB. Each symbol is the average threshold for 4 listeners; error bars indicate standard deviations of the individual mean thresholds.

Average depth discrimination thresholds are shown in Fig. 3-1D for a range of modulation frequencies. The most relevant aspect of the data for the purposes of the current study is the similarity in performance for the gated and fringe conditions, which is in contrast to the findings in the audio-frequency domain. Performance was broadly consistent across listeners, as suggested by the size of the standard deviation bars (<1.6 dB). Listener L4 was slightly more sensitive in the fringe conditions, while L2 exhibited lower thresholds in the gated conditions. Because these individual differences were similar in magnitude and stable across AM frequency for both listeners, they effectively cancelled out in the mean data.

Mean absolute thresholds [in $10 \log ((m_c^2 - m_s^2) / m_s^2)$] dropped from approximately 0 dB at 4 Hz to -4 dB at 32 and 64 Hz. This is equivalent to target-interval depths at threshold ranging from 0.35 to 0.30 in $20 \log m$ for a 25% modulated standard, and is consistent with previous studies that have found decreases in threshold at higher AM frequencies with a fixed-duration stimulus (i.e. Lee and Bacon, 1997). Thresholds in the gated condition at 16 Hz (-2.1 dB) were within 1dB of those reported by Ewert and Dau (2004), who used a 16-Hz signal with a standard depth of 22.5% (among others) imposed on a 65-dB SPL, 4-kHz carrier.

3.2.2.2 Gated and fringe AM detection thresholds and comparison with ‘static’ intensity discrimination performance

Previous studies have reported an enhancement of SAM detection thresholds at low modulation rates ($\leq \sim 10$ Hz) when a temporal fringe was used instead of gating the carriers (Viemeister, 1979; Yost & Sheft, 1997). Our finding of identical discrimination thresholds in the first experiment (with $m_s = 0.25$) appears incompatible with these earlier findings, since detection is a special case of discrimination ($m_s = 0$ for pure detection). We wanted to determine whether the listeners in the current study also exhibited a gated-continuous difference for AM detection. In this extension of the first experiment, thresholds in several related envelope-processing tasks were directly compared in an effort to better map out the differences between AM discrimination and detection, and to establish relationships between “dynamic” AM cues and “static” audio-frequency intensity discrimination. How are gated and continuous level discrimination and AM detection thresholds related? Based on the results of the first experiment, one hypothesis is that AM detection should *not* depend on the choice of gating parameters if the same mechanism underlies AM detection and AM depth discrimination.

To test this hypothesis, the dynamic AM stimuli were matched to the static level-increment stimuli in their carrier frequency (5.5 kHz), standard level (70 dB SPL), overall duration (500 ms), and onset / offset ramp duration (50 ms). Restricting the ramp duration parameter to be the same resulted in an AM stimulus that contained 5 cycles of 10-Hz SAM signal (50 ms on + 50 ms off = 100 ms period). Different combinations of gated vs. fringe presentation modes and static vs. dynamic fluctuations in the target interval resulted in the five test stimuli used in the second experiment. Schematics of the envelope waveforms in each condition are shown in Fig. 3-2B. Conditions with temporal fringes were identified as INC (static increment), AM1 (dynamic AM increment, with a DC offset), and AM2 (dynamic AM increment, no DC offset). Gated stimuli were labeled ID (static gated intensity discrimination) and AM (gated AM detection).

Mean data across 8 listeners are shown in Fig. 3-2A. Performance is defined in terms of a Weber fraction, where ΔI is determined by the difference between the

maximum and minimum value of the envelope in the target interval for the AM conditions (open symbols), and by the difference in peak intensities across the standard and target intervals in the audio-frequency level-discrimination conditions (closed symbols). In Fig. 3-2B, a two-interval task is represented for clarity (a 3AFC task was used in the actual experiment), with the target interval onset beginning at 0.5 s and the standard interval beginning at 1.5 s.

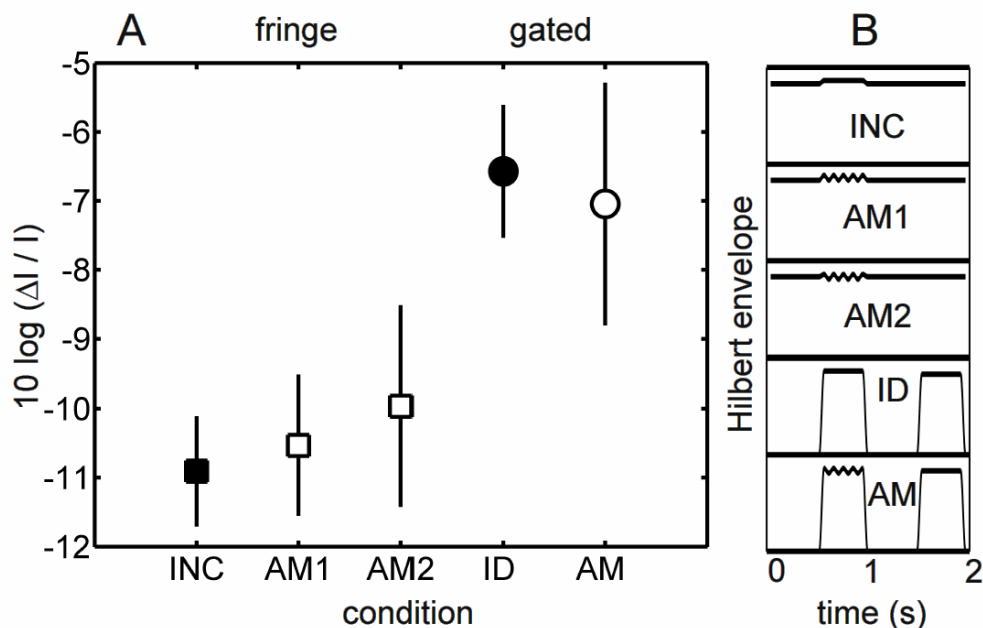


FIG 3-2. A: Mean audio-frequency level discrimination thresholds (solid symbols) and envelope-frequency detection thresholds (open symbols) under different gating conditions. Squares represent performance with a quasi-continuous carrier; circles depict thresholds with gated carriers. Conditions AM1 and AM2 are distinguishable based on the presence (AM1) or absence (AM2) of a DC component in the target interval. B: Schematic illustrations of the stimulus envelopes used in each condition. Conditions INC and ID are re-plotted from Fig. 3-1B.

One main result is that thresholds were similar for all three fringe conditions [Fig. 3-2(A)], and for both gated conditions, when performance was defined as $10 \log (\Delta I / I)$. This result suggested that the system was not more sensitive when several dynamic temporal envelope fluctuations were presented than when a fixed energy increment with a single onset and offset was used. Apparently the contribution of energy integration over 500 ms in the static conditions was approximately equal to the temporal fluctuation cues provided in the dynamic conditions (AM and AM2 had no level increment imposed by the modulation). The similarity between thresholds in AM1 and AM2 suggests that the listeners were probably not using an overall level cue in condition AM1. The finding that the INC and AM1 and AM2 stimuli produce similar thresholds supports the hypothesis that increment detection is linked to modulation detection (and not primarily based on the detection of an energy change).

Another comparison to make in Fig. 3-2(A) is across the open symbols (conditions with SAM in the target interval). In line with noise-carrier studies of Viemester (1979) and Sheft and Yost (1990), and with the tone-carrier experiments of

Yost and Sheft (1997), but seemingly at odds with our hypothesis based on the depth-discrimination task, we found that our listeners were more sensitive in a low- f_m AM-detection task with a temporal fringe than with gated carriers. Converting the thresholds to $20 \log m$, the difference between thresholds in the AM2 condition (-32 dB) and the gated AM condition (-26.5 dB) amounts to about 5.5 dB. Both the gated-continuous AM detection difference and the gated-continuous AM discrimination similarity can be qualitatively accounted for with a simple phenomenological model (see section IV. Modeling).

3.2.3 Discussion

3.2.3.1 Adaptation and change detection

The similarity between gated and quasi-continuous AM depth discrimination thresholds can be interpreted in terms of the adaptation mechanisms that have been used to qualitatively explain the audio-frequency asymmetry in performance seen in gated-carrier intensity discrimination and continuous-carrier increment detection (see Introduction). If an increased amount of adaptation in gated conditions underlies gated-continuous differences, then the current results suggest one of at least two conclusions in the envelope-frequency domain. Either there is little or no adaptation at the output of modulation-tuned channels, or if there is adaptation, then the response to an increment in AM depth must decrease with the same time course as the adaptation, so that the relative response increment remains constant as a function of time.

There is some peripheral physiology that initially appears consistent with a transformation supporting the latter interpretation. Smith *et al.* (1985) reported a decrease in the response modulation of ANFs as a function of time: the response modulation depth decreased with short-term adaptation (i.e., the effect lasted for approximately 10 ms). In contrast with the current study, Smith *et al.* (1985) used stimuli with high AM frequencies (150-600Hz), and imposed them on gated carriers with short (2.5 ms) rise-fall times. For the lower fluctuation rates (4-64 Hz) and slow (50 ms) ramp functions used here, it is unlikely that the small effect observed in peripheral physiology could have an impact on the observed similarity between gated and continuous AM depth-discrimination thresholds. This leads back to the alternative explanation, namely that there is negligible perceptual adaptation to AM stimulation.

Perceptual coding of AM is usually assumed to be strongly influenced by central processing factors. This is because modulation-tuned channels are not found in the periphery, and the temporal responses of auditory-nerve (AN) fibers can robustly follow modulations to significantly higher rates than the several hundred Hertz (Joris and Yin, 1992) that human listeners can detect as a temporal (i.e. not spectrally resolved) cue (Kohlrausch *et al.*, 2000). The responses of cells in the inferior colliculus (IC) appear to be more tightly coupled to psychophysical measures than peripheral responses (Joris *et al.*, 2004), although temporal adaptation (or a lack thereof) in IC neurons to AM stimuli with relatively long offset and offset ramps has not been previously reported.

An alternative way to conceptually account for the audio-frequency gated-continuous difference is to assume the existence of a modulation filter bank that processes the stimuli at the output of peripheral filters, generating an effective additional dimension. An increment in the intensity of a sound activates at least the low frequency modulation channels, where the amount of activity depends on the exact stimulus

characteristics and the transfer functions of the modulation filters. As recently shown by Gallun and Hafter (2006), increment detection thresholds can be quantitatively accounted for by assuming a modulation frequency selective analysis. In contrast, in the gated-carrier intensity discrimination conditions, the most effective cue is reflected in the dc component (or in the lowest available modulation filter) in such a model. The finding that a similar asymmetry between increment detection and intensity discrimination was not found in the AM domain may suggest that analogous circuitry, i.e. another “independent” (envelope) domain, does not exist, or has a negligible influence on perception.

3.2.3.2 Relation to previous work

Wojtczak and Viemeister (1999) compared intensity discrimination and low- f_m SAM detection with continuous-carrier pure tones across a wide range of standard SPLs, and arrived at an empirical formulation of the relationship between the two measures:

$$10 \log \frac{\Delta I}{I} = 0.44 \cdot (20 \log m) + D(f_m),$$

where $D(f_m)$ is a constant that depends only on

modulation frequency. For a 4-Hz signal AM, Wojtczak and Viemeister determined this constant to be 1.7; for the 10-Hz signal AM used in the current study, $D(f_m)$ would probably take on a slightly lower value. With continuous 70-dB SPL carriers, a 10-Hz modulation rate, and assuming $D(f_m)$ to be 1.7, our data are somewhat inconsistent with the proposed empirical formula: $10 \log \frac{\Delta I}{I} = -10.9$ dB (INC condition in Fig. 3-2), and

$0.44 \cdot (20 \log m) + 1.7 = -12.5$ dB (AM2 condition in Fig. 3-2). Decreasing the value of $D(f_m)$ or inserting the modulation thresholds measured with the AM2 stimuli (SAM with a DC component) would only make the equation’s predictive ability worse. The difference might be a result of the difference in onset and offset ramp durations used in the two studies: Wojtczak and Viemeister (1999) used 5-ms ramps for the increment detection task, and much longer effective ramps in the AM detection task (1/2 cycle of a 4-Hz sinusoid = 125 ms). We matched the ramp durations in both conditions to eliminate this potentially confounding factor: both INC and AM2 conditions used 50-ms ramps. Wojtczak and Viemeister (1999) also speculated that the empirical relationship might also hold for gated carriers, but did not test this hypothesis explicitly. The current data allow for such a test. In the gated intensity discrimination task (ID) here,

$$10 \log \frac{\Delta I}{I} = -6.6 \text{ dB, while in the gated AM detection condition (AM),}$$

$0.44 \cdot (20 \log m) + 1.7 = -10.0$ dB. The match to the proposed formula is worse in this condition, suggesting that it does not directly generalize to describe the relationship using gated stimuli.

3.3. EXPERIMENT II. TONE-IN-NOISE DETECTION WITH A MODULATED MASKER IN THE ENVELOPE-FREQUENCY DOMAIN

The goal of the second experiment was to determine the extent to which listeners could use slow and regular temporal fluctuations in the instantaneous depth of a stochastic masker modulation to aid in the detection of a deterministic (SAM) signal modulation. The stimuli were designed to maximize the availability of potential release-from-masking cues in an envelope-domain transposition of a typical audio-frequency CMR paradigm.

3.3.1 Methods

Details of the listeners, apparatus, and procedure were the same as in the first set of experiments. This section addresses any remaining differences, which were mainly limited to stimulus parameters.

The carrier was again a 5.5-kHz pure tone with 50-ms Hanning windows applied to the onset and offset. The overall SPL of the standard and target were normalized to have the same rms level as a 70-dB SPL pure tone. Observation intervals were separated by a 500-ms silent interval. In the two standard intervals, the tonal carrier was modulated by a Gaussian noise, which had a 120-Hz bandwidth (BW) and was geometrically centered around 32 Hz, from 8 to 128 Hz. The average masker modulation depth was -13.2 dB rms, ($m = 0.22$; for a 120-Hz BW, the noise had a ‘spectrum level’ of -34 dB). This combination of masker depth and BW was chosen to (1) ensure significant masking of the 32-Hz signal AM (presented only in the target interval), (2) avoid over-modulation (no stimuli with modulation depths greater than one were presented to the listeners), and (3) provide the opportunity for across-modulation-channel processes to enhance detection performance (by using a BW greater than that of the putative modulation filters, which are typically described as having half-power Q-values between 0.5 and 2, or effective BWs between 16 and 64 Hz for a channel tuned to 32 Hz).

Masker waveforms in each interval were independent noise realizations, generated digitally by setting the Fourier coefficients outside the desired pass-band to zero. In the baseline conditions (analogous to unmodulated conditions in audio-frequency CMR experiments), no further manipulations were made of the masker waveform before the 32-Hz sinusoidal AM (always in sine phase) was added and the resulting envelope signal imposed on the carrier. A general equation for the final stimulus is:

$$s(t) = c\{\sin(2\pi f_c t)[1 + m \sin(2\pi f_m t) + M(t)]\},$$

where c is a scalar that equalizes the overall audio-frequency power in each interval, f_c is the carrier frequency, m is the stimulus modulation depth (zero in the standard interval), f_m is the signal modulation frequency (32 Hz), and $M(t)$ is the masker waveform.

In contrast to the baseline condition, the comparison, or ‘comodulated’ masker waveforms were further processed before being combined with the signal SAM. Slow, coherent, and regular (sinusoidal or square wave) temporal fluctuations were imposed on the instantaneous masker modulation depth, $M(t)$. This resulted in a stimulus with a time-varying envelope. In all of the comparison conditions, the imposed envelope fluctuations were maximal in the sense that the nominal envelope depth of the masker varied between zero and the peak value. Examples of the time waveforms that were used to modulate the carrier are shown in Fig. 3-3; masker-alone (standard) waveforms are illustrated on the

left, signal-plus-noise envelopes are shown on the right. Baseline unmodulated masker (R), sinusoidally modulated masker (C_{SAM}), and square-wave-modulated masker (C_{SQ}) conditions are shown.

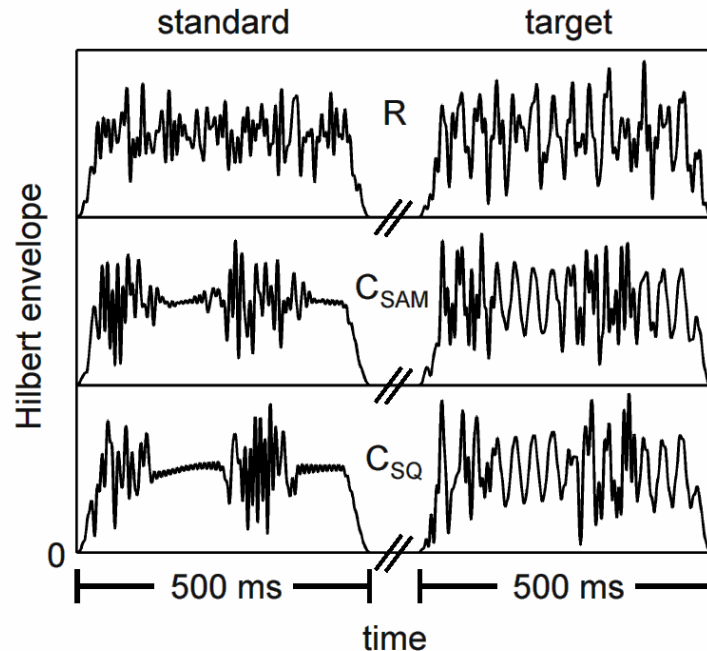


FIG 3-3. Example temporal envelopes of the stimuli used to test for AM-domain modulation masking release. Standard-interval envelopes (left) are defined by the masker-alone waveform; target interval envelopes (right) are made up of an additive combination of masker and sinusoidal signal AM. A: Baseline (unmodulated) condition. C_{SAM} : Sinusoidal envelope fluctuations. C_{SQ} : Square-wave envelope fluctuations. $f_c = 5500$ Hz; standard SPL = 70 dB; masker BW = 120 Hz, geometrically centered on the 32-Hz signal frequency; observation interval duration = 500 ms; signal depth $m = 40\%$; $f_m' = 4$ Hz.

Imposing slow fluctuations in the envelope can affect the resulting modulation spectra (i.e. sidebands are generated when the envelope is modulated, just as they are in the audio-frequency spectrum when a carrier is modulated). One way to avoid this complication is to filter the noise after it is modulated; the trade-off when using this strategy is that the temporal waveform can be changed, usually in the form of ringing caused by the band-limiting. To control for this issue, thresholds were measured when the masker envelope bandwidth was limited either before (condition C_{SAM}) or after (condition C'_{SAM}) imposing the slow envelope fluctuations in both the baseline and comparison conditions.

In the first part of the experiment, the envelope modulation rate, f_m' , was fixed at 4 Hz (two cycles were presented for each 500-ms signal), and the waveform used to modulate the (first-order) AM noise was varied, both in terms of its shape and its amplitude. In the extension of the experiment, the duration of the signals was extended to 2 s to allow for the use of even slower envelope fluctuation rates (from 0.5 Hz to 4 Hz). An equal-energy (in terms of the envelope rms), square-wave envelope modulator was used with the 2-s signals.

For comparison, thresholds were also measured for the same listeners in an audio-frequency CMR paradigm with parameters designed to (loosely) parallel those used in the envelope-frequency experiment. In both the audio- and envelope-domain experiments, detection thresholds of a mid-frequency sinusoidal signal embedded in a moderately intense and wideband (re: relevant filtering properties) Gaussian masker were measured. Slow and regular fluctuations were imposed on the masker in both domains; release from masking was defined as improved thresholds in the conditions using modulated maskers over those using noises with flat temporal envelopes. Specific audio frequency parameters were: signal frequency = 2 kHz; masker BW = 800 Hz (geometrically centered on the signal frequency); masker spectrum level = 30 dB SPL (overall rms level = 59 dB SPL); AM imposed on masker = 32 Hz sinusoid. Observation and inter-stimulus intervals were 500 ms. The tone level was adaptively varied initially in steps of 8 dB; the initial step size was halved after each of the first two track reversals occurring after consecutive correct responses until it reached 2 dB. Again, the mean of six reversals was taken as threshold for a given track.

3.3.2 Results

3.3.2.1 *No release from masking in the envelope-frequency domain for 500-ms stimuli*

The magnitude of audio-frequency CMR with a wideband masker centered on the tone frequency and fully modulated with a deterministic waveform in the comodulated conditions is illustrated in the left panel of Fig. 3-4. Thresholds were about 10 dB lower in conditions with a comodulated masker (C) than in the random (flat masker) case (R). The magnitude of the effect is close to that observed in a previous study using similar stimulus parameters (Verhey *et al.*, 1999).

The new contribution of the current experiment was to translate the parameters that result in significant audio-frequency CMR into the envelope-frequency domain. Pure-tone carrier SAM detection thresholds were measured in the presence of several types of additive modulation maskers. A release from masking would take the form of lower thresholds in the conditions with slow and regular variations imposed on the instantaneous masker modulation depth when compared to performance in the conditions with a flat-envelope (Gaussian) masker modulation.

Within the right panel of Fig. 3-4, SAM detection thresholds are shown for the four masker conditions described above. The average thresholds were all between -12 and -15 dB ($20 \log m$), and none of the comodulated condition thresholds were significantly different from those measured in the random condition (t-test p-values = 0.18, 0.34, and 0.68). In other words, the listeners were unable to take advantage of the slow and regular envelope fluctuations imposed on the first-order masker.

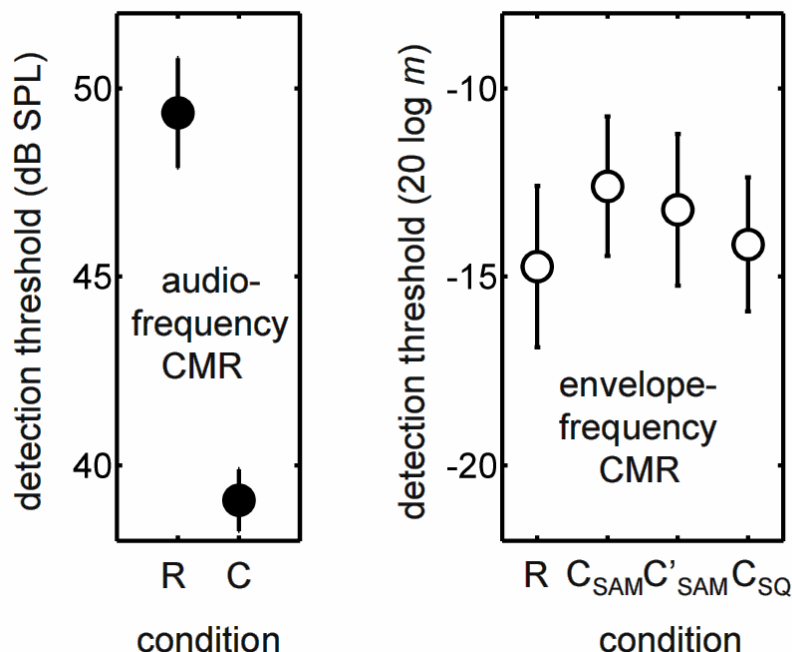


FIG 3-4. Effects of imposing slow and regular fluctuations on the masker amplitude in the envelope- and audio- frequency domains. Conditions correspond to different temporal shapes imposed on the masker amplitude. *Left panel:* Audio-frequency thresholds. R: random flat masker envelope (unmodulated). C: 32-Hz SAM masker envelope, filtered after modulation and not equalized for overall energy increment caused by modulation. *Right panel:* Envelope-frequency thresholds. R: flat envelope masker (unmodulated). C_{SAM} : 4-Hz SAM envelope, noise filtered after modulation. C'_{SAM} : 4-Hz SAM envelope, noise only filtered prior to modulation. C_{SQ} : 4-Hz square-wave envelope, noise filtered after imposing the 4-Hz fluctuations. Conditions C_{SAM} , C'_{SAM} , and C_{SQ} were compensated for the small overall increase in masker energy caused by the modulation.

3.3.2.2 Extending the time course of the slow masker fluctuations

In the square-wave envelope masker conditions above, the listeners were presented with two 125-ms segments of the unmasked pure SAM 32-Hz signal (4 complete cycles) between two 125-ms segments containing both the signal and masker modulation. This duration of pure-signal AM was insufficient to give rise to a release from masking. However, intuitively, one expects that there *must* be a release from masking if the periods of low masker depth are long enough. To further investigate the time course of the effect, it was necessary to increase the overall duration of each interval to 2 s to accommodate more than one cycle of the slow masker modulation. Square-wave envelope waveforms with rates of 0.5, 1, 2, and 4 Hz were imposed on the same modulation masker used with the 500-ms signals (120-Hz BW geometrically centered on the 32-Hz signal rate, with an average depth of -13.2 dB rms), and detection thresholds were determined again for a 32-Hz signal AM.

Detection thresholds for the 2-s stimuli are shown in Fig. 3-5. Individual thresholds are shown in addition to the mean results because of the relatively high inter-subject variability. For all four of the listeners, performance improved with decreasing envelope fluctuation rates over the range of frequencies tested. The parameters of the

stimuli used in the 4-Hz condition were identical to those used with the 500-ms stimuli (condition C_{SQ} in Fig. 3-4); mean performance improved by about 2.5 dB as a result of increasing the stimulus duration.

Thresholds asymptote near the expected pure-AM detection thresholds for listeners L1, L2, and L3; the remaining listener (L4) was less sensitive overall, and continued to show improvement between the 1 Hz and 0.5 Hz conditions. Overall, the listeners required at least 500-ms periods of unmasked SAM signal between masker bursts to reach performance near the 32-Hz SAM detection thresholds expected for a 70-dB SPL, 5.5 kHz pure tone carrier (between -25 and -30 dB; Zwicker, 1952; Fleischer, 1982; Kohlrausch *et al.*, 2000). The relatively high thresholds observed for the 2- and 4-Hz envelope frequencies suggests that the perceptually relevant decision variable is either integrated over a long interval, or that the internal representation of the signal AM is affected by preceding masker modulations for several hundred milliseconds. Anecdotally, listeners reported that the masker bursts were perceptually fused for envelope rates above 2 Hz and gradually became identifiable as temporally distinct and separable events at rates below 2 Hz.

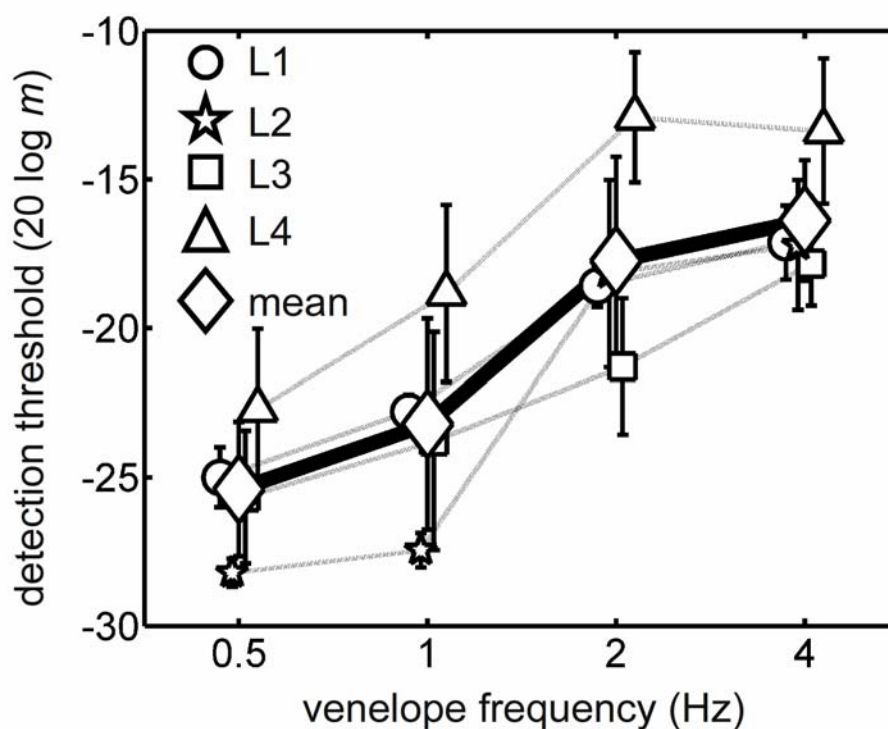


FIG 3-5. Individual and mean 32-Hz SAM detection thresholds as a function of the frequency of the square-wave envelope fluctuations imposed on the 1st-order masker modulation. Stimulus parameters were the same as those in condition C_{SQ} of Fig. 3-4, except the overall duration was increased to 2 s.

3.3.3 Discussion

The findings from Experiment II are consistent with those of Experiment I in that envelope fluctuations did not appear to contribute to performance in envelope-frequency-domain versions of either task. Both sets of experiments were designed to show that the analogous audio-frequency paradigms were strongly influenced by first-order envelope fluctuations. In other words, there does not appear to be an additional independent coding dimension that the listeners have access to in the transposed (modulation domain) experiments as there apparently is in the audio-frequency domain.

3.3.3.1 Relation to previous work

The results of our effort to measure CMR in the envelope-frequency domain are in qualitative agreement with the findings of several previous studies. Wojtczak and Viemeister (2005) also showed that a modulated envelope preceding a SAM signal imposed on the same carrier could affect detection performance for masker-probe delays of up to 200 ms. Their AM forward masking paradigm used a wideband noise carrier, a sinusoidal masker AM, and a short (50-ms) signal that was present only after the masker. The broadly similar time course of masking observed in the two studies suggests that a single mechanism could underlie both effects, and that it is independent of the statistical description of the carrier and masker. It is not immediately clear what this mechanism is; in fact, it would be reasonable to predict that neither effect would be seen based on the broadly tuned nature of the putative modulation filters: the “trade-off” for implementing broad signal-processing filters is that the response recovers quickly from stimulation.

One aspect of the stimulus that changes with the envelope fluctuation rate is the duration of the segments of pure unmasked SAM. If the listeners were ignoring the masked portions of the stimulus and basing decisions solely on a single period of unmasked SAM, then one could predict a purely duration-based improvement in thresholds at low envelope rates. Again, this argument assumes that detection is not mediated by analysis of the masked segments, or by multiple looks at the unmasked segments. While these assumptions are probably not realistic, it is worth pointing out what is known about duration effects on AM detection. Füllgrabe and Lorenzi (2003), using a 75-dB SPL, 9-kHz tone carrier and stimulus durations of 250, 500, and 2000 ms, showed that (1st-order) AM detection depends on duration, but thresholds for a 250-ms, 32-Hz signal were still < -25 dB. Thresholds for a 2-s pure SAM signal were about -32 dB. Füllgrabe and Lorenzi’s results are consistent with the idea that our listeners’ performance was affected by the masker (even at very low envelope fluctuation rates), and that the observed time course of masking was not simply a result of increasing the duration of the pure (unmasked) signal AM.

The information available to cochlear implant (CI) users is provided largely in the form of temporal envelope fluctuations imposed on the amplitude or duration of current pulses presented to stimulating electrodes. Nelson *et al.* (2003) and Nelson and Jin (2004) measured performance of CI users in a speech recognition task with a temporally modulated noise masker, and varied the “gate frequency” from 1 Hz to 32 Hz. They found that CI users did not benefit from temporal gaps in the noise masker as long as 500 ms (performance was independent of the gate frequency). Conclusions from the CI experiments can be rephrased to exactly match those from the current study: in conditions with severely impoverished spectral-frequency cues, listeners are unable to use relatively

long temporal gaps in a noise masker to aid in the detection of a signal. It is likely that the comparable quality of masker and signal in both studies leads to the fusion of temporally distinct masker events connected by a perceptually similar signal.

3.3.3.2 *Interpreting time courses*

The extended-time-course AM-detection experiment suggests a long integration time constant operating at some stage presumably central to the putative envelope-filtering process. Such a statement is consistent with “long-term” masked AM detection decision statistics that quantify responses based on an averaged representation of the processed stimulus envelope, such as envelope rms (e.g. Strickland and Viemeister, 1996; Ewert and Dau, 2000; Ewert *et al.*, 2002) or the average firing rate of a model IC cell (Nelson and Carney, in review). However, it is worth pointing out the current data set does not necessitate the assumption of such a time-averaged decision variable. It remains possible that a “local feature” decision variable (e.g. envelope max/min ratio or maximum local modulation depth) could be used, but that the listeners combine information from multiple looks (e.g., Sheft and Yost, 1990; Viemeister and Wakefield, 1991; Dau *et al.*, 1997) of the details of the local features.

Cortical physiological studies have provided evidence for long-lasting modulation of responses to envelope fluctuations that might underlie the apparent perceptual sluggishness observed here. Using pure-tone forward masking paradigms in the primary auditory cortex (A1), several groups have shown that the response to a short probe signal could be affected by the presence of a masker that preceded the probe by as much as several hundred ms or longer (e.g. Calford and Semple, 1995; Brosch and Schreiner, 1997; Ulanovsky *et al.*, 2004). If the masker had a similar audio-frequency composition to that of the probe (as it did in the current study), the response to the probe was usually suppressed. In a recent study of the unanesthetized marmoset A1 that used stimuli more similar to those used in the current psychophysical experiments, Bartlett and Wang (2005) examined the contextual dependence of AM responses on previous stimulation. Their findings were in qualitative agreement with those of the AM forward masking studies, but the observed suppression (or facilitation) of a probe AM stimulus could last longer than 1 s in some neurons and depended on the modulation properties of the preceding stimulus. To date, we are not aware of physiological results at any level that address the effect of a masker modulation (simultaneous or non-simultaneous) imposed on the same tone carrier as a deterministic signal modulation. In all of the studies mentioned above, the probe and masker were imposed on separate carriers (i.e. the stimuli were gated). It would be interesting to know whether physiological time courses of adaptation to AM are different for gated and continuous carriers.

An alternative explanation for the finding of similar thresholds in the random and comodulated envelope-domain detection task is that the listeners were basing decisions on the masked portion of the stimulus and largely ignoring the signal representation in the dips of the masker. This seems unlikely, since the signal-to-noise ratio (SNR) was optimal in the masker dips, but possible if the masker bursts were grouped together into a single auditory object. One hypothesis is that listeners chose the interval that contained the highest local modulation depth. This would likely occur near the peaks of the masker, where the instantaneous SNR was the lowest, and predicted thresholds should be qualitatively similar in the random and comodulated conditions.

3.4 MODELING

3.4.1 Methods

A physiologically motivated processing model developed to predict peripheral and central neural responses to pure SAM tones (Nelson and Carney, 2004) and psychophysical responses to masked SAM tones (Nelson and Carney, in review) was used to simulate responses to the stimuli used in the current study. The peripheral model was a modification of previous AN models (Carney, 1993; Zhang *et al.*, 2001; Heinz *et al.*, 2001a), and the final model output can be compared to pure-tone onset responders in the IC. Interactions between fast excitation and slow inhibition give rise to modulation tuning in the model IC cells; because the two inputs are matched in audio-frequency CF, the model is referred to as the same-frequency inhibition and excitation (SFIE) model as in Nelson and Carney (2004). In the current study, the relative strength of inhibition with respect to excitation at the level of the model IC cells ($S_{\text{INH,IC}}$) was set equal to 1.0. This parameter was important for determining the threshold modulation depth required to elicit firing in the model cells: values of $S_{\text{INH,IC}} \leq 1$ resulted in lower depth thresholds than $S_{\text{INH,IC}}$ values greater than one (i.e. stronger inhibition re: excitation). The time constants of inhibition (τ_{inh}) and excitation (τ_{exc}) were chosen to yield a cell that was tuned to the signal f_m of interest (see Nelson and Carney, 2004).

3.4.2 Results and Discussion

Simulation results are described and discussed next with three specific psychophysical observations in mind: (1) audio-frequency level-discrimination thresholds depend on the choice of gating mode (Experiment I), (2) AM detection thresholds depend on gating mode (Experiment I extension) but AM depth-discrimination thresholds do not (Experiment I), and (3) masked SAM detection thresholds do not readily improve when the masker is comodulated (Experiment II). The first finding is examined most carefully with the model, and those results are used as justification for the assumptions made with the remaining sets of data. In general, the modeling work is meant to qualitatively test the ability of an existing envelope-processing model to account for broad and basic features of the data. This approach intentionally lies between speculating on potential mechanisms and explicitly predicting listeners' thresholds with a specific (i.e. fitted) model.

3.4.2.1 Audio-frequency level discrimination with gated and continuous carriers

By definition, modulation-tuned neurons are also envelope change detectors, and the properties that underlie AM responses in these neurons can also qualitatively explain the audio-frequency gated-continuous difference. To illustrate this point and to provide a concrete example of a component of a realistic neural circuit that predicts a heightened sensitivity to increments over gated tone bursts, we applied the SFIE model to the audio-frequency stimuli used in the current study.

Instantaneous firing rate (IFR) functions are shown in Fig. 3-6 for two levels of model responses. The functions are comparable to physiological peri-stimulus time histograms (PSTHs) obtained at the level of the AN (i.e. Harris and Dallos, 1979) and the IC (i.e. Langner and Schreiner, 1988), and were generated for eight illustrative conditions. Both AN and IC model responses are included in each of the panels of Fig.

3-6, which correspond to a specific combination of gating mode (gated or fringe) and standard SPL (20 dB or 60 dB). The timing of the standard (S) and target (T) stimulus presentation is marked by the two horizontal bars from 0.5-1 s and 1.5-2 s. A 3-dB level increment in the target interval was used. Other parameters of the stimuli were matched to those used in the psychophysical experiment. The IC model time constants ($\tau_{exc} = 10$ ms; $\tau_{inh} = 20$ ms) were chosen to yield a cell tuned to the effective 10-Hz modulation rate caused by the 50-ms onset and offset ramps.

First, consider the model outputs in response to a 20-dB standard-interval SPL tone (upper panels of Fig. 3-6). For our purposes, the most critical differences between the modeled AN and IC responses are in the steady-state portion: ANFs respond with sustained firing to pure-tone stimulation, while the IC model only fires when the stimulus envelope elicits a change in the peripheral response. This is most clearly seen at envelope transitions with rising slopes, but offset adaptation in the peripheral model also results in a small response at the offset of the gated stimuli in the IC model. Note also that the IC model responds to both standard and target intervals when the intervals are gated, but only to the target interval in the fringe conditions. The AN model responses are always non-zero when a stimulus is present, although there is no change in the response from the baseline to the fringe-condition standard interval.

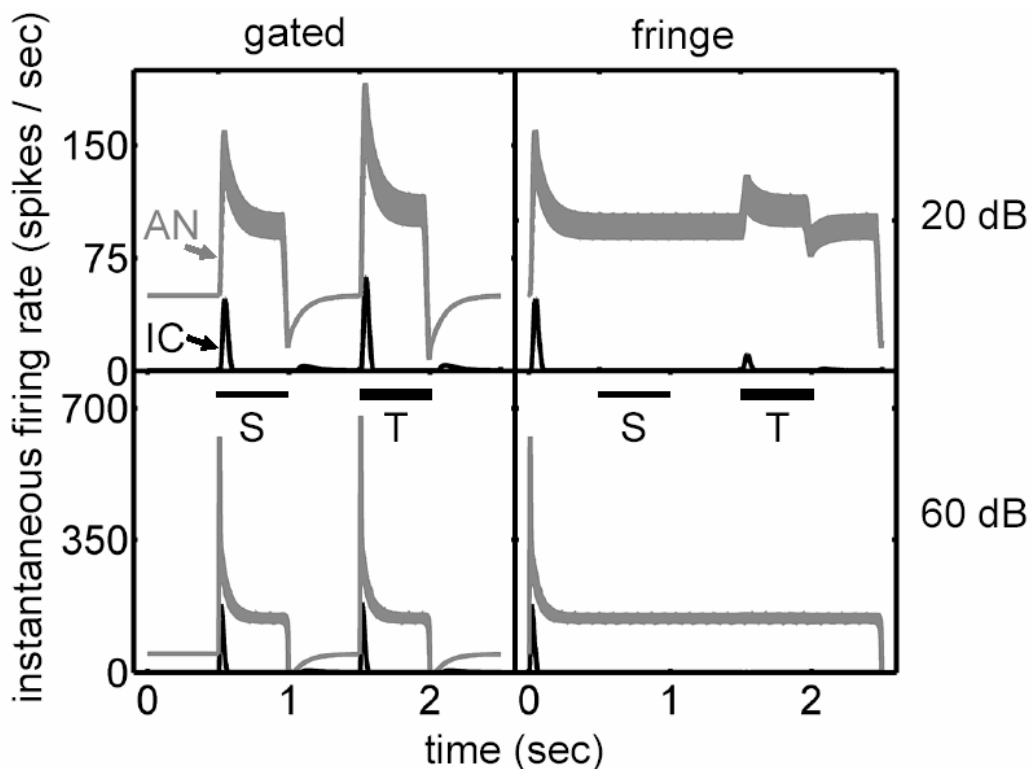


FIG 3-6. Simulated responses to standard and target stimuli for the AN and IC levels of the SFIE model. Upper panels: 20 dB SPL standard level; lower panels: 60 dB SPL standard level. The target interval level was 3 dB higher than the standard. Left panels: gated stimuli; right panels: fringe presentation mode.

Because of rate saturation effects in the AN model, the pattern of model responses was quite different when a 60-dB standard tone SPL was used (Fig. 3-6, lower panels). Specifically, the target interval increment did not elicit a change in the fringe-condition responses of either the AN or IC model. As a result, a model consisting only of high-spontaneous rate (SR), on-CF ANFs at medium to high SPLs predicts unrealistically high level discrimination thresholds (Colburn *et al.*, 2003). There are two popular ways to account for psychophysical performance at high SPLs and high frequencies. One is to unevenly and heavily weight the contribution of low-SR ANFs (Winslow and Sachs, 1988; Delgutte, 1987; Viemeister, 1988). This approach is not completely satisfying, because such high-threshold, wide dynamic range ANFs make up only ~15% of the total population in cat (Lieberman, 1978) and only exist at high CFs (> 1500 Hz, Winter and Palmer, 1991). Another aspect of the response to high-level tones that may provide information for discrimination is in the spread of excitation across a population of neurons (Viemeister, 1972; Florentine *et al.*, 1987; Heinz *et al.*, 2001b; Colburn *et al.*, 2003). To address this issue, standard and target stimuli were presented to a group of model cells with different CFs.

Responses across the population were quantified in terms of their average rate over the entire 500 ms of the stimulus. Peripheral (AN) differences in the model's rate responses are shown in the upper panels of Fig. 3-7, and central (IC) rate differences are plotted in the lower panels. The parameter in each panel of Fig. 3-7 is the standard level. As in Fig. 3-6, gated and fringe conditions are illustrated in the left and right panels, respectively. For all four combinations of gating mode and model level, the biggest differences in rate between the target and standard interval moved progressively away from the tone frequency (5500 Hz) as the standard SPL was increased. This is consistent with previous studies using simulations of high-SR ANFs (e.g. Siebert, 1968; Heinz *et al.*, 2001b; Colburn *et al.*, 2003), and is caused entirely by saturation in the present model (a linear basilar membrane model was used). A model version with level-dependent bandwidth and gain (i.e. a time-varying compressive nonlinearity, Heinz *et al.*, 2001a) was also tested, and a similar pattern was obtained, suggesting that effects caused by saturation dominate those caused by compression when low-threshold, high-SR ANFs are used to estimate the population response.

Another feature of the differences in model rates was that the shapes of the profiles were similar at both levels of the model. This was a direct result of the simplified nature of the SFIE IC model neurons. Absolute values of the rate differences were significantly higher in the AN model (note the different scales for the upper and lower panels); this was not surprising given the sustained nature of responses to pure tones in the AN model and transient characteristics of the IC model responses. Finally, comparing across the gated and fringe conditions, the changes in model rates were not strongly dependent on the mode of gating for either the AN or IC model population.

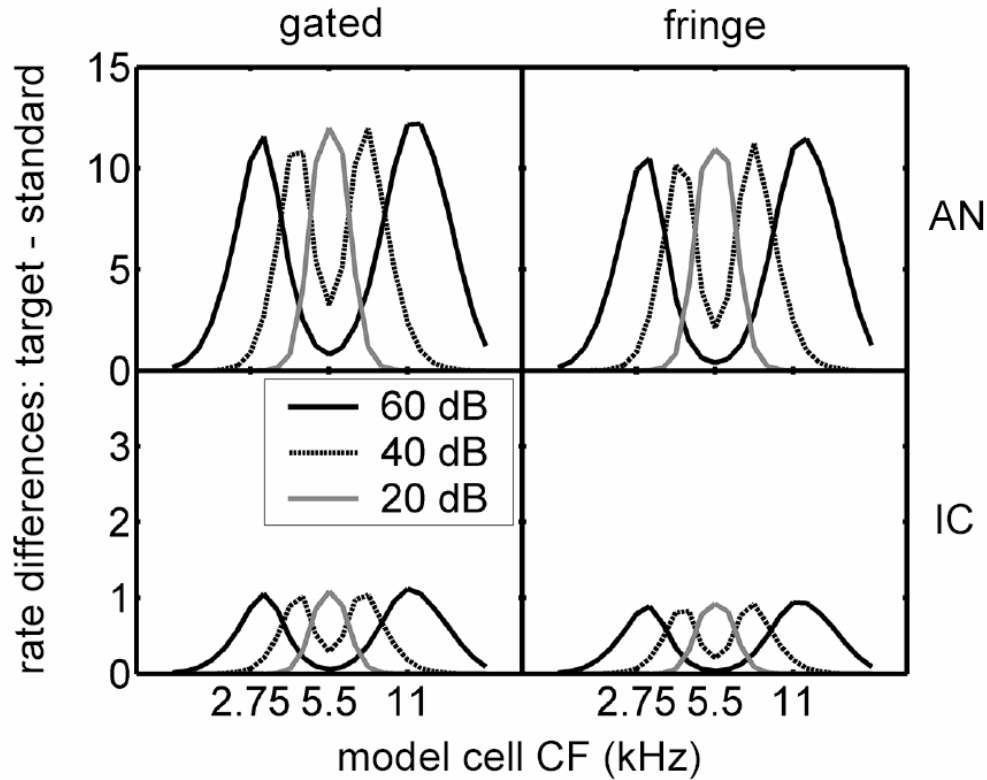


FIG 3-7. Rate difference profiles for different gating modes (left and right panels) and levels of the model (upper and lower panels) in response to a 3-dB level increment. Each curve represents changes in the model rate responses for a 5500-Hz tone with a fixed standard SPL. Twenty-five model cells, log-spaced from 1,375 Hz to 22,000 Hz, were simulated for each standard level and gating mode.

The similarity of absolute rate differences for the gated and fringe conditions at both model levels does not provide a compelling explanation of the gated-continuous difference. It does, however, lead us to a consideration of another feature of neural responses that must be known (or assumed) before predicting performance: the variability of rate estimates (i.e. Siebert, 1965). In actual ANFs, rate variability can be reasonably described as Poisson, with spike-count variance approximately equal to the mean count (at least at low rates, see Young and Barta (1986) and Winter and Palmer (1991) for a description of the reduced-variance deviation from Poisson at high rates). The situation is less clear in more central processing stations, but for simplicity, we will assume Poisson variance at both levels of the model responses. These variance characterizations allow for a relatively simple formulation of the information provided by each frequency channel in the population rate response (following the approach of Siebert, 1965; for details and derivations, also see Heinz *et al.*, 2001b):

$$\partial^2(\text{icf}) = \frac{[(\text{rate}_T - \text{rate}_S) / I_{inc}]^2}{\sigma_{rate}^2},$$

where $(\text{rate}_T - \text{rate}_S)$ is the rate difference term plotted in Fig. 3-7, I_{inc} is the size of the target-interval increment (3 dB in the examples shown), and σ_{rate}^2 is the variance of the

rate response. Our Poisson assumption allows for a simple estimate of σ_{rate}^2 : it is simply equal to the average rate (across both standard and target responses).

Information profiles, which incorporate both changes in rate and contributions of assumed neural variability, are shown in Fig. 3-8, in a format identical to that used to visualize the rate differences alone in Fig. 3-7. The limits of the y-axes are identical in all four panels. When the response variability is taken into account, the AN population model still predicts no advantage in the fringe condition relative to the gated presentation mode. In contrast, the envelope-change-detecting IC model clearly predicts a heightened sensitivity to the fringe condition, for all three tested standard levels. In addition, the overall summed population d' (related to the area under the information profile curve) is higher for the fringe-stimulus IC model rate responses than for the peripheral AN model responses. The gated-continuous difference in the IC model is strongly influenced by the lower average rate in the fringe condition, which translates into lower assumed variability and higher values of d' .

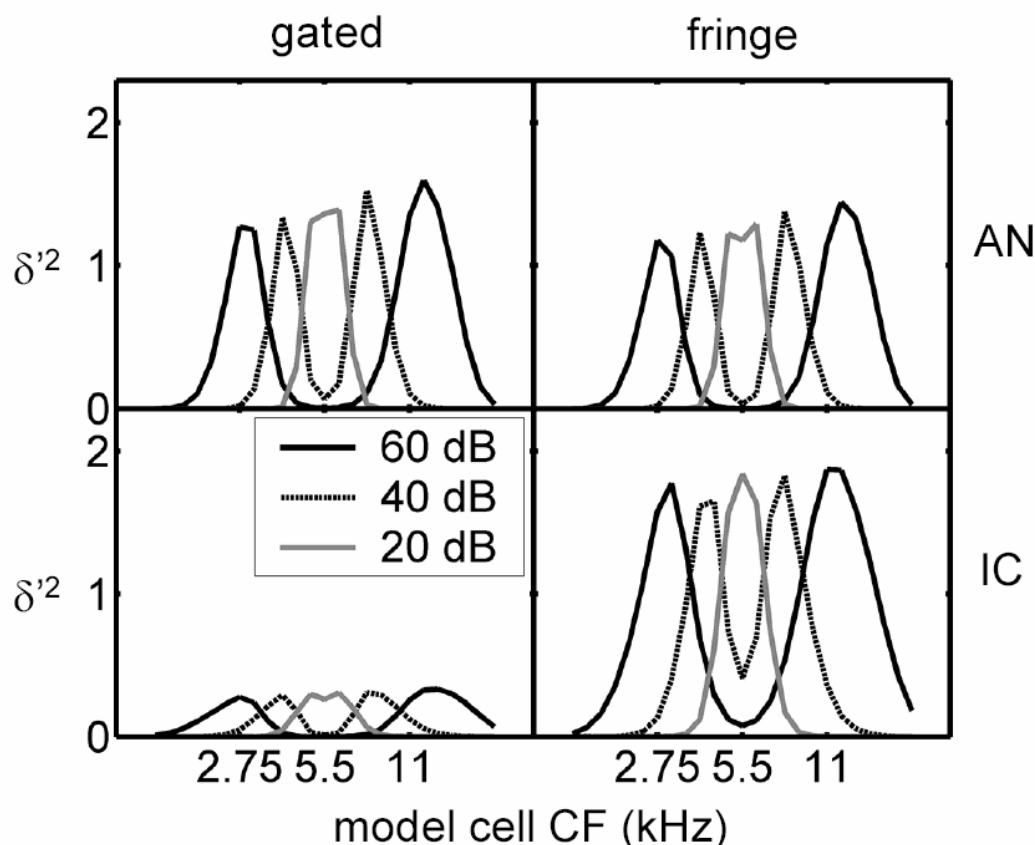


FIG 3-8. Across-frequency information profiles, arranged in the same format as Fig. 3-7. This measure of sensitivity takes both neural variance and changes in average rates into account.

While the exact values of predicted d' and thresholds depend on details of the parameters of the model and the statistical description of the chosen internal noise assumption, overall trends and the difference between gated and fringe conditions for the SFIE model with equal amplitude inhibition and excitation do not. One of the key features of the IC model that underlies the current explanation of the audio-frequency gated-continuous difference is the fact that there is some response to both intervals in the

gated condition, and only a response to the target interval in the fringe condition. The other critical assumption is an internal noise process that predicts response variability that increases with average rate. Such a change-detection mechanism could in theory be independent of peripheral adaptation, although there is an interaction between the two in the model, and some interplay probably exists in the real system as well.

3.4.2.2 Envelope-frequency detection and discrimination with gated and continuous carriers

AM detection thresholds (at least at low modulation rates) depend on whether the carrier is gated or quasi-continuous (Fig. 3-2 of the current study; see also Viemeister, 1979; Sheft and Yost, 1990; Yost and Sheft, 1997): thresholds are significantly higher when a gated carrier is used. In contrast, we found AM depth discrimination thresholds ($m_s = 0.25$) to be statistically identical with gated and quasi-continuous carriers. Based on the results from the preceding section, we will focus the analysis of the model in this section to off-CF IC model responses, where the biggest differences between standard and target were found.

IC model IFR functions are shown in Fig. 3-9 for 10-Hz SAM detection (top two panels) and 10-Hz SAM depth discrimination (bottom panels) paradigms. The observation intervals were 0.5 s; the standard interval started at 0.5 s, and the target at 1.5 s. Labels in the upper left corner of each panel indicate the gating mode for each response. Simulated PSTHs for the AM-detection paradigm were similar to the IC model responses in Fig. 3-6 for audio-frequency level discrimination, in that the gated stimuli elicited a response in both the standard and target interval, while the fringe stimulus resulted in a model response only in the target interval. Again, if differences in both rate and variance are considered, model responses predict an enhanced sensitivity to the fringe condition compared to the gated condition (see preceding section).

In contrast, for AM depth discrimination (lower panels of Fig. 3-9) the IC model responded to both standard and target interval in the gated and fringe conditions. Since both rate differences and assumed rate variability are similar in these conditions, the IC model predicts little or no difference in thresholds between the gated and fringe presentation modes (as observed in Experiment I).

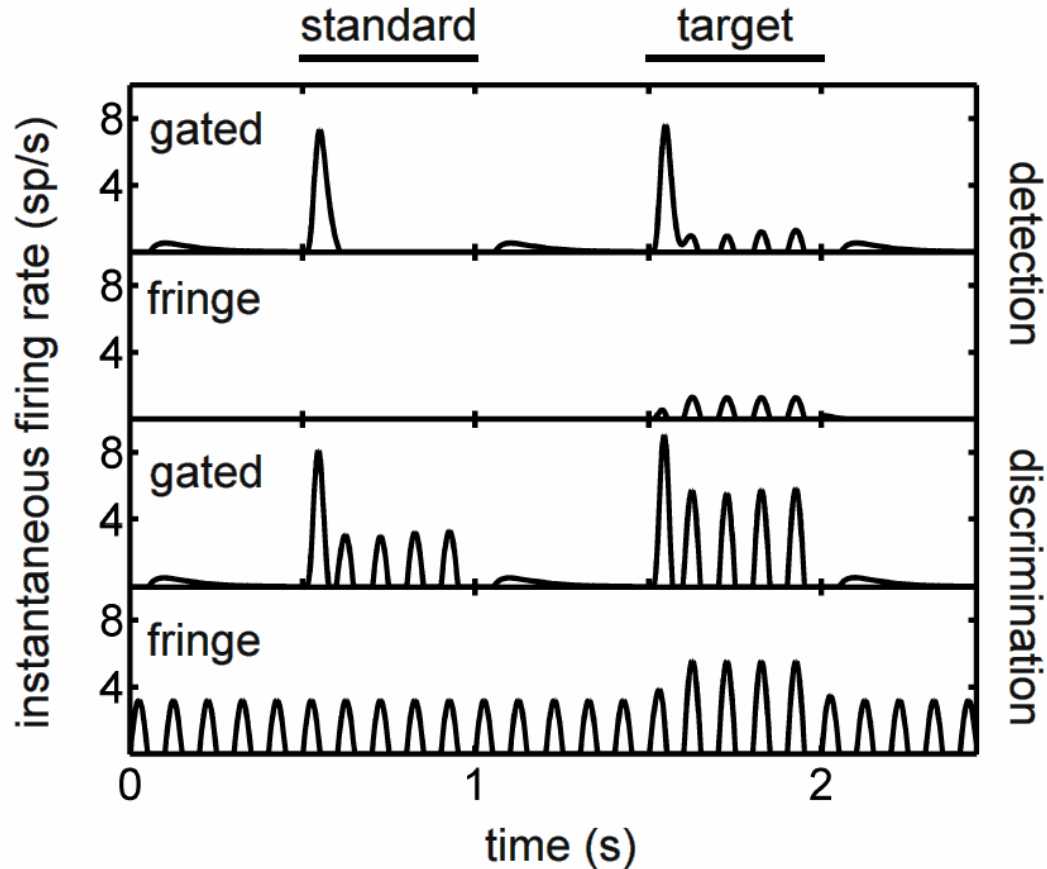


FIG 3-9. SFIE model responses are qualitatively consistent with a fringe advantage in AM detection and no fringe advantage in AM depth discrimination. Model responses are shown for a 2.5-s window centered on the presentation of a standard followed by a target modulation. Simulated PSTHs are shown for an AM detection paradigm (top two panels) and an AM depth discrimination task (bottom two panels); gated and fringe conditions are included for both tasks. Stimulus parameters: $f_c = 5.5$ kHz; SPL = 70 dB; $f_m = 10$ Hz; detection $m_t = -20$ dB; discrimination $m_s = -12$ dB, $m_t = -7$ dB. Key model parameters: $\tau_{exc} = 10$ ms; $\tau_{inh} = 20$ ms; $S_{INH,IC} = 1$; AN CF = 2000 Hz.

3.4.2.3 CMR experiment

Two questions were investigated concerning the ability of the model responses to qualitatively predict psychophysical trends observed with the stimuli used in the envelope-frequency-domain CMR experiment. First, does the model correctly predict the absence of a release from masking with the 4-Hz envelope modulation rate and 32-Hz signal SAM rate, as used in the base experiment? Second, is there an effect of envelope modulation rate over the range considered in the extension to the base CMR experiment?

To address the first question, a fixed-level signal SAM (-15 dB in $20 \log m$) was added to the masker in the target interval; model responses were quantified in terms of their average firing rates (across ten independent noise waveforms) in the random (R) unmodulated condition and in the 4-Hz square-wave comodulated masker condition for both standard and target intervals. If the model was consistent with the listeners'

thresholds, the difference in the response to the standard and target intervals should be independent of the masker condition (random or comodulated at 4 Hz). Fig. 3-10 shows that this was not the case; average rates in the random condition (gray symbols) were more similar in the standard (o) and target (□) intervals than the rates in the 4-Hz comodulated condition (right-most connected points). The disparity is caused by reduced response magnitude in the standard intervals of the comodulated conditions; target-interval rates were largely independent of the fluctuation patterns imposed on the masker. The overall long-term envelope rms energy was identical in all of the noise-alone (standard) intervals shown in Fig. 3-10; the suppression in rate for the comodulated condition was therefore caused by a non-linear relationship between envelope rms and model IC cell average rate. The main factor contributing to this relationship was the change in the slope of the rate vs. stimulus modulation depth function: at low m , the rate of change in rate was lower than at higher m . For example, when the modulation depth of a pure 32-Hz SAM signal was varied and presented to the cell simulated in Fig. 3-10, the slope of the function was ~ 0.1 sp/s/dB for $-30 \text{ dB} < 20 \log m < -25 \text{ dB}$, and ~ 0.9 sp/s/dB for $-5 < 20 \log m < 0 \text{ dB}$. In other words, responses to small effective modulation depths (such as the “ripples” caused by post-modulation filtering of the masker, or fluctuations in the masker away from the cell’s best modulation frequency) were strongly attenuated in the model, which resulted in a reduced overall response to the modulated standard-interval stimulus. To reiterate, the rate responses of the model IC cell were not consistent with the listeners’ inability to use the fluctuations in the masker to improve thresholds in the masked detection task. The schematic IFR functions included in Fig. 3-10 along with the rate quantifications show that the signal representation in the temporal responses of the model IC cells also suggest an increased salience of the signal in the comodulated conditions.

The second question posed at the beginning of this section was whether the model responses varied as the envelope frequency changed from 0.5 Hz to 4 Hz, as the psychophysical thresholds did. The long effective time constants apparently necessary to explain the time course of release from masking observed in the extension of Experiment II (on the order of hundreds of ms) are not included in model IC cells tuned to a 32-Hz signal frequency. This statement is supported by the simulations, which did not show any evidence for long time constants: rates in the standard and target intervals for the comodulated conditions were independent of the envelope fluctuation rate (connected symbols in Fig. 3-10.)

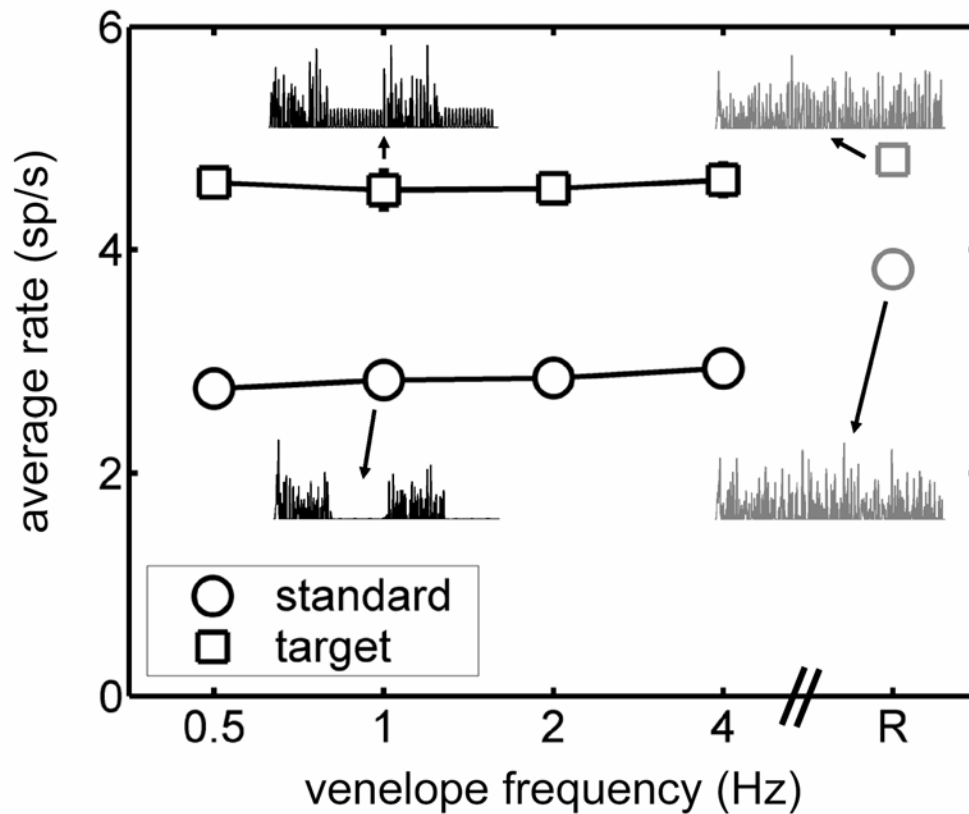


FIG 3-10. Model IC cell average rates and example IFR functions in response to the stimuli used in the envelope-frequency CMR paradigm. Model parameters were the same as those used to generate the responses in Fig. 3-9, except the AN CF = 2250 Hz, $\tau_{\text{exc}} = 3$ ms and $\tau_{\text{inh}} = 10$ ms, which resulted in a cell rate-tuned to the 32-Hz signal AM frequency. The signal depth was fixed at -15 dB (20 log m) in the target interval, and a square-wave envelope was imposed on the masker in the comodulated conditions (as in the extension to Experiment II). For comparison, rates and IFRs elicited by the random (R), or unmodulated condition, are also included in the plot (the duration of stimuli and IFRs were 2 s).

Although not explicitly tested with the stimuli used here, it is worth noting that the signal-processing-style envelope power-spectrum model (EPSM, Ewert and Dau, 2000) can also conceptually account for certain aspects of the current data. For instance, the absence of a gated-continuous difference for depth discrimination can be predicted by the EPSM, as there is no explicit adaptation or other context-dependent features included in the model. Also, the lack of a release from masking in the CMR task is predictable based on a calculation of the long-term envelope energy (as in the EPSM), since the rms modulation depth of the masker was identical in the random and comodulated conditions. However, the slow time course of masking release observed in the extension of Experiment II cannot be accounted for without modifying the most recent versions of the EPSM to include some relatively long temporal dependence of the model response or its quantification.

3.5 SUMMARY AND CONCLUSIONS

Two specific audio-frequency paradigms were transposed into the modulation domain to assess the perceptual salience of envelope cues in continuous-carrier depth discrimination and SAM detection in the presence of a slowly varying masker. The experiments described here suggest a minor effect of temporal structure on performance in the two envelope-processing tasks. Several specific conclusions can be drawn from the empirical data and modeling results:

- (i) Tone-carrier SAM-depth discrimination thresholds are not dependent on the gating mode of the carrier (i.e. gated or quasi-continuous), for a standard modulation depth of -12 dB ($m_s = 0.25$) and modulation frequencies from 4 to 64 Hz. This contrasts with audio-frequency level-discrimination results, which clearly indicate a heightened sensitivity to intensity increments when compared to gated-carrier threshold measurements.
- (ii) SAM detection thresholds (or discrimination with a standard depth $m_s = 0$) are approximately 5-6 dB lower when a quasi-continuous carrier is used than when the observation-interval stimuli are gated (for $f_m = 10$ Hz).
- (iii) Masked detection thresholds of a 32-Hz signal AM do not improve when the masker is slowly and coherently modulated with a 4-Hz envelope fluctuation rate. This is true for both sinusoidal and square-wave shaped comodulation. Audio-frequency tone detection thresholds, on the other hand, are strongly affected by the properties of the masker modulation.
- (iv) To observe CMR in the envelope-frequency domain, the time course of the coherent modulation must be lengthened until the masker bursts occur as perceptually distinct events (i.e. envelope fluctuation rates ≤ 1 Hz for a 32-Hz signal).
- (v) A simple model developed to predict responses to SAM tones in the auditory midbrain can qualitatively account for several of the results, including the gated-continuous difference for pure-tone level discrimination and AM detection and the gated-continuous “similarity” for AM depth discrimination. The model does not, however, explain the listeners’ inability to use relatively slow fluctuations in the instantaneous masker modulation depth to improve performance in the envelope-domain masked-SAM detection task. Higher-order processing, possibly associated with auditory grouping and/or segregation mechanisms, may need to be considered to account for results in the CMR task.

ACKNOWLEDGMENTS

This research was supported by NIH-NIDCD F31-7268 (PCN), NIH-NIDCD R01-01641 (LHC, PCN) and the Danish Research Council (TD, SDE).

CHAPTER 4

Rate and timing cues for neural detection and discrimination of amplitude-modulated tones in the awake rabbit inferior colliculus

ABSTRACT

Neural responses to amplitude-modulated (AM) tones in the unanesthetized rabbit inferior colliculus (IC) were studied in an effort to establish explicit relationships between physiological and psychophysical measures of temporal envelope processing. Specifically, responses to variations in modulation depth (m) at the cell's best modulation frequency were quantified in terms of average rate and synchronization to the envelope over the entire perceptual dynamic range of depths. Significant variations in the metrics were used to define neural AM detection and discrimination thresholds. Synchrony emerged at modulation depths comparable to psychophysical AM detection sensitivities in some neurons, while the lowest rate-based neural thresholds could not account for psychoacoustical thresholds. The majority of rate thresholds (85%) were -10 dB or higher (in $20 \log m$), and 16% of the population exhibited no systematic dependence of average rate on m . Neural thresholds for AM detection did not decrease systematically at higher SPLs (as observed psychophysically): thresholds remained constant or increased with level in 22 of the 33 cells tested at multiple SPLs. At depths higher than the rate-based detection threshold, some rate modulation-depth functions were sufficiently steep with respect to the across-trial variability of the rate estimates to predict depth discrimination thresholds as low as 1 dB (comparable to the psychophysics). Synchrony, on the other hand, did not vary systematically with m in many cells at high modulation depths. We conclude that the neural representation of AM in the IC likely transitions from a temporal code at low depths to a rate or synchronized-rate code at high depths. A simple computational model was extended to reproduce many features of the modulation frequency- and depth-dependence of both transient and sustained pure-tone responders. *Note: A version of this chapter will be submitted to the J. Neurophysiol., as a paper with the same title, by P. C. Nelson and L. H. Carney.*

4.1 INTRODUCTION

Temporal envelope fluctuations abound in natural acoustic landscapes, and the preservation of psychophysically relevant amplitude modulations is essential for robust vocalization perception (Drullman 1995; Shannon et al. 1995), auditory grouping and stream segregation (Bregman 1990), and signal detection in the presence of competing maskers (e.g. Hall et al. 1984). Although the auditory system is often described in terms of its audio- or spectral-frequency tuning properties, many central neurons are actually more sensitive to degradations of the stimulus temporal envelope than to smearing of spectral content (e.g. Nagarajan et al. 2002, Theunissen and Doupe 1998).

Basic psychophysical envelope-processing tasks have received renewed attention recently, due in large part to the success of a model of the “effective” signal processing of

the auditory system (Dau et al. 1997) in predicting behavioral data that are difficult to interpret unless one assumes the existence of a bank of filters tuned in the amplitude-modulation (AM) frequency domain (e.g. Bacon and Grantham 1989, Dau et al. 1997, Ewert and Dau 2000, Houtgast 1989). Such a conceptual framework is fundamentally different from the assumptions of established models of AM perception, which describe the putative central processor as a low-pass filter (e.g. Viemeister 1979).

Physiological studies of responses to AM provided some of the motivation for psychophysical investigations of frequency selectivity in the AM domain (e.g. Creutzfeldt et al. 1980, Langer and Schreiner 1988), but specific hypotheses concerning the relationships between physiological responses and perceptual signal-processing models have not been adequately examined. One possibility is that single neurons in the auditory midbrain function as modulation filters. A qualitative scan of the relevant literature suggests this may be a reasonable hypothesis, since many neurons in the inferior colliculus (IC) systematically change their responses with variations in modulation frequency (f_m) and modulation depth (m) (Krishna and Semple 2000, Langner and Schreiner 1988, Mueller-Preuss et al. 1994, Rees and Moller 1983). The next portion of this introduction provides a more detailed description of the published physiological responses to AM that should be considered to thoroughly examine this hypothesis, with a focus on the gaps in evidence that the current experiments were designed to fill.

Neural representations of sounds with dynamic temporal envelopes change dramatically as the auditory neuraxis is ascended. Much of our understanding about this transformation comes from studies of physiological responses to stimuli with systematically varied AM frequencies. From this body of work, a reasonably consistent picture has emerged: peripheral neurons appear to carry envelope-frequency information in a temporal (phase-locked code), with average rates that do not change with stimulus f_m (Joris and Yin 1992). More central neurons, in contrast, often exhibit average firing rates that are strongly dependent on f_m , and a reduced ability to synchronously follow faster fluctuations (for a review, see Joris et al., 2004). This frequency-focused description of neural responses to AM leaves a fundamental issue pertaining to the relationships between physiology and psychophysics unclear. Specifically, modulation transfer functions (MTFs) only provide information about responses to a single (usually high) modulation depth; as a result, direct comparisons to behavioral data are difficult because the goal of much of the relevant psychoacoustics is to determine the smallest detectable or discriminable m . A major objective of the current study was to obtain neural responses to stimuli with a wide range of modulation depths, including depths near psychophysical AM-detection thresholds.

A few studies have reported physiological responses to variations in stimulus modulation depth. In auditory-nerve fibers (ANFs) and most ventral cochlear nucleus (VCN) units, synchronization to the envelope increases monotonically with depth, with average rates that are largely depth-independent (AN: Joris and Yin 1992; VCN: Rhode 1994). Such generalizations cannot be made about AM-depth processing at higher levels of the central auditory system because of the striking response diversity. For instance, in the superior olivary complex (SOC), changes in the response with m are strongly correlated to the units' pure-tone response properties. Sustained pure-tone responders in the periolivary nuclei of the SOC tend to be similar to ANFs in terms of the shape of rate- and synchrony- modulation depth functions (rMDFs and sMDFs), while offset responders

exhibit monotonically increasing rMDFs and saturating sMDFs with narrow dynamic ranges (Kuwada and Batra 1999).

Krishna and Semple (2000) have provided the most complete account of neural responses at the level of the IC across a range of modulation depths. They found that average firing rates in many cells varied monotonically with m , especially near the cell's preferred modulation frequency. The change in rate could be an increase or a decrease with m , depending on the presence of regions of excitation and suppression in the cell's rate modulation transfer function (rMTF) and their relationship to the chosen stimulus modulation frequency. Temporal response patterns also changed with m in the anesthetized gerbil. A minimum depth was required to elicit significant synchrony in individual cells; this value ranged from as low as 10% (the lowest depth tested) in some of the neurons to 70% in others (Krishna and Semple 2000). Changes in vector strength above the minimum m were less stereotypical: synchrony in some neurons varied over a wide range of depths, but remained constant in most cells. Results from other (less systematic) studies are in qualitative agreement with the single-unit IC modulation-depth dependence description of Krishna and Semple (Mueller-Preuss et al. 1994, Nelson et al. 1966, Rees and Moller 1983). Relatively little is known about cortical responses to variation in m ; Eggermont (1994) and Liang et al. (2002) measured tone-carrier modulation transfer functions (MTFs) in primary auditory cortex neurons at several depths from 25% to 100%, and concluded that (1) MDFs were monotonic, and (2) neural best modulation frequencies were essentially independent of m .

A survey of this previous work allows for a qualitative description of neural responses to variations in modulation depth, but a direct and quantitative comparison of physiological responses at any level of the pathway to basic psychophysical AM detection and discrimination performance is still lacking. Two requirements for such a comparison to be made are met in the current study. First, the stimulus parameter space used in the physiology was designed to match that of the psychophysics. Specifically, m was varied from below detection threshold to 100%, in some cases using step sizes smaller than the behavioral just-noticeable difference (jnd). Second, a description of the statistical variability of the neural responses was included to quantify the significance of small changes in a given response metric (e.g. average rate and synchrony).

The IC is an inherently interesting nucleus in which to study AM processing. Structurally, it occupies a critical position in the subcortical processing pathway, as an almost obligatory ascending synapse (Aitkin and Phillips 1984, Malmierca et al. 2002, Ramon y Cajal 1904) and a receiving station for both inhibitory and excitatory inputs converging from afferent (Warr 1982, Winer et al. 1995), efferent (Winer 2004), and intrinsic and commissural connections (Saldana and Merchan 2004). In contrast to the established anatomical description of the IC and its connections, the functional representation of modulated sounds in the IC is still a matter of debate (Joris et al. 2004), but it is clear that both magnitude (rate) and phase (synchrony) information is present in the responses of single neurons (e.g. Krishna and Semple 2000, Langner and Schreiner 1988, Rees and Moller 1983). Thus, the IC apparently plays a transitional role between the temporal representation of AM in the periphery (Joris and Yin 1992) and a more rate-based code in the cortex (Liang et al. 2002).

Here, we show that changes in the average firing rates of single IC neurons in the awake Dutch-belted rabbit are generally poor predictors of human behavioral

performance in psychophysical AM detection tasks. Synchronization to the envelope, on the other hand, can emerge and change at modulation depths much closer to psychoacoustical thresholds. At supra-threshold depths, the situation is different; changes in average rates can, in some neurons, account for psychophysical sensitivity in masked AM detection and AM depth discrimination.

4.2 METHODS

4.2.1 Animal preparation

AM responses in the IC were obtained in three unanesthetized female Dutch-Belted rabbits (*oryctolagus cuniculus*). All procedures were approved by the Syracuse University Institutional Animal Care and Use Committee, and conformed to National Institutes of Health guidelines and protocols. Our preparation was developed based on techniques used in several previous studies of the awake rabbit IC and superior olivary complex (e.g. Batra et al. 1989, Kuwada et al. 1987). Before recordings began, two separate aseptic surgeries were performed to allow for chronic access to the midbrain in daily 2-hour recording sessions. In both procedures, the animals were anesthetized with ketamine (66 mg/kg) and xylazine (2 mg/kg) delivered intramuscularly, and supplemental doses were administered to maintain areflexia.

In the initial surgery, a 15-mm inner-diameter stainless-steel cylinder and brass headbar (aligned parallel to the sagittal suture) were centered on the midline and affixed to the exposed skull with dental acrylic and screws threaded into the skull. The rostral edge of the cylinder was aligned with bregma and a wall of dental acrylic was built up under the posterior side of the cylinder to compensate for the slope of the skull.

Each animal was given several weeks to recover from the first surgery before it was gradually adapted to sitting in the recording chamber and exposed to auditory stimuli. The rabbit was restrained with a snug blanket around the body and placed in a plexiglass chair positioned in front of a clamp used to fix the headbar. Daily sessions were increased in duration over the course of 2-3 weeks until the animal was acclimated to sitting quietly for 2 hours.

A small (approximately 3-4 mm diameter) craniotomy was made in the skull in the second anesthetized surgery. The medial edge of the hole was approximately 2 mm lateral of the midline, and the rostral edge was slightly forward of the middle of the cylinder. The exposed dura was rinsed with sterile saline, treated with a topical antibiotic (Bacitracin), and the cylinder was filled with a sterile silastic elastopolymer cap between sessions. A 1-2 day recovery period was allowed before removing the silastic cap and attempting electrode penetrations. Additional craniotomy surgeries were occasionally performed to extend the existing hole, or to provide access to the opposite IC.

After every session, new dural scarring was removed with forceps before re-applying Bacitracin and filling the cylinder with the polymer plug. In one rabbit, the dura was also treated with an anti-mitotic compound (5-Fluorouracil) before sealing the cylinder to discourage scar tissue from forming between sessions (Spinks et al. 2003). Using these daily cleaning techniques, recording sessions yielded reasonable success rates in a single IC for 3-6 months.

4.2.2 Acoustic stimuli

Sound stimuli were generated digitally and converted to analog signals using a Tucker-Davis System II D-A converter (TDT DA 3-4). The stimuli were anti-alias filtered at 20 kHz (TDT FT6) and attenuated (TDT PA4), before being passed to a headphone buffer (TDT HB6) and finally to a pair of Beyer-Dynamic speakers (DT-48). The speaker outputs were delivered through custom-made soft plastic (Hal-Hen Perform) earmolds, and a probe tube allowed daily calibration of the closed acoustic system before each session with an Etymotic ER-7C probe microphone system. Calibration tables based on the frequency shaping introduced by the system to wideband (100 Hz – 20 kHz) noises were used to determine the attenuation values required to describe the sound levels in dB SPL (dB re: 20 μ Pa). Monaural (usually contralateral) or diotic stimuli were presented, depending on the properties of each individual unit (see response classification below).

4.2.3 Recording methods

Single-unit extracellular responses were recorded using glass-insulated tungsten microelectrodes (Bullock et al. 1988). Electrode impedances between 10 and 30 M Ω measured at 135 Hz were usually required for the successful isolation and holding of neurons, but measures of impedance were only marginally reliable as predictors of electrode performance. The electrode signal was amplified (Grass Instruments), filtered (700 Hz – 3 kHz), and AC-coupled with a TDT PC1 spike conditioner before being passed to a spike discriminator (TDT SD1) and event timer (TDT ET1). Isolated spike times were recorded with respect to a stimulus-onset-triggered reference with a resolution of ± 10 μ s.

Before lowering the electrode, a topical anesthetic (Lidocaine) was applied to the dura to desensitize the surface. The position of the electrode was set with a stereotaxic system (Edmund), which was mounted on the cylinder affixed to the rabbit's skull. A sharply beveled, sterile, stainless-steel guide tube (23xx gauge) was used to pierce the dura and protect the electrode tip. The guide tube was lowered by hand until its sharpened end was about 2-3 mm from the proximal (dorsal) surface of the IC. From there, the electrode was lowered independent of the guide tube until a unit was isolated. Because of the tonotopic organization of the structure (low frequencies were encountered at shallower depths) and the limited recording time, the distribution of best frequencies (BFs) of the neurons described here was biased toward lower frequencies (94% of the population had a BF < 10 kHz). Electrodes were advanced from outside the double-walled soundproof booth with a hydraulic microdrive (Kopf Instruments, Tujunga CA); stimulus presentation and online data analyses were also controlled from outside the booth.

At the conclusion of the recordings in each rabbit, electrolytic lesions were made in the approximate center of the 3-D coordinates that described the spatial distribution of the population of well-studied neurons from that specific IC. Standard histological techniques were used to confirm that the recording sites were likely within the central nucleus of the IC (ICC). However, the prolonged duration of recording from each IC made it impossible to definitively state that every unit was positioned within the ICC.

4.2.4 Response classification and analysis

Parameters of AM stimuli were designed for each neuron based on its responses to a battery of simpler sounds. Specifically, to study a cell's sensitivity to changes in modulation depth, it was necessary to determine the appropriate binaural configuration, tone carrier frequency, sound-pressure level (SPL), and modulation frequency. This section describes the stimuli and response quantifications used to make those decisions.

To search for driven activity, a 10-dB SPL spectrum level (50-dB SPL rms), 500-ms Gaussian wideband (100-10,000 Hz) noise with 10-ms \cos^2 ramps was presented binaurally every 1.5 s. The interaural time difference (ITD) for each presentation was randomly chosen from a uniformly distributed range from -300 μ s (contralateral ear leading) to +300 μ s (ipsilateral ear leading) in steps of 100 μ s.

Once a unit was encountered and isolated using the search stimulus, its binaural configuration preference (contralateral, ipsilateral, diotic, or silence) was quantified by counting the number of spikes elicited by each configuration in response to a 500-ms noise (or silent interval) presented once per second. The level and bandwidth of the noise was the same as the search stimulus, and each condition was repeated five times. Next, the unit's BF and threshold were estimated by manually controlling the frequency and level of 100-ms pure tones (10-ms \cos^2 ramps) separated by 500-ms inter-stimulus intervals (ISI).

Based on the audio-visually determined estimates of BF and threshold, a response area (RA) was measured at two SPLs, 10 and 40 dB above threshold (Ramachandran et al. 1999), and at 15 log-spaced frequencies from an octave below to an octave above BF. Average rates were measured over the entire duration of the 200-ms tones, which were presented once per second and windowed with 10-ms \cos^2 onset and offset ramps. Usually just one repetition was sufficient to determine the frequency that elicited an excitatory response at the lowest SPL (defined as BF).

All stimuli presented after the RA had BF tone carriers. First, a rate-level function (RLF) was obtained, usually over a 70-dB range starting about 10 dB below the threshold estimate that was determined audio-visually. Ten repetitions per level of each 100-ms tone burst (including 10 ms \cos^2 ramps) were presented with 400 ms ISIs. Rates were measured over the entire 100-ms stimulus presentation window, and a peri-stimulus time histogram (PSTH) was constructed using a bin size of 0.5 ms. From these responses, cells were classified based on their PSTH type (onset, sustained, on+sustained, or other; similar to Krishna and Semple 2000 and Le Beau et al. 1996), RLF shape (monotonic, saturating, nonmonotonic, or other), and mean first-spike latency (FSL) across the 10 repetitions. PSTH type and FSLs were often level-dependent; the SPL used to classify responses was the level used for AM stimulation (below).

Next, 100%-modulated sinusoidally AM (SAM) tone responses were recorded, usually at 15 modulation frequencies log spaced from 2 to 311 Hz. The overall SPL was fixed (i.e. there was no level increment caused by modulation), and chosen to correspond to a level on the ascending portion of the RLF. This convention was followed unless the strongest response to tones was suppression of firing rate below spontaneous activity, in which case an SPL was chosen that clearly elicited such suppression (this occurred in 9 neurons). To accommodate several cycles of low- f_m stimuli, a 2-s BF tone including 50-ms \cos^2 ramps (a common ramp duration in AM psychophysics) was used as the carrier. Modulation was applied for the entire duration of the carrier (including the ramps).

Three repetitions of each stimulus were presented with an ISI of at least 1 s. Time permitting, additional (usually higher) SPLs and f_{ms} were presented.

Four metrics were used to quantify the responses to fully modulated SAM tones. The average firing rate was computed excluding the first 100 ms to avoid onset effects, although there was usually negligible temporal adaptation to AM stimulation. Synchronization or vector strength (VS, Goldberg and Brown 1969) to the modulation period was calculated from period histograms, which were constructed with a fixed number of bins per AM cycle (64). To quantify the amount of modulation in the response with respect to that imposed on the stimulus envelope, modulation gain [$20 \log(2 \cdot VS/m)$] was also computed (e.g. Frisina et al. 1990). Synchronized rate (e.g. Sachs et al. 1983) was defined as the product of vector strength and average rate. Synchrony and phase values were plotted only if the vector strength was significant (Rayleigh statistic > 13.8 ; equivalently, a test of uniformity $p < 0.001$). In addition, at least 5 spikes across all three stimulus repetitions were required before a response was designated as being significantly synchronized. Envelope-locked response descriptions were computed only for the component synchronized to the stimulus f_m [see footnote 1 in Krishna and Semple (2000) for a brief discussion of this issue, and Khanna and Teich (1989) for ANF responses examined at other stimulus-related frequencies]. Quantifications based on average rate (rMTF), synchrony (sMTF), synchronized rate (srMTF), and response phase (pMTF) provided a modulation-frequency focused description of AM responses.

Several aspects of the MTFs were extracted for comparisons across the population, and for making decisions concerning the stimulus parameters to be studied at lower modulation depths. rMTFs, sMTFs, and srMTFs were classified as all-pass, low-pass, band-pass, band-reject, or high-pass (over the range of f_m tested), based on a 70% change criterion in the response above or below the cell's best modulation frequency (BMF, f_m resulting in an excitatory peak in the MTF) or worst modulation frequency (WMF, f_m eliciting the strongest response suppression flanked by excitatory regions). sMTFs and srMTFs were almost exclusively band-pass or low-pass, while rMTFs could take on any of the five shapes (see Results).

Because of time limitations in the unanesthetized preparation and our goal to study a wide range of modulation depths, a single f_m was used to study the response dependence on m (as opposed to obtaining complete MTFs at several depths). Modulation depth functions (MDFs) based on rate, synchrony, synchronized rate, and phase were measured at a stimulus f_m set equal to the frequency at the peak of the srMTF, regardless of whether the srMTF was strictly defined as band-pass or low-pass based on the 70% drop criterion. The srMTF peak was chosen as a compromise between pure rate and pure timing analyses; time permitting, additional MDFs were obtained at other interesting f_{ms} (e.g. a rate-based WMF).

Modulation depths from -35 dB to 0 dB in $20 \log m$ ($0.018 < m < 1$) were tested in 5-dB or 1-dB steps. Other than m , the stimulus parameters were identical to those used in recording the MTF. Rate and synchronization analyses were also broadly similar, except the initial 500 ms of the response was discarded, and the remaining 4.5 s (1.5 s x 3 reps) was separated into 9 500-ms segments when determining a mean and variance of the rate estimate. The 500-ms window was used because it matched that used in much of the AM psychophysical literature. Ignoring the onset at low m was more crucial for avoiding artifacts than with fully-modulated stimuli because a pure-tone onset response

could result in artificially high values of vector strength if the duration of the onset response interacted with the period of the modulating waveform.

To determine neural detection and discrimination thresholds, responses to different stimulus depths were tested for significant differences between one another. Neural rate-based detection threshold (θ_{rate}) was defined as the lowest m that elicited a rate different from the rate in response to the lowest tested depth (paired t -test $p < 0.05$). An additional condition was imposed: responses to depths higher than θ_{rate} were also required to elicit significantly different rates compared to those in response to the standard depth. This requirement rarely changed the resulting thresholds in practice, but it did eliminate the effect of spurious changes in rate caused by motion artifacts. When calculating rate-based neural discrimination (as opposed to the special case of detection) thresholds, the responses to each depth were treated as responses to a standard; the distance between the lowest comparison depth resulting in a significantly different response determined the predicted just-noticeable difference in depth based on the rate of a single cell.

Synchrony-based detection threshold (θ_{sync}) was defined as the lowest depth that resulted in a significant value of vector strength (Rayleigh statistic > 13.8). This criterion is commonly used in physiological studies (e.g. Liang et al. 2001), and almost always resulted in thresholds matching those determined qualitatively by visually inspecting the period histograms at each depth. The statistical criteria employed in the detection and discrimination analyses were intentionally biased to be strict for synchronization measures (Rayleigh statistic $p < 0.001$) and less stringent for the rate-based predictions (t -test $p < .05$).

In addition to the tests of sensitivity to pure-SAM stimuli, neurons were also tested for their ability to represent deterministic (SAM) envelope fluctuations in the presence of a competing stochastic masker modulation. The equation for the stimuli in the masked-detection task is:

$$s(t) = c\{\sin(2\pi f_c t)[1 + m \sin(2\pi f_m t) + M(t)]\},$$

where f_c is the carrier frequency (set equal to the neuron's BF) and $M(t)$ is the masker waveform (zero when measuring absolute thresholds). Masker level was defined in terms of the root-mean-square (rms) of $M(t)$. The compensation factor c was included so the overall power in both intervals was fixed as m and $M(t)$ varied. The response to a range of signal m was measured for three values of the overall depth of the masker: -23, -18 and -13 dB rms. The frequency composition of $M(t)$ was defined in terms of its bandwidth, which was set to $\frac{1}{2}$ of the neuron's srBMF, and center frequency, which was equal to the cell's srBMF. This was chosen to mirror a psychophysical paradigm, which used a single AM frequency (64 Hz) and a masker bandwidth of 32 Hz (Nelson and Carney, in review). The signal AM was varied as it was in the "unmasked" MDFs, but was restricted to depths below -5 dB ($20 \log m$) for masker levels of -23 and -18 dB rms and below -10 dB for a masker level of -13 dB rms to avoid overmodulation (i.e. overall modulation indices greater than 1 were not presented). Data analyses and other stimulus parameters were identical to those used for the unmasked MDFs, and allowed for a description of *masked* neural θ_{rate} and θ_{sync} .

4.2.5 Implementation of the computational model

The structure and implementation details of the phenomenological model tested here were recently described, and its responses quantitatively compared to previously published AM physiological responses (Nelson and Carney, 2004). Briefly, same-frequency inhibitory and excitatory (SFIE) inputs interact with one another to give rise to model neurons that are rate-tuned to f_m . There are three key parameters at the level of the model IC cells that significantly change the model's overall AM response properties. By choosing appropriate values of the time constants associated with the successive low-pass filtering properties of inhibition (τ_{inh}) and excitation (τ_{exc}), the model cell's rBMF can be adjusted to match single-unit recordings (Nelson and Carney, 2004). The relative strength of inhibition with respect to excitation ($S_{INH,IC}$) determines the degree of suppression observed in the SFIE model cell at low and high f_m s (away from BMF). This parameter was not systematically studied in the initial modeling study, but it was crucial to account for the different AM response types that we observed in the current study in groups of neurons with different pure-tone response properties.

4.3 RESULTS

Our population consisted of 198 single units characterized through the MTF and 164 cells held for at least one MDF. These cells were isolated in 4 ICs of 3 animals. All of the neurons in the sample were responsive to pure-tone carrier SAM stimuli, although a small number of isolated cells were not characterized in detail because of poor responses to the short tone bursts used in the search mode or response area measurements. Furthermore, all of the neurons exhibited significant synchrony to at least one modulation rate, in contrast with cortical responses, which are often not phase-locked to the envelope over a range of f_m s similar to the range used here (e.g., Liang et al., 2002).

4.3.1 Population pure tone responses and correlations with 100% SAM responses

The heterogeneity of BF pure-tone responses in the IC is impressive when compared to the responses observed in lower brainstem structures such as the cochlear nucleus (e.g. Blackburn and Sachs, 1989), and remarkable when compared to the highly stereotypical nature of ANF responses (e.g. Kiang et al. 1965). Because of this diversity, any classification scheme is somewhat arbitrary, as the number of potential categories is essentially unlimited and at least partially subjective. We have chosen to use a small number of classifications (4) for both PSTH type and RLF shape, including one all-encompassing "other" category.

Distributions of PSTH type across the population are shown in Fig. 4-1A. Sustained pure-tone responses (without a clear onset component) were the most common PSTH type (43%), while only 13% of the neurons were pure onset responders. Nearly a third of the population (30%) exhibited some combination of an onset and sustained response. The PSTHs of the remaining 14% of the cells did not fall neatly into one of the three other categories. This PSTH group included offset responses (N=2), pauser-buildups (N=6), combined onset and offset responses (N=3), responses with regularly spaced peaks of discharge not related to the stimulus periodicity (choppers, N=2), suppression below spontaneous rate without an excitatory region (N=9), and unusual histograms (N=5). Because of the paucity of pauser response types in our population

(N=6), they were not identified as a separate class (as in Le Beau et al. 1996 and Krishna and Semple 2000).

Many PSTH type classifications were strongly dependent on the tone SPL; the high proportion of pure sustained responses is in part due to the fact that the neurons were grouped based on their responses at a relatively low SPL (usually on the initial ascending portion of the RLF). Typically, pauser-buildup and offset PSTHs emerged only at higher SPLs. The proportion of chopper units was probably underestimated, as the number of repetitions of the short pure tones was insufficient to perform reliable regularity analyses. Responses phase-locked to the carrier frequency were not observed, even in the 7 neurons with BFs < 600 Hz, although more repetitions may have been required to observe significant vector strength to the fine structure.

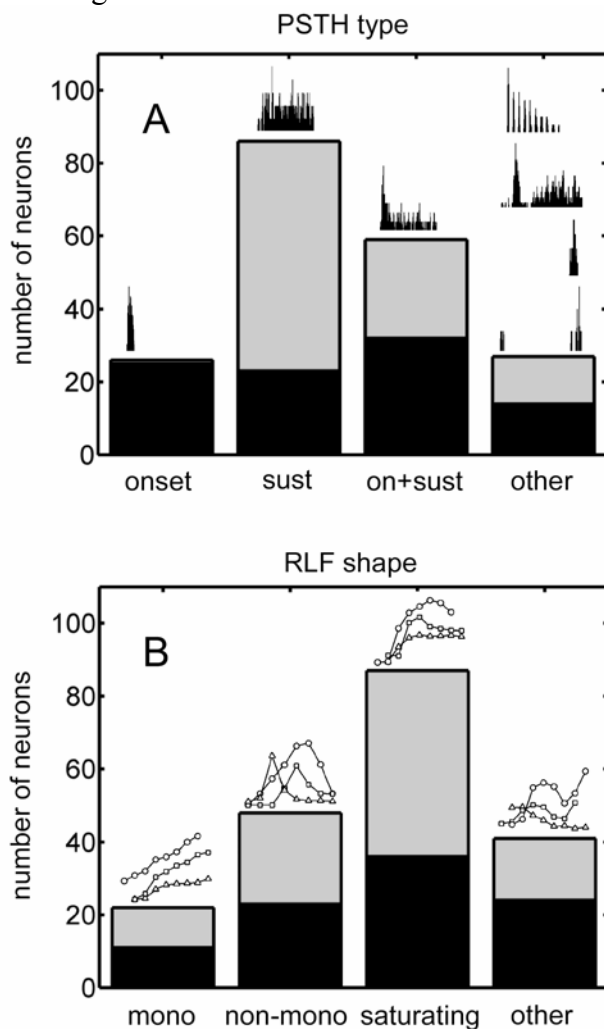


FIG 4-1. Pure-tone response characteristics across the population of 198 neurons. A: Distribution of peri-stimulus time histogram (PSTH) types. B: Distribution of rate-level function (RLF) shapes. Example PSTHs and RLFs are shown above the bars in each category. The dark lower area of each bar represents the neurons that were band-pass rate-tuned to modulation frequency; the rate modulation transfer functions (rMTFs) of the remaining units (gray regions) did not have a single or prominent peak (see Fig. 4-5 for a further breakdown of MTF shapes).

A characterization of the population based on the shapes of single-unit BF RLFs is illustrated in Fig. 4-1B. A 50% rate drop at high SPLs relative to the peak response was required for a RLF to be classified as non-monotonic (as in Aitkin 1991); 24% of the neurons had a single peak and such a rate drop at high SPLs. Some units that met the 50%-drop criterion were placed in the “other” RLF shape category because of multiple peaks in the RLF, or a rebound from an initial rate drop at the highest levels tested (20% of the units). The remaining cells exhibited monotonic (11%) or saturating (44%) RLF shapes over the range of levels tested (almost always 70 dB).

The pure-tone response properties examined in Fig. 4-1 (PSTH type and RLF shape) were broken down further with respect to the corresponding neurons’ fully modulated SAM response properties. Neurons with band-pass (BP) rMTFs (which made up 47% of the population) are shown with the dark portions of the bars in Fig. 4-1; cells with non-BP rMTFs are represented by the light upper segments of each bar. rMTFs in onset pure-tone responders always revealed a band-pass shape (25/26 onset cells were classified as BP over the 2-312 Hz f_m range; one onset neuron had a rBMF of 312 Hz), while most rMTFs in sustained pure-tone responders (63/86) were not BP. Classifications of on+sustained or “other” pure-tone responses were not predictive of the rMTF shape: approximately half of each category was BP tuned. Similarly, RLF shape was not predictive of the presence of a single region of excitation in the rMTF (i.e. a BP shape).

One aspect of BF pure-tone responses that has been shown to be correlated with AM responses in previous studies is the mean FSL: IC neurons with longer FSLs tend to have lower BMFs (Heil et al. 1995, Krishna and Semple 2000, Langner et al. 1987). The current data set corroborates the finding of a weak inverse FSL-BMF correlation (FSL-rBMF Kendall’s $\tau = -0.25$, $p < 0.001$; FSL-sBMF $\tau = -0.26$, $p < 0.001$; FSL-srBMF $\tau = -0.23$, $p < 0.001$). Despite the significant correlations, Fig. 4-2 makes it clear that FSL is in general an unreliable predictor of rBMF (the same is true for srBMF and sBMF). The different symbols in Fig. 4-2 denote the various PSTH types; the three major groups (sustained, onset, and on+sustained) contain neurons with a similar range of rBMFs, but the longest FSLs were found in sustained pure-tone responders. Below the scatter plot in Fig. 4-2 is a histogram of the FSL values across the entire population (only neurons with BP rMTFs were included in the scatter plot). There was not a significant correlation between BMF and BF (Kendall’s $\tau < 0.1$, $p > 0.1$ for rBMF, sBMF, and srBMF correlations with BF) (consistent with Krishna and Semple 2000). It is worth reiterating the fact that the FSLs reported here were based on responses at a relatively low SPL, as opposed to the minimum mean FSL across the entire range of levels tested with pure tones. One might expect that the use of a single low SPL to derive FSLs would result in higher estimates of latency, but this was not strictly true. Some neurons in the IC exhibited an increase in FSL with level, an effect that has been termed the “paradoxical latency shift” (Sullivan 1982). In our population, 16% of the cells revealed such a latency level-dependence, in the form of a mean FSL at a higher SPL greater than one standard deviation higher than the mean latency at the (lower) SPL used for the population analysis.

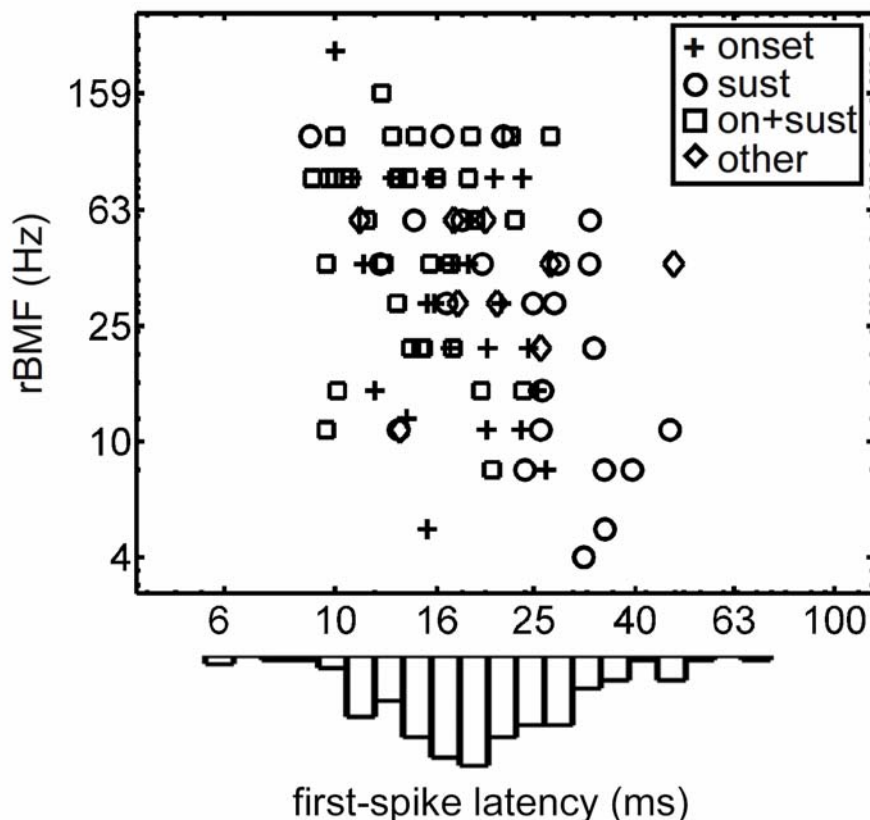


FIG 4-2. Among the neurons with a well-defined rate best modulation frequency (rBMF), cells with a longer first-spike latency (FSL) tended to have a lower rBMF. PSTH types are indicated by the symbols (categories are the same as in Fig. 4-1A). The histogram of FSLs shown below the axis is a description of the entire population with a measurable FSL ($n=182$), not just neurons with rBMFs. The peak histogram value corresponds to a count of 27 neurons.

The range of observed rBMFs was qualitatively similar to the range of sBMFs and srBMFs, but the exact value of BMF could vary considerably across the different response quantifications in a single neuron. Figure 4-3 illustrates these points, with a comparison of BMF histograms for the three metrics (Fig. 4-3A) and scatter plots of the three combinations of rBMF, sBMF, and srBMF (Fig. 4-3B), which are summarized with a correlation coefficient. BMFs were typically between 10 Hz and 100 Hz, although the upper bound of computed BMFs was partially biased by the range of f_m s used (2 – 311 Hz). The correlation between rBMF and sBMF ($R=0.38$) was significant ($p < 0.01$), but considerably lower than the correlations between the srBMFs and the other two metrics (rBMF x srBMF $R = 0.88$, $p < .0001$; sBMF x srBMF $R = 0.58$, $p < .0001$). This is not surprising, since synchrony and rate are, by definition, correlated with synchronized rate. In any case, the even spread of points above and below the diagonal (rBMF = sBMF) in the upper panel of Fig. 4-3B suggests that the relatively low correlation between sBMF and rBMF was not caused by a consistent bias in either direction (i.e. sBMFs were not reliably lower than rBMFs or vice versa).

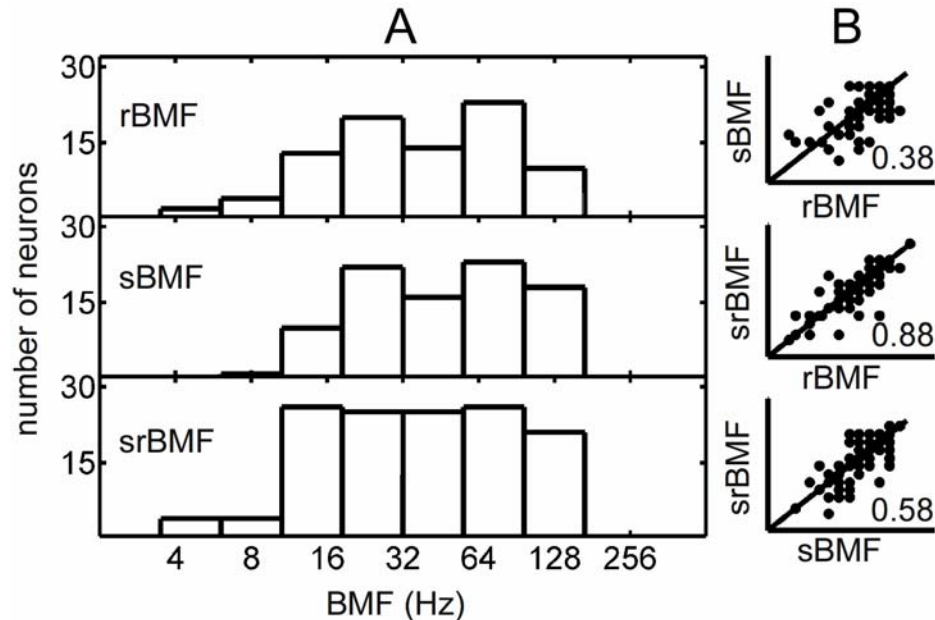


FIG 4-3. Most values of rBMF, synchrony BMF (sBMF), and synchronized rate BMF (srBMF) were between 10 Hz and 128 Hz, but BMFs based on the three quantifications were not always the same in single neurons. A: Distributions of rBMF, sBMF, and srBMF. B: Scatter plots showing the three combinations of BMF based on one quantification versus BMF based on another quantification. Correlation coefficients are included with each scatter plot.

Sharpness of tuning was quantified across the population of BP-tuned rMTFs with a quality factor $Q = \text{BMF}/\text{BW}$, where the bandwidth (BW) was defined as the width of the rMTF at 70% of the peak rate evoked at BMF. The resulting values are plotted as a function of the neuron's rBMF in Fig. 4-4. Most (76%) of the Q -values were less than one, and only three were greater than 2. Onset responders to pure tones (+ symbols in Fig. 4-4) tended to be more sharply tuned (mean $Q = 1.15$) than the other unit types (mean $Q = 0.61$). There was no correlation between the quality factor and rBMF, suggesting a constant relative BW across the range of relevant f_m . Both the overall average value ($Q=0.76$), and the independence of Q on BMF are qualitatively consistent with psychophysical measures of AM-frequency tuning curves and masking patterns (e.g. Houtgast 1989, Ewert and Dau 2000, Ewert et al. 2002, Wojtczak and Viemeister 2005).

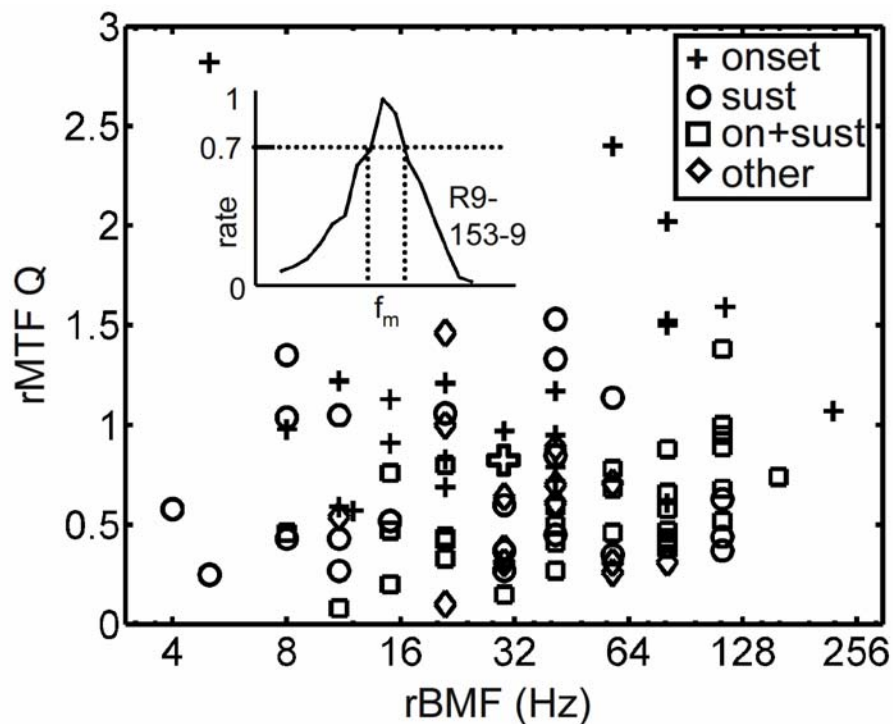


FIG 4-4. Q-values based on rMTFs with clearly defined peaks. Symbols indicate PSTH type as in Fig. 4-2. The inset illustrates calculation of bandwidth at 70% of the peak rate; the open cross shows the corresponding Q and rBMF for the neuron from the inset.

The diversity of MTFs in the IC cannot be fully appreciated by limiting analyses to a determination of the cell's BMF, because many neurons do not have a well-defined BMF (according to the 70% drop criterion both below and above BMF over a range of stimulus f_m from 2 to 312 Hz). Example rMTFs representing the five shapes encountered in the rabbit IC are shown in Fig. 4-5, along with the distribution of the different shapes across the population. Most common were BP-tuned rMTFs, followed by low-pass (LP), high-pass (HP), band-reject (BR), and all-pass (AP) shapes. The scarcity of AP rMTFs in the IC lies in sharp contrast to peripheral (ANF) rMTFs, which are almost exclusively AP (Cooper et al. 1993, Joris and Yin 1992).

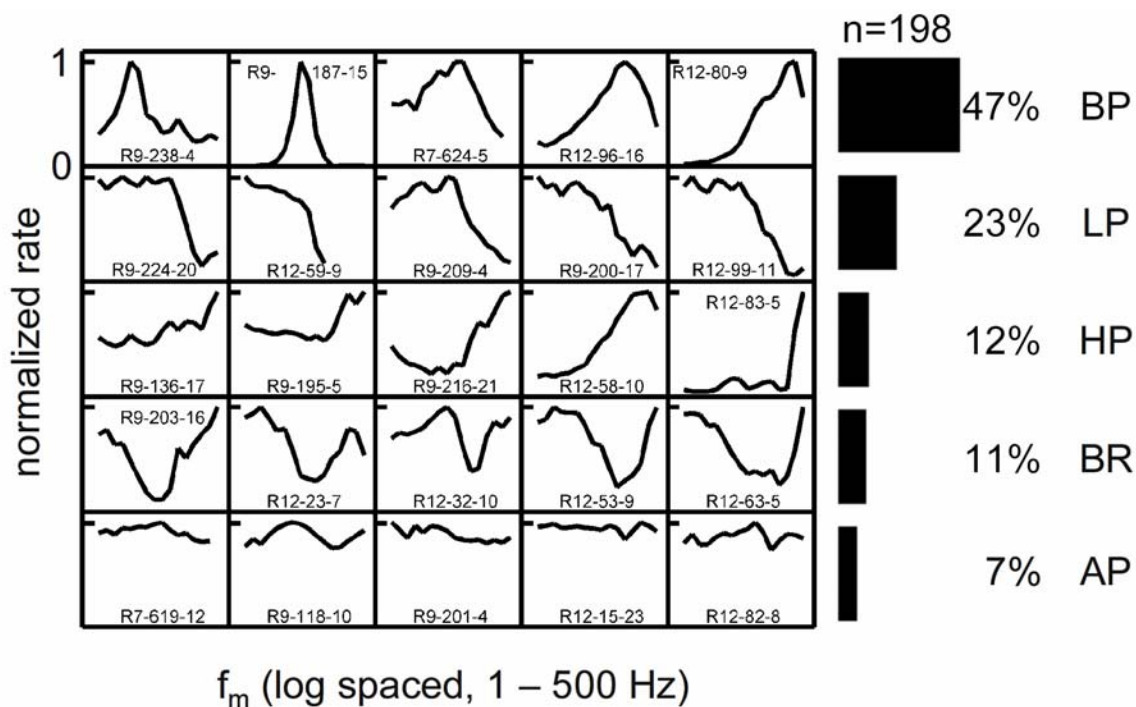


FIG 4-5. Example rMTFs and relative proportions of neurons exhibiting each rMTF shape. Unit identification numbers are included with each rMTF; representative band-pass (BP), low-pass (LP), high-pass (HP), band-reject (BR), and all-pass (AP) neurons are shown in each row.

Because IC neurons are limited in their capacity to follow envelope fluctuations faster than several hundred Hertz (Burger and Pollak 1998, Krishna and Semple 2000, Langner and Schreiner 1988, Rees and Moller 1983), MTFs based on synchrony and synchronized rate were never HP or AP; most sMTFs were either BP (49%) or LP (42%), and 70% of srMTFs were BP. It is worth noting that although synchrony always decreased at high modulation rates, 35% (69/198) of the neurons maintained significant synchrony at the highest f_m tested (usually 311 Hz; limited to 222 Hz in 30/198), and 80% (159/198) were phase-locked to the envelope at $f_m > 100$ Hz. These proportions are comparable to those reported by Krishna and Semple (2000) in the anesthetized gerbil.

4.3.2 Example MTFs and MDFs

Detailed AM responses of four example neurons are highlighted in this section; they were chosen as representatives of each of the four categories of pure-tone responses (sustained, onset + sustained, onset, and other).

4.3.2.1 Representative onset pure-tone response

Without exception, onset units exhibited BP-tuned rMTFs; Fig. 4-6 characterizes such a neuron in more detail. Fully modulated AM responses are shown in Fig. 4-6A, revealing rate tuning to stimulus f_m between 30 and 100 Hz. The mismatch between the peaks in the rMTF and the sMTF resulted in a srBMF (58 Hz) between the rBMF (81 Hz) and the sBMF (41 Hz). The sharpness of rMTF tuning was relatively high in this example ($Q = 1.5$). Raw period and PST histograms are also included in Fig. 4-6B and 4-6C. Two features of the histograms in Fig. 4-6 were consistently observed across

neurons. First, the response phase changed near BMF (the period histograms were constructed starting at a fixed point in time). Second, there was negligible temporal adaptation observed over the 2-s stimulation period in the PSTHs in response to the AM tones.

Stimulus AM depth was varied over a 35-dB range in 5-dB steps to obtain the responses illustrated in Fig. 4-6D. This corresponded to linear modulation depths ranging from 1.8% to 100%. For reference, human tone-carrier AM detection thresholds at medium SPLs and modulation rates below 150 Hz are -20 dB ($m=10\%$) or lower (e.g. Kohlrausch et al. 2000). First, consider the rMDF (the curve with error bars in Fig. 4-6D), which was typical of many neurons in the population in terms of its flat (m -independent) characteristic for depths from -35 dB to -15 dB. Rates that were significantly different from the rate response to the -35 dB AM tone are indicated with open squares. The lowest depth that elicited such a significant change (see Methods) was defined as the neuron's rate threshold ($\theta_{\text{rate}} = -10$ dB).

Plotted alongside the rMDF in Fig. 4-6D is the srMDF (closed circles), which only includes values that were computed with a significant synchrony coefficient. A consistent offset between the rMDF and srMDF indicates a constant value of vector strength across depth. This is confirmed with the sMDF (open circles) shown below the rate- and synchronized rate- MDFs (vector strength values were all between 0.6 and 0.8). Another way to interpret depth-independent response synchrony is in terms of a modulation gain that decreases with increasing m . Single-cell synchrony-based neural thresholds were defined as the lowest modulation depth that evoked a significantly envelope-locked response ($\theta_{\text{sync}} = -10$ dB for the neuron shown in Fig. 4-6). Visual inspection of the period histograms in the Fig. 4-6E suggests that the statistical criteria used to define synchrony threshold were reasonable: response modulation clearly emerges at $m = -10$ dB. Also, the period histograms indicate that phase of the response did not change appreciably as m was varied. As with the PSTHs shown for different modulation frequencies (Fig. 4-6C), there was no evidence for gross, slow temporal adaptation in the PSTHs at different modulation depths (Fig. 4-6F).

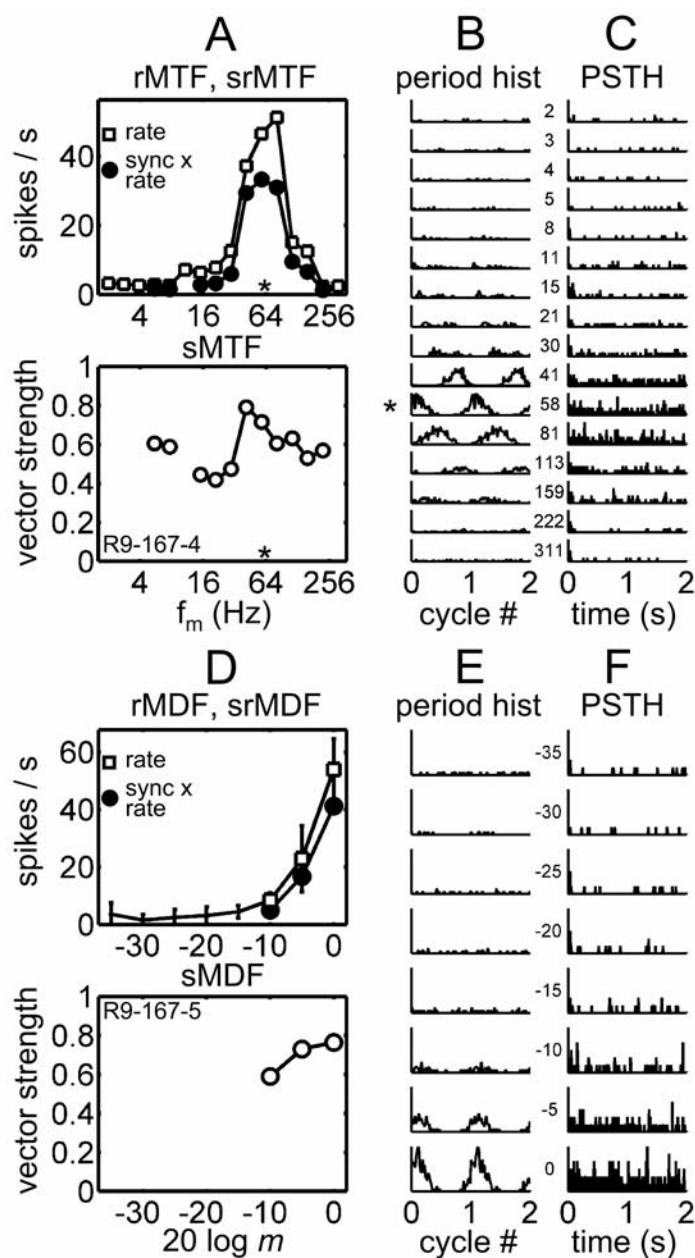


FIG 4-6. AM response properties of a neuron with a pure onset response to best frequency (BF) tones as a function of modulation frequency (f_m , A-C) and modulation depth (m , D-F). A: MTFs based on rate (rMTF), synchrony (sMTF), and synchronized rate (srMTF). Asterisks (*) correspond to the srBMF (and the f_m used in D-F). Quantifications including synchrony are only plotted for significant values of vector strength. B: Period histograms for each f_m represented in the MTF. Spikes were binned into one period of the modulation, and the resulting histogram was plotted twice for clarity. C: PSTHs for each tested f_m . D: Modulation depth functions (MDFs), again based on rate, synchrony, and synchronized rate. Rate responses plotted with open squares (\square) were significantly different ($p < 0.05$) from the rate response to the lowest tested m (-35 dB). E and F: period histograms and PSTHs, respectively, for each tested m . SPL = 20 dB, $f_c = 2000$ Hz.

4.3.2.2 Representative on+sustained pure-tone response

Several aspects of the AM responses of the on+sustained pure-tone responder shown in Fig. 4-7 are fundamentally different from those of the pure onset responder described above. Perhaps the most salient difference was that the entire range of 100%-AM stimuli elicited synchronized firing (from 2 Hz to 311 Hz). The resulting rMTF reveals weaker tuning ($Q = 0.81$), with a peak at 81 Hz (the same rBMF as the representative onset neuron). Vector strength reached a maximum of 0.56 at 113 Hz, and synchronized rate peaked at 58 Hz. Period histograms at low f_m s show that the probability of firing remained relatively constant for a longer portion of the corresponding stimulus waveform than the onset neuron, with a weak cycle-by-cycle onset adaptation component. The shape of the period histogram at the srBMF (58 Hz) was somewhat more complex, with two peaks near the onset of adaptation of the response during each cycle. This multimodal period histogram shape emerged only at higher modulation depths (Fig. 4-7E). PSTHs plotted in Fig. 4-7C and Fig. 4-7F again suggest minimal adaptation over a time scale on the order of hundreds of ms.

The rMDF measured at 58 Hz (Fig. 4-7D) shows that the peak in the rMTF became apparent only at high modulation depths ($\theta_{\text{rate}} = 0$ dB). In contrast, synchronization to the period of the modulating waveform was significant at -20 dB, and synchrony increased monotonically with increasing m . Correspondingly, the period histograms were modulated, and the timing of spikes became more phase-locked to a particular phase of the envelope as m was varied between -20 dB and 0 dB (Fig. 4-7E).

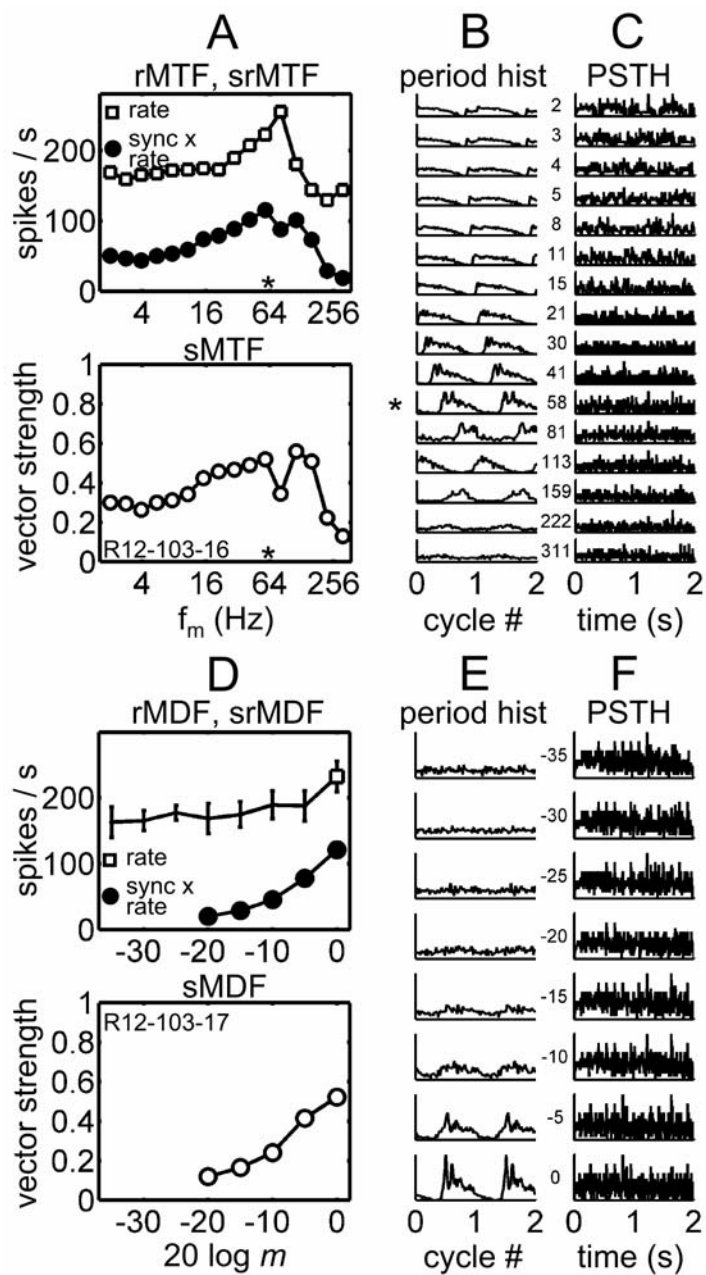


FIG 4-7. AM responses of a representative neuron with an onset + sustained BF tone PSTH. Format is the same as Fig. 4-6. SPL = 40 dB, $f_c = 3900$ Hz.

4.3.2.3 Representative sustained pure-tone response

Pure sustained responders to short tone bursts were usually associated with rMTFs that were not BP (Fig. 4-1A). An example of such a cell is described in terms of its MTFs and MDFs in Fig. 4-8. The main feature of the rMTF of Fig. 4-8A was the presence of a broadly tuned suppressive region (i.e. a BR rMTF); a complementary (BP) sMTF had a peak within the region of rate suppression. Taking the product of synchrony and rate resulted in a srMTF with both a region of suppression at lower f_{ms} , an excitatory region at higher f_{ms} , and a srBMF that matched the sBMF of 113 Hz (Fig. 4-8A). The period histograms in Fig. 4-8B indicate that the drop in rate was largely mediated by a suppression of firing after the onset response elicited in each cycle, and that the rate recovery at higher f_{ms} was not strongly synchronized (although VS remained significant up to 311 Hz).

Because the location of the srBMF was within the region of rate suppression, increasing the stimulus m in the MDF protocol resulted in a drop in average rate. This rate drop is quantified in the rMDF of Fig. 4-8D. In terms of information-carrying capacity, a rate code that is based on a decrease in firing rate is just as effective as one that requires an increase in rate as the stimulus depth is increased. However, the cell's θ_{rate} of -10 dB based on a drop in rate was still much higher than human AM-detection abilities. As with the first two example neurons, a timing-based metric such as synchrony was more sensitive to low-depth AM stimulation: the cell's θ_{sync} was -25 dB. Also, the value of VS increased with depth; this trend was mainly observed in neurons that responded to pure tones with a substantial sustained rate (see Fig. 4-13B).

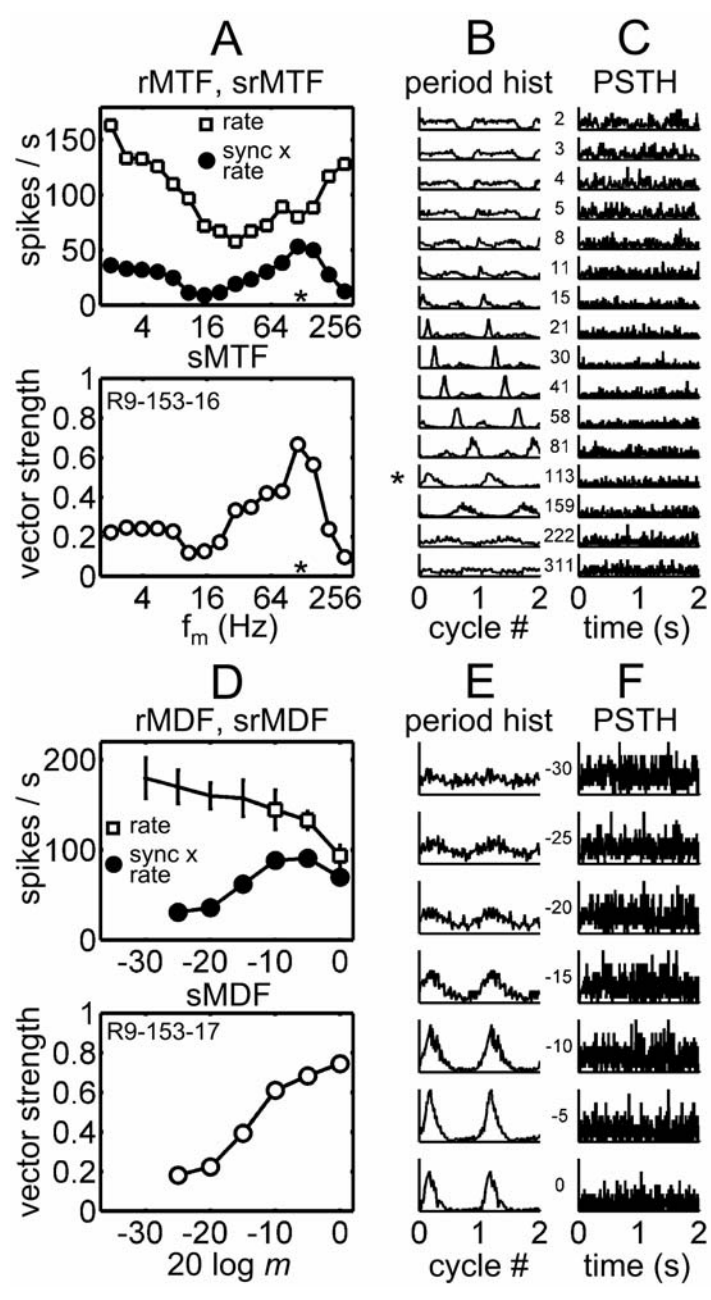


FIG 4-8. AM responses of a representative neuron with a sustained BF tone PSTH. Format matched to Figs. 4-6 and 4-7. SPL = 20 dB, f_c = 3900 Hz.

4.3.2.4 Representative inhibited pure-tone response

Nine neurons in the population did not exhibit a clear excitatory response at any tested frequency or SPL using pure tones [a small island of excitation may have potentially been revealed had we used finer SPL-resolution; these neurons probably belong to the Type O RA category as defined in Ramachandran et al. (1999)]. The BFs of the units belonging to this subset were defined accordingly as the tone frequency that evoked an inhibitory response (below spontaneous rate) at the lowest effective SPL. Figure 4-9 illustrates that the responses of such neurons can be highly dependent on the modulation properties of the stimulus. All three of the MTFs shown in Fig. 4-9A were LP in shape; rate and synchronized rate dropped to 70% of their peak values at about 65 Hz, while synchrony was at least 70% of its peak value ($VS_{\max} = 0.8$) up to 107 Hz.

The characteristic phase [the low- f_m y-intercept of the phase-MTF (not shown)] was approximately 180° out of phase with respect to those observed in the neurons illustrated in Figs. 4-6, 4-7, and 4-8. This can be qualitatively verified by comparing the period histograms at the lowest tested f_{ms} across the four example neurons, and suggests that the neuron was released from inhibition at times corresponding to the valleys of the modulating waveform. Also in contrast to the first three examples, a weak form of slow adaptation was observed in the PSTHs: onset inhibition was slowly released over a time course of about 1s (this is most clearly illustrated by the 113-Hz PSTH). A similar slow release from inhibition was observed in several other neurons, but was usually found only in response to stimulus f_{ms} above the cell's primary excitatory or suppressive regions.

Since there was not a clear peak in the srMTF of Fig. 4-9A, a modulation of 40 Hz was chosen for the MDF simply as an AM frequency within the pass-band of all three MTFs. Interestingly, based on the MDFs alone, the neurons characterized in Fig. 4-6D and Fig. 4-9D were essentially indistinguishable (despite their obvious differences in pure-tone responses). The rMDF revealed an increase in firing rate with increasing depth (and longer effective times of release from inhibition), with a θ_{rate} of -10 dB (Fig. 4-9D). The synchrony-based threshold was, once again, lower than the rate threshold ($\theta_{\text{sync}} = -15$ dB). Although VS values were similar for $-15 \text{ dB} \leq m \leq 5 \text{ dB}$, the corresponding period histograms were quite different. This illustrates one of the limitations of the synchronization coefficient alone as a general description of temporal response characteristics.

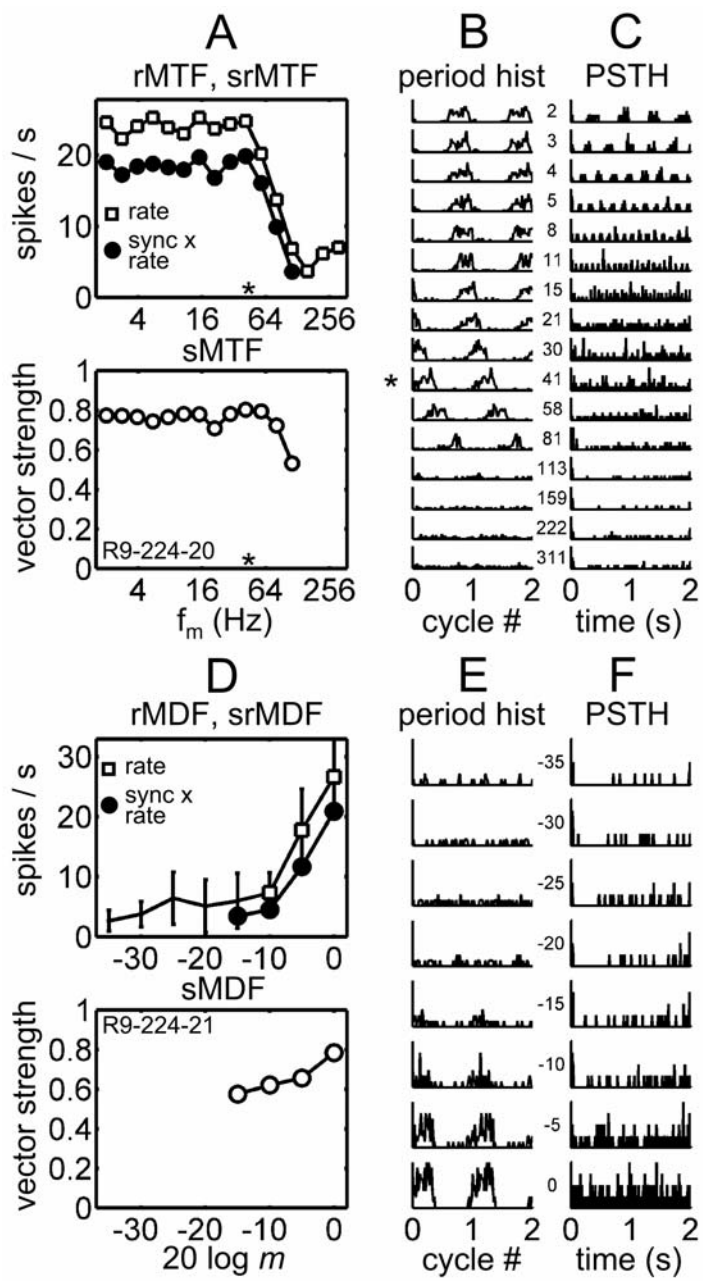


FIG 4-9. AM responses of a representative neuron that responded with a decrease in rate when stimulated using pure tones. Format is the same as Figs. 4-6, 4-7, and 4-8. SPL = 10 dB, $f_c = 12885$ Hz.

4.3.3 Single-unit rate- and synchrony-based AM thresholds

Performance in three basic psychoacoustic AM tasks was predicted based on changes in neural responses at different modulation depths that were quantified in terms of average rate and synchrony. The three psychophysical paradigms are: (1) pure SAM detection, (2) masked SAM detection, and (3) SAM depth discrimination.

4.3.3.1 Pure SAM detection

Human listeners can discriminate the difference between a pure tone and a SAM tone at modulation depths lower than -30 dB (e.g. Zwicker 1952). At a given SPL, performance does not systematically depend on f_m or f_c for modulation frequencies between 10 Hz and 150 Hz and carrier frequencies above 1000 Hz (Kohlrausch et al. 2000). Sensitivity to SAM is best at higher SPLs, but thresholds can remain lower than -20 dB at low sensation levels (Kohlrausch et al. 2000). There is indirect psychophysical evidence suggesting that listeners probably use audio-frequency channels other than that of the carrier to perform the task at high SPLs, where the effective level is lower and peripheral saturation and compression are less likely to have a strong influence on performance (Kohlrausch et al. 2000, Ruggero et al. 1997).

Neural rate and synchrony SAM detection thresholds across the population of 164 neurons are shown in Fig. 4-10A as a function of the stimulus modulation rate.

Consistent with the example neurons described in the previous section, the vast majority of rate thresholds (open circles) were -10 dB or higher (139/164 neurons). If the rate did not change across the entire range of m , thresholds were deemed immeasurable; this group of neurons is identified with the X on the axes in Fig. 4-10. Five neurons had a rate threshold of -20 dB, and 20 responded with a significant change in rate at -15 dB. The histogram of rate thresholds to the left of Fig. 4-10A reinforces the fact that rate changes in single IC neurons were, in general, poor predictors of human SAM sensitivity. There was not a strong relationship between rate thresholds and the f_m of stimulation, which was set equal to the most prominent peak in the cell's srMTF.

Synchrony-based thresholds were more evenly distributed across the perceptually relevant dynamic range (shown in the histogram to the right of Fig. 4-10A) than the values of θ_{rate} . Twenty-eight percent (46/164) of the neurons had synchrony thresholds of -20 dB or lower; three units were significantly phase-locked at a modulation depth of -30 dB. Examination of the individual neural thresholds (x symbols in Fig. 4-10A) reveals no obvious trends either in maximum sensitivity or threshold distribution as a function of the stimulus f_m .

Figure 4-10B shows a feature of the data that is suggested by but not explicitly contained in Fig. 4-10A: θ_{sync} was almost always lower than θ_{rate} on a neuron-by-neuron basis. In the scatter plot of Fig. 4-10B, this aspect of the data takes the form of almost all of the points lying above the diagonal. The three neurons that had the most sensitive synchrony thresholds ($\theta_{sync} = -30$ dB, the three left-most points in Fig. 4-10B) had corresponding rate thresholds of -15 dB, -5 dB, and one immeasurable θ_{rate} .

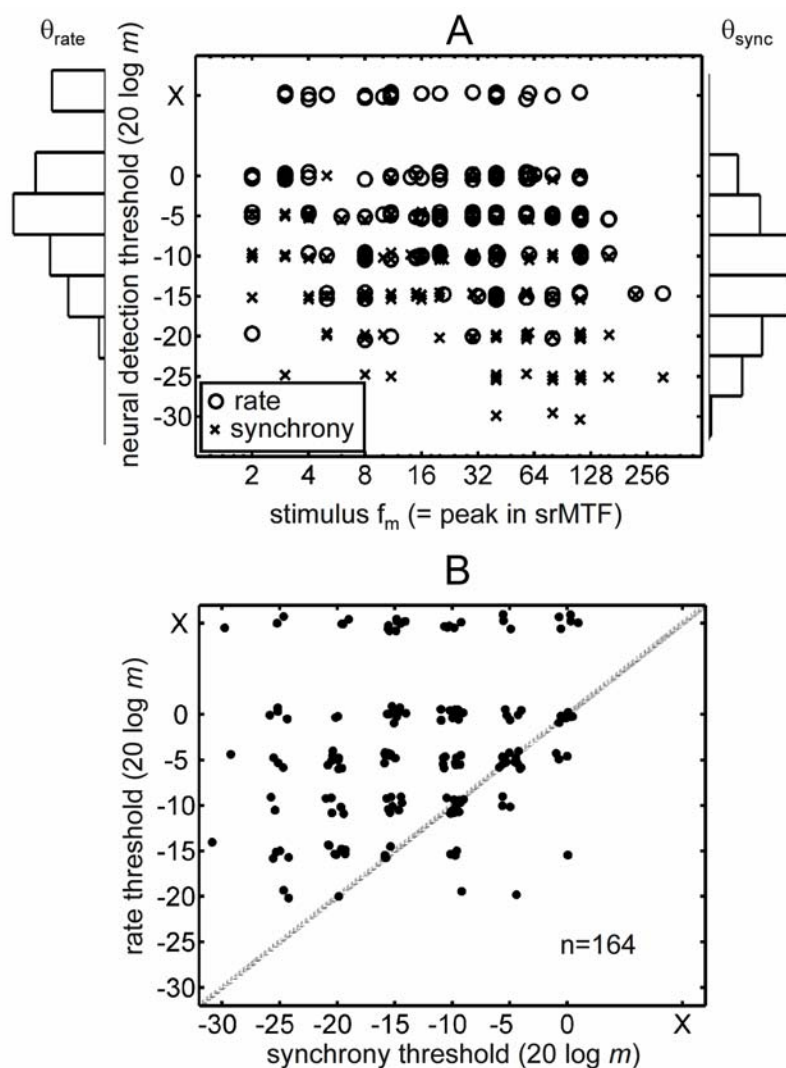


FIG 4-10. Neural AM detection thresholds based on rate and synchrony. Axis labels marked with an X indicate that thresholds were immeasurable, and that average rate was depth-independent over the entire dynamic range (all neurons responded with significant synchrony at $m = 0$ dB). A: The lowest AM depth that elicited a significant change in rate (open circles) or a significant value of vector strength (crosses) for each of the 164 tested neurons. Histograms framing the y-axes in A show the distribution of rate thresholds (left axis; peak histogram value = 47 neurons) and synchrony thresholds (right axis; peak histogram value = 39 neurons). B: Rate threshold versus synchrony threshold scatter plot. Points located above the diagonal of equality in the scatter plot indicate that the cell's rate threshold was higher than its synchrony threshold. In both A and B, a uniformly distributed random number between -1 and 1 was added to the value of threshold to facilitate visualization of the responses of all of the neurons. For reference, human tone-carrier AM detection thresholds at supra-threshold SPLs and modulation frequencies below approximately 120 Hz are between -20 dB and -30 dB (e.g. Kohlrausch et al. 2000).

As mentioned above, overall SPL can affect behavioral SAM detection thresholds. The neural responses summarized in Figs. 4-2 through 4-10 were obtained at an SPL chosen based on individual RLFs: the SPL used for AM stimulation was set to a level on the ascending portion of the RLF, or at its peak if it was sharply non-monotonic. One obvious question is whether the relatively poor neural rate-based thresholds might improve if a higher SPL were chosen. In 33 neurons, this question was addressed by re-measuring MTFs and MDFs at a level typically 20-40 dB higher than that used for the low-SPL responses. The stimulus f_m for each MDF was determined based on the peak in the srMTF, which could vary across SPL (for a more detailed discussion of the level-dependence of MTF shapes, see Krishna and Semple 2000).

The resulting rate- and synchrony-based thresholds for this subset of the population are shown for both tested SPLs in Fig. 4-11. Thresholds that increased (i.e. sensitivity got worse) or remained the same with increasing SPL are plotted with solid lines; the remainder of the neurons, which exhibited a decrease in threshold at the higher SPL, are represented by dashed lines. The majority of rate thresholds (22/33) did not improve at high SPLs (Fig. 4-11A), and 11 of the comparisons revealed an increase in θ_{rate} at higher SPLs. The 11 cells that did exhibit an improvement in rate-based sensitivity with level still did not approach human detection thresholds at comparable SPLs (the lowest θ_{rate} above 40 dB SPL was $20 \log m = -15$ dB). Synchrony-based thresholds (Fig. 4-11B) were also more likely to show a decrease in sensitivity at high SPLs (16/33) than an improvement (11/33). The average θ_{rate} increased by 0.9 dB at the higher SPL, and the average θ_{sync} increased by 1.1 dB at the higher level. Overall, the trends illustrated in Fig. 4-11 suggest that the use of a relatively low SPL in the population analysis (e.g. Fig. 4-10) probably did not bias the results toward higher thresholds.

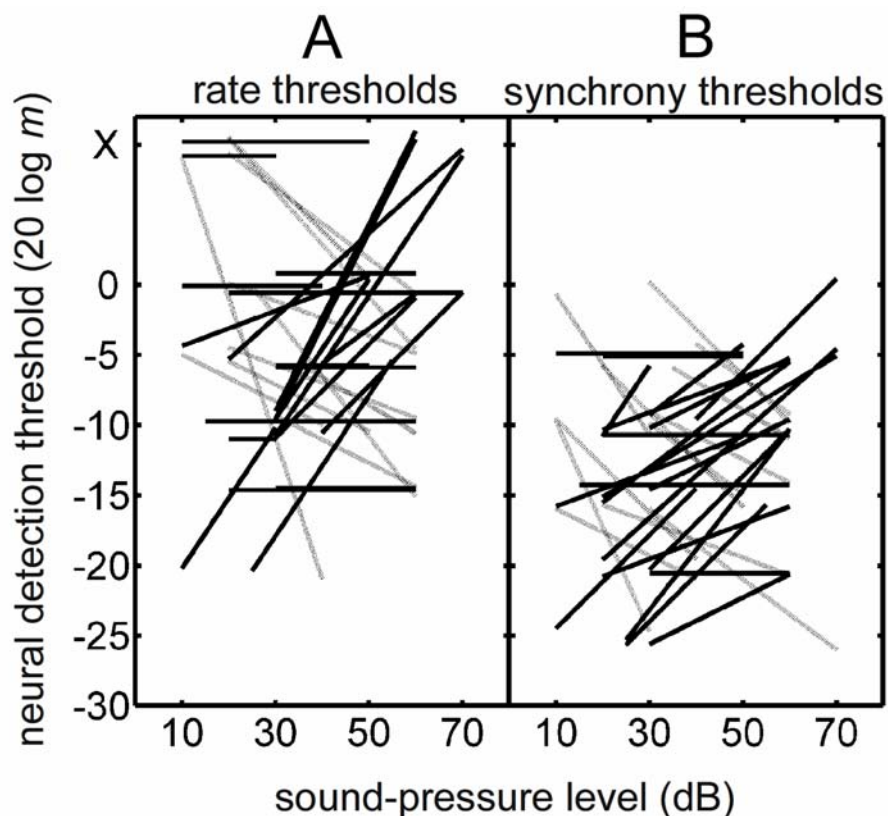


FIG 4-11. Increasing SPL does not systematically improve AM sensitivity in IC neurons. A: Rate-based thresholds in 33 neurons as a function of the SPL of stimulation. B: Synchrony-based thresholds in the same population. In both A and B, neurons with thresholds that were level-independent or increasing with SPL are depicted with solid black lines (22/33 for both rate and synchrony), and sensitivities that improved with level are plotted with the lighter gray lines. As in Fig. 4-10, thresholds for each neuron were plotted with a small amount of jitter added so that all of the lines were visible in the figure.

We conclude that, for pure SAM detection, some temporal aspect of the envelope-locked response (i.e. synchrony) at the level of the IC must be taken into consideration to account for psychophysical detection thresholds based on the responses of single neurons.

4.3.3.2 Masked SAM detection

Most naturally occurring sounds have complex modulation spectra; to assess the generality of the conclusions made in the preceding section, a similar rate- and synchrony-based analysis was applied to responses elicited by a SAM signal modulation embedded in a competing masker modulation. Psychophysical experiments using similar stimuli have shown effects of the frequency relationship between masker and signal (Bacon and Grantham 1989, Ewert and Dau 2000, Ewert et al. 2002, Houtgast 1989, Strickland and Viemeister 1996) and of the “level,” or modulation depth of the masker (Bacon and Grantham 1989, Nelson and Carney, in review, Strickland and Viemeister 1996). Because a main focus of the current set of experiments was to establish the modulation-depth dependence of responses in the IC, neural masked thresholds were

determined with a narrowband Gaussian masker (centered on the sinusoidal signal frequency) at several masker depths.

Over a 10-dB range of masker depths from -23 dB rms to -13 dB rms, psychophysical SAM detection thresholds in a similar task (with $f_m = 64$ Hz, masker BW = 32 Hz, $f_c = 5500$ Hz, and SPL = 65 dB) increased monotonically as the masker fluctuations became stronger (more details and thresholds in a wider range of stimulus conditions can be found in Nelson and Carney, in review). The behavioral thresholds are shown in Fig. 4-12 along with neural thresholds obtained in 28 units, again based on rate and synchrony to the signal envelope frequency. The signal f_m was set to the peak in the cell's srBMF, and the masker bandwidth was fixed at half of the SAM signal frequency. To avoid over-modulation (i.e. a modulation depth greater than 1), the signal depth was restricted to values ≤ -5 dB for masker depths of -23 and -18 dB rms, and to values ≤ -10 dB for the -13 dB rms masker. As a result, there were no predicted neural thresholds at 0 dB for any masker level, and no thresholds at -5 or 0 dB ($20 \log m$) for the -13 dB rms condition.

Figure 4-12A shows that, in general, average rates were more successful in predicting masked thresholds than they were at predicting pure SAM detection (i.e. Fig. 4-10). A small group of cells (8) exhibited rate thresholds within approximately 5 dB of the listeners' data at one or more masker depths. This is weakly suggestive that the functional contributions of rate and synchrony may depend on the range of relevant modulation depths in a given task. We will come back to this idea in the following section, which considers a general discrimination task across the entire perceptual modulation depth dynamic range.

Synchronization to the signal SAM was significant in some neurons at depths even lower than the behavioral thresholds, most notably at the highest tested masker modulation depth, where 7/24 neural thresholds were below the psychophysical data (Fig. 4-12B). As with the pure SAM detection population analysis, the distribution of synchrony thresholds was more uniform than the rate-based distribution (which was skewed towards higher or immeasurable predictions of threshold). Trends in threshold across masker depth for individual neurons are not shown in Fig. 4-12 for clarity, since there was not a consistently observed increase or decrease in predicted sensitivity with increasing masker level (as there was in the psychophysical data).

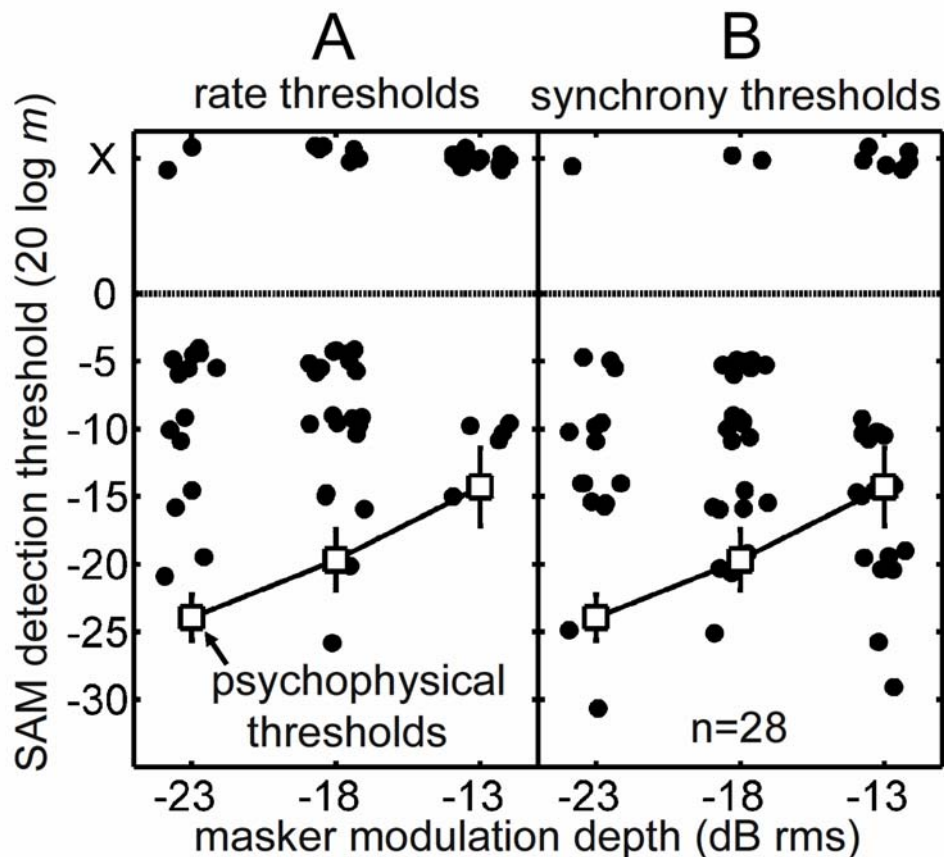


FIG 4-12. Neural AM detection thresholds in the presence of a competing masker modulation. Thresholds are plotted as in Fig. 4-10A, but some neurons are represented by more than one point in these panels, if they were tested at several masker depths. Human psychophysical thresholds are plotted for reference in both panels. A: Rate thresholds. B: Synchrony thresholds. See text for details concerning the lack of thresholds at 0 dB (and -5 dB for the -13 dB rms masker modulation depth).

4.3.3.3 SAM depth discrimination

Another fundamental measure of envelope processing in psychoacoustics is SAM depth discrimination, which describes the ability of the system to resolve small changes in m . Pure SAM detection is a special case of this more general paradigm: the standard depth (m_s) for detection is set to 0, and the comparison depth (m_c) is adjusted until it is just noticeably different from the standard interval. The same procedure can be repeated for any value of m_s . Psychoacoustic measurements of pure-tone carrier SAM-depth discrimination reveal thresholds that are approximately a constant fraction of m_s for supra-threshold standard depths (e.g. Ewert and Dau, 2004). In other words, Weber's law is a reasonable first approximation to the data; a 1- or 2-dB difference in SAM depth is required to discriminate m_s and m_c , over a range of standard depths from about -23 dB to -3 dB. Performance worsens at very low standard depths, where the task effectively becomes one of detection (i.e. the modulation is sub-threshold in the standard interval and slightly above threshold in the comparison interval; Ewert and Dau 2004).

To determine whether SAM depth discrimination performance based on the responses of single IC neurons was comparable to perceptual thresholds, it was necessary

to use higher resolution along the modulation depth axis than was used for the detection paradigms. In 20 neurons, a high-resolution MDF was obtained with step sizes smaller than behavioral jnds (1 dB in 19/20, 2 dB in 1/20, Fig. 4-13B). The entire 35-dB range was not usually sampled; instead, the initial coarse MDF determined the range of depths that elicited a change in rate, and the 1-dB step size analysis was limited to that range. Neural threshold was defined for a range of m_s by treating the mean and SD of each rate response as the response to a standard depth, and searching for the lowest comparison depth that elicited a significantly different (t -test $p < 0.05$) rate response. The neural m_c at threshold was required to be higher than m_s ; further, all of the rate responses above threshold were required to be significantly different from the standard response.

Psychophysical SAM depth-discrimination thresholds are plotted in Fig. 4-13A (connected open circles, from Ewert and Dau 2004), along with the predictions based on significant changes in average rate in the 20 neurons (black lines without symbols). Perhaps the most salient aspect of the neural predictions is the noisiness of the functions; this reflects the relatively lenient statistical criteria that were used to identify significant changes in rate. Nonetheless, 17/19 neurons exhibited rate-based discrimination thresholds as low as 1 dB (the lowest possible value, given the 1-dB step-size), which corresponds to a Weber fraction ($10 \log(m_c^2 - m_s^2) / m_s^2$) of approximately -6 dB. All of the 1-dB depth-discrimination threshold predictions occurred for standard depths higher than -18 dB. This is consistent with the finding that all of the rate thresholds for AM detection were -20 dB or higher.

Figure 4-13B contains the raw rMDFs (gray lines with error bars) corresponding to the predicted performance shown in Fig. 4-13A. The rMDFs were all monotonic (one was monotonically decreasing), and 12/20 neurons showed no evidence for rate saturation at high AM depths. There was no obvious dependence of rate variability on the average rate, as observed in a Poisson process (this point is addressed in more detail the Discussion). Also included in Fig. 4-13B are sMDFs computed from the same responses (black lines). In contrast with the rMDFs, 17/20 sMDFs saturated at depths near 100% modulation. This suggests that rate may be a more reliable single-neuron predictor of depth discrimination performance at high standard depths.

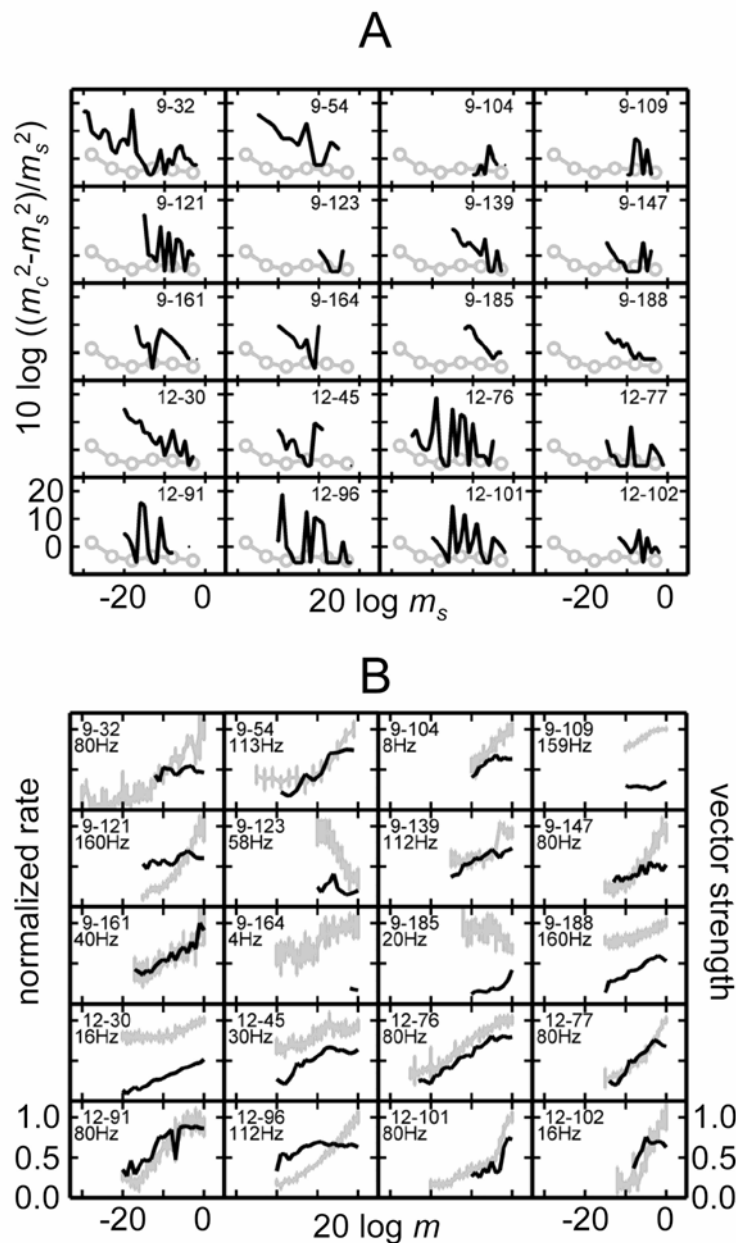


FIG 4-13. Changes in average rate can account for psychophysical SAM depth discrimination, but only at high standard depths. **A:** Comparison of neural predictions of depth discrimination performance in 20 neurons (black lines) and human psychophysical thresholds for the same task (gray connected circles, from Ewert and Dau 2004). **B:** Raw rate MDFs (gray lines with error bars) and synchrony MDFs (black lines) for the same neurons examined in A.

4.3.4 Extending a computational model

The prevalence of sustained pure-tone responses in the unanesthetized rabbit IC, which were often associated with weak rate-tuning to AM (Fig. 4-1A), calls into question the generality of a simple phenomenological model that was developed to simulate pure onset tone responses and strong tuning to f_m in the IC (Nelson and Carney 2004). In this

section, we show that the adjustment of a single model parameter can give rise to model cells with a wide range of pure-tone and AM response properties. The effects of this parameter were not examined in the original modeling study.

Effectively a temporal filter, the IC stage of the same-frequency inhibition and excitation (SFIE) model assumes the convergence of inhibitory and excitatory inputs that have low-pass or weakly band-pass sMTFs and all-pass rMTFs (typical of AN fibers and many cell types in the VCN). By assuming sluggish inhibition (a lower sMTF corner frequency) with respect to excitation, the final output of the model is rate-tuned to a modulation frequency that depends on the assumed time constants of the membrane and synaptic low-pass filtering effects. The relative strength of inhibition at the level of the IC model cells ($S_{\text{INH,IC}}$) was fixed at 1.5 in Nelson and Carney (2004) but was varied here.

Predicted AM responses of the SFIE model are shown in Fig. 4-14 along with the responses of two example neurons, chosen to represent the two broad classes of cells with a steady-state response to pure tones and those without a steady-state response (i.e. onset responders). Actual data are plotted with black lines throughout Fig. 4-14; model predictions for matched stimulus conditions are shown with gray lines. In the rMTFs and rMDFs shown in Fig. 4-14A and D, several model predictions are shown; $S_{\text{INH,IC}}$ was the parameter varied to obtain the various simulated results (the value is indicated next to each function).

The rMTF of the on+sustained neuron depicted in Fig. 4-14A has a peak around 20 Hz, but does not drop to near zero at lower and higher f_m . In general, higher inhibitory strengths in the model resulted in stronger rate-tuning and higher rate-based AM detection thresholds. Comparing the simulated rMTFs revealed that a shape similar to the data could be obtained in the model when $S_{\text{INH,IC}}$ was set to 0.95. Allowing this parameter to be less than one also resulted in a sustained portion of the model response to pure tones. When modulation depth was varied (for a fixed f_m of 20 Hz), the value of $S_{\text{INH,IC}}$ that provided the best match to the data was again 0.95. Corresponding period histograms for the data and the version of the model that best fit the rate functions are shown Figs. 4-14B and 4-14C. As a function of f_m (Fig. 4-14B), the model histograms are qualitatively similar to the data, although the relative phase changes faster in the data near rBMF (20 Hz) than in the model. Figure 4-14C illustrates the period histogram dependence on modulation depth; in both data and model, there was a response at the lowest tested m (i.e. a sustained response to an effectively unmodulated pure tone) and synchronization to the envelope emerged at modulation depths below the rate-based AM threshold.

A similar comparison between data and model for a pure-tone onset responder with a rBMF of 80 Hz is shown in Fig. 4-14D-F. To accurately predict the negligible responses at low and high f_m , and at low m , it was necessary to assume inhibition whose overall strength was greater than that of the excitatory input. A reasonable fit to both the rMTF and rMDF was achieved when $S_{\text{INH,IC}}$ was set equal to 1.1 (Fig. 4-14D), and the period histograms suggested that the model was also capable of predicting the temporal aspects of the responses of onset neurons in the IC. Note the lack of a consistent sustained response at modulation depths below -15 dB in both the data and model.

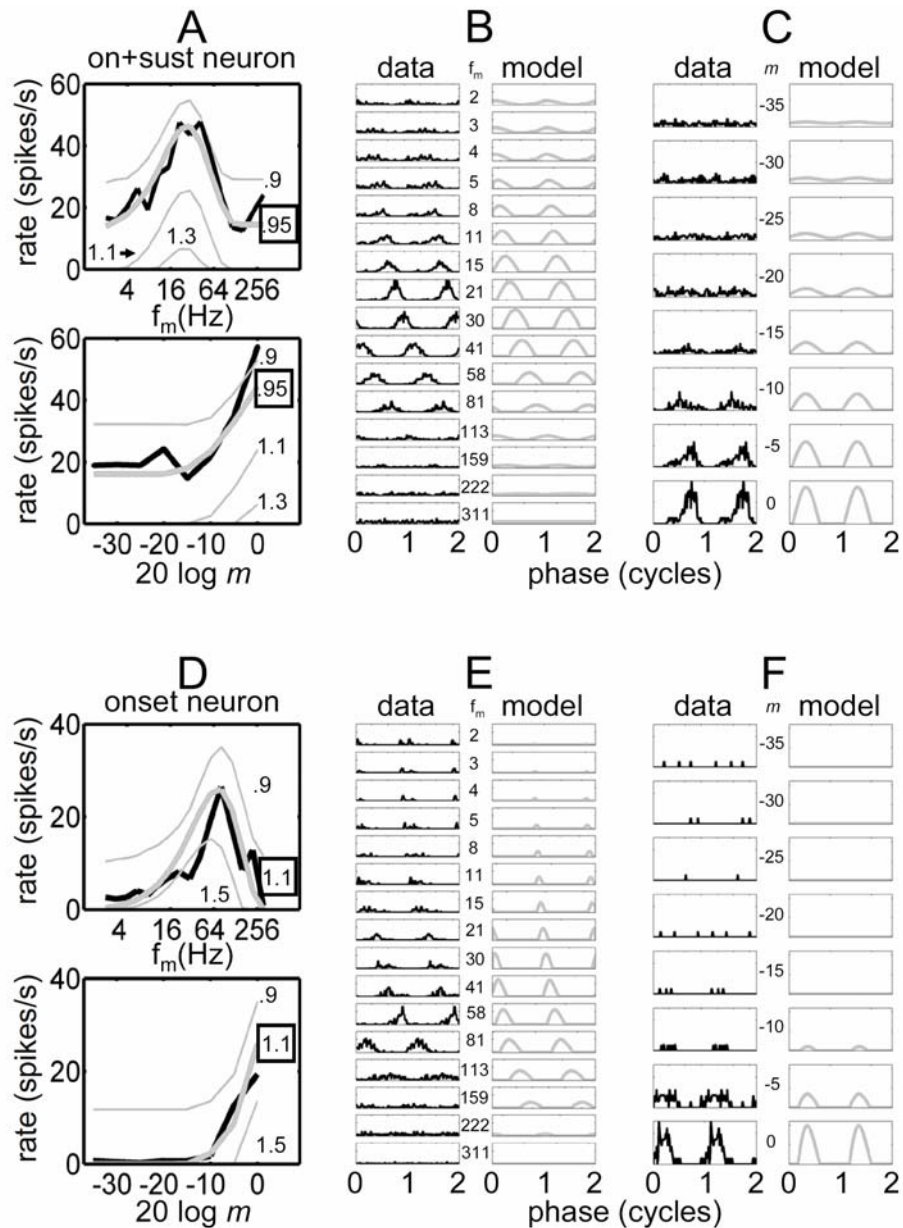


FIG 4-14. Comparison of SFIE model (gray lines) and neural data (black lines). A-C: onset + sustained responder to unmodulated tones (R9-179B); D-F: pure onset responder to unmodulated tones (R12-86A). A: rMTFs and rMDFs for the representative IC neuron and different implementations of the SFIE model. The model parameter governing the relative strength of excitation with respect to inhibition ($S_{INH,IC}$) is indicated next to the corresponding model rMTF or rMDF. B: Period histograms (as in Fig. 4-6) for data and model ($S_{INH,IC} = 0.95$) and for each tested f_m . C: Period histograms for each tested m . Other model parameters: $\tau_{exc} = 3.5$ ms, $\tau_{inh} = 6$ ms, rate scalar = 2.5. Stimulus parameters: SPL = 20 dB, $f_c = 800$ Hz, f_m for MDF = 20 Hz. D-F: same as A-C, but the value of $S_{INH,IC}$ was set to 1.1 for the period histogram comparison. Other model parameters: $\tau_{exc} = 1$ ms, $\tau_{inh} = 2$ ms, rate scalar = 0.5. Stimulus parameters: SPL = 40 dB, $f_c = 2145$ Hz, f_m for MDF = 80 Hz.

4.4 DISCUSSION

Relationships between single-unit neural responses in the auditory midbrain of the unanesthetized rabbit to controlled envelope fluctuations and behavioral thresholds in corresponding psychophysical tasks were examined. The stimuli used to establish neural sensitivity to AM were designed on a neuron-by-neuron basis, taking into account potential selectivity for binaural presentation mode, carrier frequency, SPL, and modulation frequency. The ranges of the resulting stimulus parameters mirrored the limits of parametric psychophysical investigations of temporal modulation processing. We now turn to a discussion of the implications of the results in the context of previous work with possible alternate interpretations in mind.

4.4.1 Temporal versus rate information

The current analysis of the average rate and envelope-locked temporal information available in neural responses over a wide range of modulation depths provides some insight into the functional contributions of these different aspects of the response at the level of the IC. To account for psychophysical pure-tone AM *detection* thresholds with a variation in the responses of single cells, a temporal metric (e.g. synchrony to the envelope period) must be considered. At the upper end of the modulation-domain dynamic range ($m > -15$ dB), where vector strength is saturated in many neurons, changes in average rate are sufficient in some neurons to predict AM depth *discrimination* performance.

4.4.1.1 Temporal information: synchrony, phase, and intervals

Hints of low-depth envelope synchronization can be found in previous studies of central AM processing. Significant synchronization coefficients have been measured at AM depths as low as 5% (-26 dB) in the SOC (Kuwada and Batra 1999) and IC (Rees and Moller 1983), although this aspect of the response was not emphasized in the earlier studies. Rhode (1994) and Krishna and Semple (2000) did not systematically use AM depths lower than 10% (-20 dB), but some neurons in the cat VCN and gerbil IC did respond with phase-locked spikes at that depth.

Saturation of the vector strength metric at high modulation depths in the IC has been previously reported (Krishna and Semple 2000, Rees and Moller 1983, but see aggregate trends in multi-unit cluster activity in Mueller-Preuss et al. 1994), although the earlier studies were typically focused on stimulus m well above detection thresholds. A limited dynamic range does not necessarily preclude the use of changes in synchronization to code modulation depth; a population of neurons with staggered thresholds and small dynamic ranges could potentially transmit sufficient information for discrimination. This possibility remains a viable option, given the even spread of synchrony-based thresholds across the perceptually relevant AM dynamic range (Fig. 4-10).

The presence of synchronized spikes at low modulation depths leads to the question of whether changes in synchrony could account for AM depth discrimination performance for standard depths below -15 dB. A quantitative analysis of this question was not attempted because only 9 neurons in our sample were characterized with 1-dB MDF resolution *and* responded with significant synchrony at depths below -15 dB. Qualitatively, the best low-depth discrimination thresholds based on single-cell

synchrony would probably be found in sustained pure-tone responders, which were most likely to have monotonically increasing sMDFs. The sMDFs of these neurons were similar to those of ANFs (Joris and Yin 1992) and VCN neurons (Rhode 1994). An alternative population-based scheme for the coding of AM depth with synchronized spikes might use the spread of synchrony across neurons with different BMFs to represent an increase in stimulus m (again, also a reasonable hypothesis given the distribution of θ_{sync} illustrated in Fig. 4-10).

The synchronization coefficient is a measure of the precision of phase locking, but it is not a complete characterization of the temporal response of IC neurons. This is especially true when the shape of the period histogram changes with f_m or m , which is typical of responses in the IC (Figs. 4-7 and 4-8 illustrate clear examples of such phenomena). One step towards a more complete characterization of a neuron's temporal response properties is to include a description of the dependence of phase on the relevant stimulus parameters (e.g. Anderson et al. 1971). For our purposes, the parameter of interest is the AM depth; response phase is plotted as a function of stimulus m (phase modulation depth functions, pMDFs) in Fig. 4-15 for 154 neurons. The main feature to be extracted from the population of pMDFs is the lack of a systematic dependence of phase on AM depth; this suggests that depth is probably not encoded by variations in phase, at least within single cells stimulated at or near their BMF. Nonetheless, the possibility remains that higher-order processing could make use of phase variations with depth across neurons with different BMFs (such that the stimulus f_m is not equal to the BMF for all of the inputs to the more central processor). An evaluation of this hypothesis is not possible with the current data set, since MDFs were usually only obtained at a single f_m .

Extraction of phase information at higher levels of processing is one potential mechanism that could make use of synchronized spikes. An inter-spike interval code could also provide a representation of the stimulus AM frequency or depth that assumes phase locking. Inter-spike interval histograms (ISIHS) derived from the MTFs of a subset of the population exhibited a prominent peak at $1/f_m$, especially at modulation rates near the sMTF peak in neurons with higher sBMFs. In general, such ISIH features are observed in spike trains with some degree of entrainment to the modulating waveform and suggest the possible existence of a first-order interval representation of AM in some cells at the level of the IC.

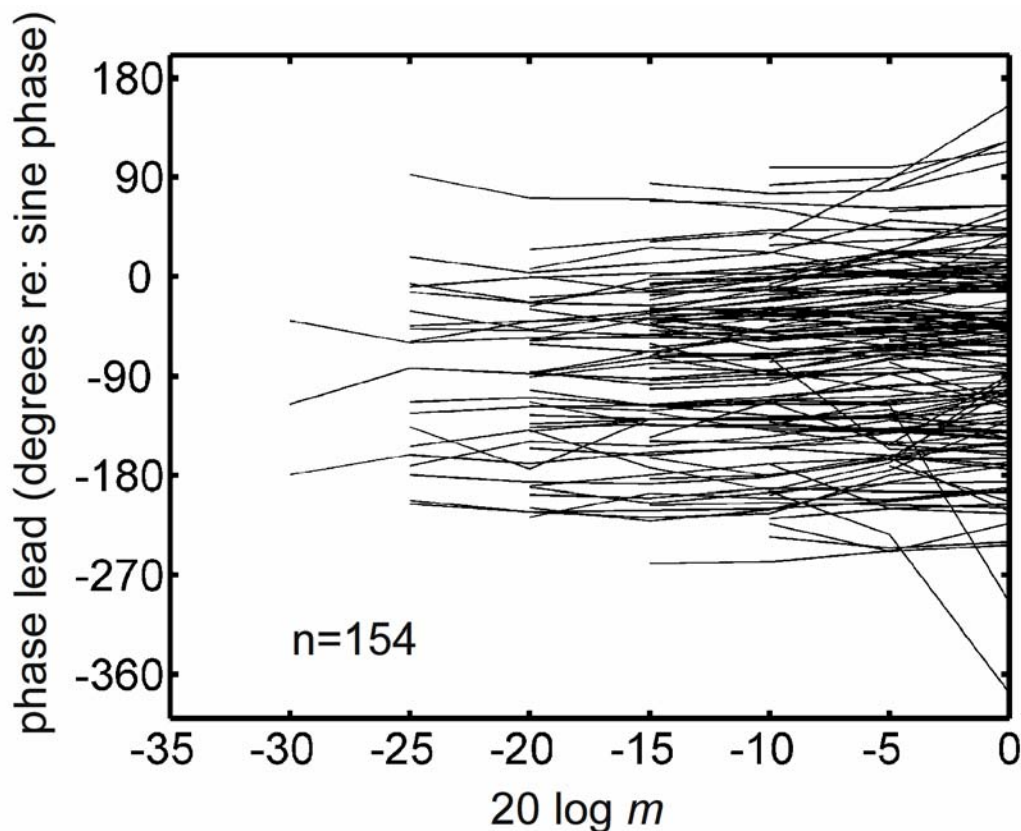


FIG 4-15. Response phase at srBMF does not change systematically with m . Phase MDFs for 154 neurons are shown; a significant value of vector strength was required for phase to be plotted. Not shown are the y-axis phase MTF intercepts (the characteristic phases), which were mainly clustered between 0 and 45 degrees, indicating a strong response at low f_m s on the ascending portion of the stimulus envelope.

4.4.1.2 Rate information: dynamic ranges and pooling

The limitations of a single-neuron rate code for AM depth in the IC are very different from the difficulties associated with rate codes for SPL in the AN (see Colburn et al. 2003). In the periphery, single-fiber rate-based pure-tone detection thresholds in quiet are in line with behavioral thresholds (e.g. Young and Barta 1986). However, predictions of supra-threshold intensity discrimination performance based on single ANF average rates are often much higher than psychophysical thresholds because of saturation and increased variance at high SPLs (e.g. Colburn et al. 2003, Heinz et al. 2001). The situation in the IC for AM coding is in a sense reversed: SAM detection cannot be accounted for with changes in rate (Figs. 4-10 and 4-11), but rate predictions of SAM depth discrimination in some neurons are consistent with psychophysical measures, at least for standard AM depths greater than -15 dB (Fig. 4-13).

An increase of the masker modulation depth in the masked SAM-detection task can be interpreted as an effective shift of the stimulus envelope fluctuations to a higher operating point on the MDF. The ability of average rate changes in several neurons to account for performance at the higher masker modulation depths was therefore not surprising, given the depth-discrimination predictions based on rate at higher standard

depths. It is also possible that the noise masker served as a sub-threshold pedestal for the signal AM, allowing for more frequent crossing of the neurons' putative instantaneous depth threshold in the signal interval but remaining below threshold in the masker-alone interval. This is one explanation for stochastic resonance-type effects observed in biological systems (Ward et al. 2002), and a related mechanism can account for trends in psychophysical masked SAM-detection performance across a wide range of masker depths (Nelson and Carney, in review).

One way to compensate for single-cell rate-based thresholds that are higher than behavioral performance is to assume the existence of a population of neurons that respond with similar rate functions: the pooling of information across a group of cells reduces the effective variance of the rate estimate, resulting in lower thresholds. An assumption that must be met for a pooling scheme to work efficiently is that the neural responses be independent from one another; this requirement has not been established at the level of the IC. In addition, previous studies that tested the efficacy of pooling strategies in the auditory system have simulated a population response by assuming a stereotypical rate function with a fixed relationship between the mean and variance of the rate estimate (e.g. Delgutte 1987, Viemeister 1988). An analysis of the current data set revealed no systematic relationship between the mean spike count and the variance of the count estimate in the unanesthetized rabbit IC (not shown), suggesting that the assumption of a rate function with stereotypical statistical properties at the level of the IC is probably not appropriate (i.e. variance cannot be reliably predicted from the rate).

4.4.2 Comparing animal physiology and human psychophysics

Ideally, neural and behavioral responses would be obtained and compared in the same animal, but AM sensitivity has not been determined psychophysically in the rabbit or any other lagomorph. AM detection thresholds measured with noise carriers in the chinchilla (a rodent) are better than -20 dB for modulation frequencies below 64 Hz (Salvi et al. 1982). Because pure-tone carrier AM detection thresholds in humans (e.g. Kohlrausch et al. 2000) are lower than thresholds measured with noise carriers at medium to high SPLs (e.g. Viemeister 1979), it is reasonable to assume that the use of a deterministic carrier in the chinchilla experiments would yield lower thresholds as well. Furthermore, the low false-alarm rate in the animal study suggests that the chinchillas were using a relatively conservative criterion (Salvi et al. 1982), another factor that could result in elevated thresholds. Given the absence of directly relatable rabbit behavioral data, the chinchilla study provides the most compelling argument against a species difference causing the discrepancy between rate-based AM detection thresholds and human psychophysical results.

4.4.3 Level effects

The improvement in AM detection performance at high SPLs (e.g., Kohlrausch et al. 2000) is not easily explained with single ANF AM responses, since synchronization to the envelope decreases and average rate tends to saturate at high SPLs (e.g. Joris and Yin 1992). The synchronized responses of some neurons in the VCN to AM are less susceptible to degradation at high SPLs than for ANFs (Frisina et al. 1985), although vector strength still decreases in almost all VCN neurons at SPLs commonly used in psychophysical studies (e.g. Rhode 1994, Rhode and Greenberg 1994). In the IC,

variation in SPL can have complex effects on the shapes of MTFs, including sMTF transitions from low-pass to band-pass, and shifts in rBMF of an octave or more (Krishna and Semple 2000).

Figure 4-11 shows that AM depth sensitivity does not systematically depend on the SPL of stimulation in the IC, even when shifts in BMF with level are taken into account, and when quantifications based on both average rate and synchrony are considered. This result suggests that the psychophysical improvement in AM detection performance at high SPLs is likely due to a spread of excitation across a population of neurons (e.g. Kohlrausch et al. 2000). In the IC, the complexity and level-dependence of tuning in both audio frequency (Ramachandran et al. 1999) and modulation frequency (Krishna and Semple 2000) serves as a reminder that the strategy used by the auditory system to code AM over a wide range of SPLs is not likely straightforward in the context of existing models of AM processing (Dau et al. 1997, Hewitt and Meddis 1994, Nelson and Carney 2004).

4.4.4 Temporal adaptation to AM

Another question relevant to psychoacoustics is whether responses to AM in the IC adapt over time. Perceptual adaptation to AM has been demonstrated through the measurement of decreased sensitivity (higher AM-detection thresholds) following a period of prolonged (20-30 min) exposure to modulation (Kay and Matthews 1972, Tansley and Suffield 1983). More recently, Wojtczak and Viemeister (2005) showed that similar effects could be observed using a short-duration adaptor (i.e. a forward masker) and probe imposed on the same carrier. The effect was maximal when the masker and probe were close to one another in both frequency and time, and exhibited time constants of 100 ms or longer.

Presumably, any single-cell responses consistent with the psychophysical AM forward masking effect would also show signs of temporal adaptation at the onset of stimulation. The PSTHs shown in Figs. 4-6, 4-7, and 4-8 suggest negligible variation in the magnitude of the response over time courses of several hundred ms (or any other time course visible on the scale illustrated), even using fully modulated stimuli at BMF. This general finding of weak or nonexistent onset adaptation to AM was consistently observed across most neurons in our population (but see the neuron described in Fig. 4-9). The use of typical psychophysical ramp durations (50 ms) was probably one key contributing factor, although many neurons did not respond with a clear onset component to pure tones with 10-ms ramps either (i.e. the 47% of the population identified as sustained pure-tone responders).

Preliminary results using a stimulus paradigm designed to match the psychophysical study of Wojtczak and Viemeister (2005) also suggest that the effects of previous stimulation observed in psychoacoustics are not salient in the responses of IC neurons: responses to a short (50-ms) AM probe are largely independent of the presence of a preceding 500-ms masker modulation. In other words, AM responses in IC neurons appear to recover too quickly to explain the behavioral effect. Cortical neurons, in contrast, can exhibit time constants that are more in line with the psychophysical results (e.g. Bartlett and Wang 2005, Ulanovsky et al. 2004). The apparent transformation between the colliculus and the cortex can be interpreted either as increased sluggishness or as enhanced context dependence. A better understanding of the functional

consequences of this transition is likely to provide new insights into relationships between single-neuron physiology and psychophysics.

4.4.5 Implications for models of AM processing

Both the structure and response quantification of the original signal-processing style modulation filterbank model for perceptual AM processing (Dau et al. 1997) have been simplified in its most recent implementations (Ewert and Dau 2000, 2004). Specifically, the original resonance filters (which allowed an envelope DC component to pass) were replaced with second-order band-pass filters, and the template-based optimal detector that was used to quantify responses and predict thresholds was replaced with a long-term rms energy integrator (Ewert and Dau 2000). These simplifications were psychophysically justified because they still allowed for reasonable predictions of the relevant AM behavioral data. They were also appealing because the assumption of a simple decision variable instead of a template-correlation approach allowed for a more concrete description of perceptually relevant aspects and dimensions of the stimulus.

Interestingly, a fundamental description of many IC neurons in the current study is more consistent with the predictions of the original versions of Dau's model. Most of the neurons in the unanesthetized rabbit IC responded with a sustained (DC) response to pure tones; the resonance signal-processing temporal filter predicts such a response, but the band-pass filter does not. Responses of the small proportion (13%) of onset neurons are more similar to the output of a band-pass envelope filter. With respect to the model quantification, we have shown that behavioral thresholds can be more accurately predicted if some aspect of the temporal response is taken into account; the optimal detector includes such information, but the rms quantification disregards temporal structure in the response (similar to an average-rate metric in physiology).

The physiologically motivated SFIE model was shown to be flexible enough to account for the AM responses of both sustained and transient pure-tone responders as a function of f_m and m . In addition, the model response exhibited envelope locking at modulation depths lower than those required to elicit an increase in rate. This feature was also observed in almost all of the neural responses. The key model parameter that allowed for the successful prediction of these aspects of the data was the strength of the putative inhibitory input to the post-synaptic model IC cell. Pharmacological studies of AM responses in the IC have shown that blocking inhibition typically does not affect the rBMF of the neuron; instead, an overall rate increase (across f_m) is more common (Burger and Pollak 1998, Caspary et al. 2002, Zhang and Kelly 2003). SFIE model rMTFs (Fig. 4-14) are also in line with these expectations: decreasing the inhibitory strength results in a uniform shift toward higher rates in the rMTF without changing the rBMF.

4.4.6 Future directions

A question that deserves further consideration is, "How do higher processing centers extract and use the envelope-locked timing information present in the responses of IC neurons?" One way to address this question would be to use the SFIE model responses as inputs to higher-order model cells. A preliminary hypothesis regarding the assumptions built in to more central model neurons is that the temporal convergence of synchronized responses (perhaps from cells with different BMFs) could give rise to a rate-based representation of low modulation depths. An alternative approach for understanding how the system might use synchrony information would be to

simultaneously record responses from a group of neurons with different properties to variations in AM frequency and depth. Because of the relatively broad tuning to AM observed across the population (Fig. 4-4), many neurons are likely to respond to a given modulation rate; an analysis of the synchronized responses across cells with different BMFs and depth thresholds might yield new ideas for coding schemes that could in turn suggest specific tests for modeling or psychophysical studies.

In addition to suggesting directions for modeling and population analyses, the current study also has implications for improving predictions of responses to complex sounds such as vocalizations based on responses to simpler stimuli such as SAM tones. In general, speech and other vocalizations have dynamically fluctuating envelopes with a wide range of modulation frequencies and depths. Incorporating an appropriate description of the AM-depth dependence of the response is essential to providing a complete picture of SAM-tone responses in the IC, and is likely to be similarly important for making reasonable predictions of vocalization responses.

ACKNOWLEDGMENTS

We thank Anita Sterns and Lorraine Pawson for valuable technical assistance. We are also indebted to Shigeyuki Kuwada and Blagoje Filipovic for their generous contributions of time and electrode-making advice. This research was supported by NIH-NIDCD F31-7268 (PCN) and NIH-NIDCD R01-01641 (LHC).

CHAPTER 5

General discussion and summary

The central focus of this thesis was the establishment of explicit connections between auditory physiology, psychoacoustics, and computational modeling. Findings from each of the approaches had ongoing influences on the interpretation and design of experiments from the other methods. For example, a physiological modeling study (Nelson and Carney 2004; Appendix of this document) suggested a realistic mechanism that could account for a psychophysical effect that was otherwise not straightforward to interpret (Ch. 2). The negative masking or stochastic resonance phenomenon in masked AM detection was accounted for in the model by a hard modulation-depth threshold. Perceptual AM processing was investigated further in Ch. 3, providing additional tests of our specific physiological model and new constraints for envelope-processing models in general. The physiological experiments described in Ch. 4 were undertaken because of the specific disparity in the parameter spaces, especially concerning modulation depth, used in the existing AM physiological literature and the related psychophysical studies (e.g. those described in Chapters 2 and 3) that was brought to light only after attempting to directly relate simulated neural responses to performance in psychophysical paradigms.

5.1 APPLICATIONS

Understanding basic auditory function at both behavioral and neural levels (often with the help of models) is critical for the advancement of several important fields of applied research. One area is the improvement of devices designed to assist the hearing impaired. Current strategies of both cochlear implant (CI) and hearing aid (HA) signal processors are strongly dependent on the envelope (i.e., the AM composition) of the incoming acoustic signal. Most CI processors use the envelope of a bank of audio-frequency channels to modulate the amplitude (or duration) of current pulses that are sent to electrode pairs at set points along the cochlea. Variations on existing processing schemes have largely focused on restoring “normal” audio-frequency auditory-nerve filtering properties and/or the stochastic properties of peripheral neural spike trains. Given the fundamental assumptions of the overall CI strategy, perhaps more attention should be paid to restoring the physiologically appropriate neural AM response properties, with some basic AM perceptual findings in mind.

Hearing aids pose a very different kind of problem, since the goal is to modify the incoming signal such that the output of the damaged but partially functioning cochlea is more similar to that of a normal cochlea. One universal component of all HA (and CI) processors is a compressive nonlinearity, designed to compensate for the reduced dynamic range that has been repeatedly observed in hearing-impaired listeners. The

assumed compression is fast-acting; gain or attenuation may be applied to envelope fluctuations in the range of the modulation frequencies studied in this thesis on a cycle-by-cycle basis. In addition to limiting the maximum and minimum amplitudes passed to the impaired cochlea, the effect of such compression in terms of AM is to reduce the instantaneous modulation depth of the signal. Our analysis of central responses to AM stimuli suggests that the system uses different neural strategies to represent sounds with low versus high modulation depths. A possible unwanted side effect of dynamic-range compression is the transformation of high-depth envelope features into effective low-depth fluctuations. Again, a physiological investigation of neural AM response properties in impaired animal models would shed more light on the possible side effects of artificially applied compression in terms of the representation of modulation.

Another application area that might benefit from the perceptual and physiological studies described in the preceding chapters is the development of artificial systems that make use of the signal-processing strategies used by the brain for problems such as noise reduction, data scaling and compression, and speech recognition and identification. One angle that has been exploited to some degree is the potential signal-processing benefits that a modulation-frequency analysis would provide (Greenberg and Arai 2001, Atlas and Shamma 2003, Sukittanon et al. 2004).

Our results provide further suggestions for additions and constraints on the structure of the models that underlie the signal-processing algorithms. The current combination of psychophysics and modeling suggests that the actual system may incorporate a hard modulation-depth threshold nonlinearity (Ch. 2), along with envelope change-detectors that respond only to amplitude transients (Ch. 3). Findings from both physiology and psychophysics are consistent with the existence of a representation of low AM depths in the form of temporal fluctuations in the output of modulation-tuned channels, as opposed to increases in long-term (average) responses (Ch. 2; Ch. 4). Also, the system does not appear to emphasize temporal changes in the signal's instantaneous modulation depth; that is, there is not a strong second-order envelope change detector (Ch. 3), and responses to an ongoing modulation do not adapt significantly over time (Ch. 4).

5.2 PERCEPTUAL VERSUS PHYSIOLOGICAL ENVELOPE PROCESSING

5.2.1. Similarities

A fundamental driving hypothesis behind the physiological experiments was that single IC neurons could be conceptualized as perceptual modulation filters. Some basic observations that apply to psychophysical modulation filters and neural IC responses to AM can be summarized as follows. (1) Tuning to AM frequency is broad. (2) A long-term decision variable (e.g. average rate or rms energy) can account for some, but not all, aspects of the response. (3) Maximal sensitivity to tone-carrier AM is essentially independent of modulation frequency (f_m), for rates below 150 Hz. Above 150 Hz, detection and discrimination performance based on temporal fluctuation cues deteriorates. (4) The mechanism for AM depth discrimination likely changes with standard depth. It is possible that the transition from AM detection to discrimination is a result of a change in both the internal representation and in the decision variable.

5.2.2. Differences

5.2.2.1 General survey

It is also relatively easy to identify some limits to the validity of the assumption that IC neurons are physiological implementations of perceptual modulation filters. (1) Many IC MTFs are not band-pass. This is true for any quantification of the response (e.g. rate, synchrony, or synchronized rate), and suggests that a subset of the population might form the substrate for perceptual envelope-domain filtering. (2) Neural AM detection performance does not systematically depend on overall SPL, while perceptual thresholds improve with increasing level. A system with the ability to scan a population of cells and identify the most sensitive (phase-locked) responses might benefit from higher SPLs simply because a larger proportion of the population is likely to respond to high presentation levels. (3) Perceptual AM processing is sluggish or context-dependent; IC responses are not as sensitive to previous stimulation. Because this final statement is potentially not as obvious as the other comparison points based on the data described in the body of the thesis, some elaboration is justified.

5.2.2.2 Time constants: long in psychophysics, short in IC physiology

The terms “sluggish” and “context-dependent” are certainly not synonyms, but the psychophysical effects that are described below could be caused by aspects of a system with either sluggishness or context-dependence (or both). In a sense, the two descriptors differ mainly in their connotations: sluggishness implies a limitation of the system, while context-dependence gives the system more credit in its ability to identify and separate auditory objects. Intuitively, a transition occurs as the sensory processing pathway is ascended, with peripheral stages responding to and following fast fluctuations, and central stations parsing and grouping incoming streams of information. The basic questions that we have addressed are, “Where does the IC fit in this transition with respect to envelope processing? Are IC responses fast or context-dependent?”

Two findings from psychoacoustics provide clear illustrations of the issue. One is the long time course of the release from masking in the envelope-frequency domain demonstrated in Ch. 3. A reasonable description of the psychophysical observation is that the presence of a competing modulation influenced the perception of a signal AM, even when the masker and signal were presented non-simultaneously. The time course of this effect was on the order of several hundred milliseconds. Another clear psychophysical effect can be described in exactly the same way. Wojtczak and Viemeister (2005) showed that the detection of a short probe SAM was made more difficult when the signal was preceded by a masker SAM imposed on the same carrier, for masker-signal delays as long as 200 ms. The correspondence of the time courses of the two effects suggests a single underlying mechanism. A simple sluggishness argument may explain the findings: long-lasting persistence or adaptation could potentially account for the non-simultaneous masking. Alternatively, the perceptual similarity between the masker and signal in both experiments may result in their grouping into a single object, making extraction of the signal more difficult (a context-dependence explanation).

We recorded neural responses in the unanesthetized rabbit to the stimuli used in Wojtczak and Viemeister (2005) in 25 cells to determine whether any long-lasting effects of the masker would alter the response to the signal AM in the IC. Several delays were tested in the physiology (as in the psychophysics). The responses of a typical neuron are summarized in Fig. 5-1, in terms of PSTHs, average rates, and an analysis based on the

temporal correlation between the control and masked signal representations. Essentially no effect of the masker was observed in terms of changes in average rate (Fig. 5-1C) or correlation (Fig. 5-1D), for any tested delay. The quantifications are supported by visual inspection of the PSTHs shown in Fig. 5-1A and 5-1B, which suggest no systematic change in the response to the signal as the delay between masker and probe was increased.

Although only a single example, the neuron illustrated in Fig. 5-1 clearly shows no evidence for sluggishness or context-dependence. An examination of the responses of the remainder of the tested units revealed qualitatively similar responses across the population. The findings are consistent with a view of the IC as a nucleus that acts as a relatively fast envelope-information transmission station that is peripheral to the stages of the system responsible for grouping (or sluggishness). In other words, responses of IC neurons predict negligible forward masking of amplitude modulation. It is likely that IC responses would also predict a substantial “release from masking” using the stimuli of the second experiment of Ch. 3, although this hypothesis was not directly tested with physiological recordings. Returning to and repeating the statement that launched this discussion, perceptual AM processing is sluggish or context-dependent; IC responses are not as sensitive to previous stimulation.

In contrast to neurons in the IC, cells in the primary auditory cortex have been shown to exhibit temporal context-dependence on time scales similar to psychophysical measurements (e.g. Calford and Semple 1995, Brosch and Schreiner 1997, Bartlett and Wang 2005). Given the strong descending (efferent) input from the cortex to the IC (reviewed in Winer 2004), it was somewhat surprising that we did not observe any significant contextual or sluggish responses. A direction for future work might include an investigation of responses to stimuli that provide evidence for perceptual context dependence in the auditory thalamus, which has historically been conceptualized as a “relay” station between the IC and cortex (mainly because of a lack of relevant data).

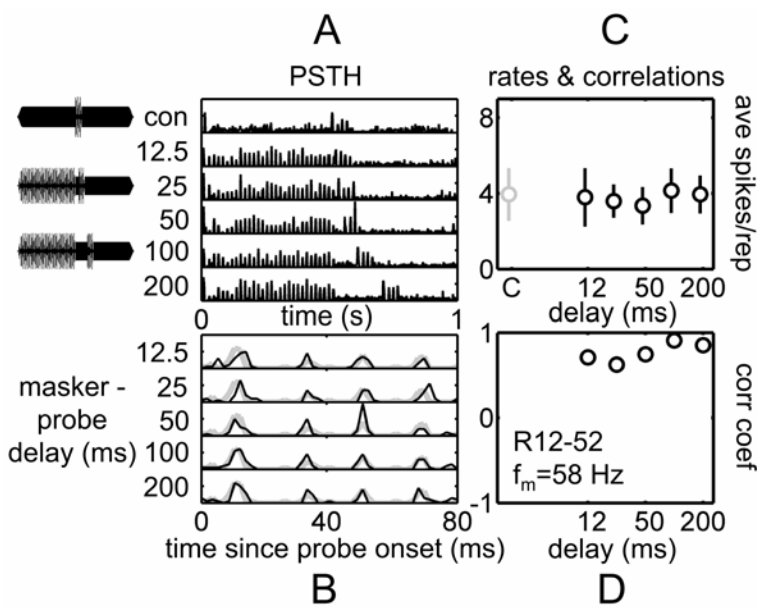


FIG 5-1. Responses of a typical IC neuron that does not exhibit sluggishness or context-dependence in an envelope-frequency forward-masking paradigm. A: PSTHs for the control response, along with the responses to each tested masker-probe delay. Schematic waveforms are shown to the left for the control, 25-ms, and 100-ms delay conditions. B: Comparison of control (thick gray lines) and “masked” (thin black lines) responses for each delay condition. Only the 80-ms window following the onset of the probe is shown (i.e. the comparison conditions were shifted to align with the timing of the probe in the control condition). C: Average number of spikes elicited in each of 20 repetitions of the probe (error bars = standard deviation) as a function of the delay between the masker offset and probe onset. The control condition is labeled with a “C” on the x-axis. D: Correlation coefficients computed between control and masked conditions over the 80-ms portion of the PST shown in B. The bin width used to generate the PSTHs was set to 1.72 ms ($1/10^{\text{th}}$ of the duration of a period of the signal AM).

5.3 OVERALL SUMMARY

To summarize, the data presented in this thesis represent the most direct comparison of basic auditory perceptual and neural AM processing described to date. The experiments suggest constraints for future models of envelope processing and provide insight into the strategies used by the mammalian auditory system to code modulations. Contributions specific to each approach (physiology, psychophysics, and modeling) include: establishment of a baseline for potential future comparisons of neural modulation-depth sensitivity across species, anesthetic states, and levels of the pathway (physiology), a description of several observations that point out limitations for conceptualizing audio-frequency and modulation-frequency perceptual processing in a single framework (psychophysics), and the testing and validation of a simple and realistic mechanism implemented computationally that accounts for several key aspects of the data (modeling). Overall, the integrative nature of the results illustrates the advantages of addressing a question from several angles, allowing the findings from one approach to drive hypothesis testing in another.

APPENDIX

A phenomenological model of peripheral and central neural responses to amplitude-modulated tones

ABSTRACT

A phenomenological model with time-varying excitation and inhibition was developed to study possible neural mechanisms underlying changes in the representation of temporal envelopes along the auditory pathway. A modified version of an existing auditory-nerve model [Zhang *et al.* (2001), *J. Acoust. Soc. Am.* **109**, 648-670] was used to provide inputs to higher hypothetical processing centers. Model responses were compared directly to published physiological data at three levels: the auditory nerve, ventral cochlear nucleus, and inferior colliculus. Trends and absolute values of both average firing rate and synchrony to the modulation period were accurately predicted at each level for a wide range of stimulus modulation depths and modulation frequencies. The diversity of central physiological responses was accounted for with realistic variations of model parameters. Specifically, enhanced synchrony in the cochlear nucleus and rate-tuning to modulation frequency in the inferior colliculus were predicted by choosing appropriate relative strengths and time courses of excitatory and inhibitory inputs to post-synaptic model cells. The proposed model is fundamentally different than others that have been used to explain the representation of envelopes in the mammalian midbrain, and it provides a computational tool for testing hypothesized relationships between physiology and psychophysics. *Note: This appendix was published as a paper with the same title: Nelson, P. C., and Carney, L. H. (2004). J. Acoust. Soc. Am. 116, 2173-2186.*

A.1 INTRODUCTION

Physiological responses to amplitude-modulated (AM) stimuli have provided a basic description of how controlled fluctuations in a signal's temporal envelope are represented at several levels of the auditory pathway. The same class of stimulus has also been used extensively in psychophysical experiments, as a convenient means to investigate perceptual temporal processing capabilities and limitations. Sinusoidally amplitude-modulated (SAM) sounds are often used in both approaches because they can be described by a limited parameter set, and they are logical links between simple (tonal) stimuli and more spectro-temporally complex and biologically relevant signals such as speech. The abundance of data describing neural responses to SAM stimuli provides an opportunity to thoroughly test specific hypotheses concerning the physiological mechanisms that underlie envelope coding. Parallel psychological studies offer chances to better understand neural correlates of AM perception.

The current study uses computational modeling as a tool to quantitatively test realistic neural encoding hypotheses that may be used by the auditory system to code envelope modulations. Model responses will be directly compared to those reported in physiological studies of periodicity coding in the periphery, brainstem, and midbrain (auditory nerve, cochlear nucleus, and inferior colliculus, respectively). Some specific hypotheses that this approach will allow us to test are that: (1) simple interactions

between inhibition and excitation can give rise to modulation-tuned cells, and (2) physiologically realistic model parameter variations can account for the diversity of physiological AM responses. The broad goals of this study are to gain insights on coding strategies. As a result, the model is phenomenological in nature; we have not tried to include details of the underlying biophysical mechanisms in the model structure. Limitations of the predictive capabilities of the model we present here can provide useful clues for developing future experiments.

A.1.1 Extracellular physiological responses to AM stimuli

Typically, neural responses to AM stimuli are quantified based on (1) average firing rate and (2) synchrony (Goldberg and Brown, 1969) to the modulation period. These metrics are usually reported as a function of the stimulus modulation frequency (f_m), resulting in a physiological rate modulation transfer function (rate-MTF, average rate vs. f_m), or a synchrony MTF (sync-MTF, synchrony vs. f_m). For reference, a half-wave rectified sinusoid and a pure sinusoid (i.e. the envelope of a fully-modulated stimulus) are described by synchronization coefficients (SC) of 0.784 and 0.5, respectively (or modulation gains of +3.9 dB and 0 dB for a fully modulated stimulus).

At the level of the auditory nerve (AN), the most complete set of data describing responses to SAM tones is that of Joris and Yin (1992) in the cat. High-characteristic frequency (CF) AN sync-MTFs for low to moderate stimulus levels are uniformly low-pass, with corner frequencies between 600 Hz and 1 kHz and peak modulation gains from 0 dB to +4 dB. Modulation corner frequencies are positively correlated with CF, indicating that temporal responses are dependent on the spectral interactions between AM signal components. Synchrony increases significantly with increasing stimulus modulation depth in AN fibers, and there is an optimal (moderate) stimulus level that elicits the most synchronous response (Joris and Yin, 1992; Smith and Brachman, 1980; Cooper *et al.*, 1993). High-spontaneous-rate (SR) fibers, which make up the majority of the AN-fiber population (Lieberman, 1978), show little or no variation in rate with f_m , while average rates of most low-SR fibers tend to drop slightly as f_m increases (Joris and Yin, 1992). Encoding of AM in the AN is therefore widely assumed to be based on a temporal code, with stimulus modulation information carried in the periodicities of the post-stimulus time (PST) histogram. However, in contrast to psychophysical measures, which are robust across level, AN PST histograms tend to flatten out at high stimulus levels, causing *decreased* response synchrony (Joris and Yin, 1992; Smith and Brachman, 1980). Central processing either recovers the diminished synchrony information at higher stimulus levels (possibly by combining information across audio frequencies), or uses some other feature of the AN neural response to encode envelopes.

Neurons in the cochlear nucleus (CN) can receive convergent information from multiple AN fibers, as well as from collaterals and descending inputs (both excitatory and inhibitory) (e.g. Ryugo and Parks, 2003; Schofield and Cant, 1996). The resulting MTFs are more complex than those measured in the AN. Sync-MTFs can be either low-pass or band-pass, and synchrony to the modulation period is often enhanced relative to AN fibers (Rhode and Greenberg, 1994, cat; Frisina *et al.*, 1990, gerbil). Corner frequencies of CN sync-MTFs can be considerably lower than those measured in high-CF AN fibers, and are generally not correlated with the unit's CF. Most rate-MTFs remain relatively flat at the level of the ventral CN (VCN), with the exception of onset-choppers, which

can exhibit rate-tuning to f_m (Rhode and Greenberg, 1994). Also, some AM response characteristics (i.e. shape of MTF and maximum synchrony values) are correlated with tone-based physiological characterizations of individual neurons (Rhode and Greenberg, 1994). For instance, many onset and chopper units are characterized by sync-MTFs that change from low-pass at low SPLs to band-pass at high SPLs (Frisina *et al.*, 1990; Rhode and Greenberg, 1994). Existing physiological models of AM processing rely on chopper cells in the VCN (e.g. Hewitt and Meddis, 1994); in the current study, we simulate bushy cells (which would be classified as primary-like or primarylike-with-notch based on their responses to tones) at the level of the VCN as an alternative pathway along which temporal envelope information could flow. The specific AM response characteristics that we have achieved in our physiologically-realistic model bushy cells are (1) enhanced synchrony and (2) lower sync-MTF corner frequencies (re: AN inputs; Frisina *et al.*, 1990; Rhode and Greenberg, 1994). The low-pass to band-pass transition in the shape of the sync-MTF as sound level increases was not modeled, as this transition is not evident in VCN bushy cells (Frisina *et al.*, 1990; Rhode and Greenberg, 1994).

The inferior colliculus (IC), like the CN, is an almost obligatory synapse between lower brainstem nuclei and higher processing centers. Interactions between ascending and descending inputs potentially complicate the AM response properties of IC neurons. A systematic and thorough investigation of gerbil IC cell responses to SAM tones provided by Krishna and Semple (2000) has refined earlier work described by Langner and Schreiner (1988) in the cat. IC rate-MTFs often indicate rate tuning to specific modulation frequencies, and many cells exhibit suppressive f_m regions (Krishna and Semple, 2000). IC units are often highly synchronized to f_m , with most sync-MTFs taking low-pass or band-pass shapes. Best modulation frequencies (BMF, f_m eliciting largest response) based on rate are significantly lower (most are less than 100 Hz) than those observed in the VCN or AN based on synchrony. Increasing stimulus modulation depth generally increases the firing rate near rate-BMF, while changing the overall stimulus level has complicated and less systematic effects on rate-MTF shape (Krishna and Semple, 2000).

A.1.2 Intracellular responses and studies of inhibition

Inhibition could be an integral component of the neural processing that underlies the apparent transition from a temporal AM code in the periphery to a rate-based scheme at higher levels (Grothe, 1994; Faure *et al.*, 2003). Several aspects of the inhibitory circuitry in the VCN and IC have been previously investigated. In a pharmacological study of neurons in the chinchilla VCN, Caspary *et al.* (1994) selectively blocked inhibitory inputs and measured response areas (RAs) before and after neurotransmitter blockade. They found that response rates were most often affected near or at CF (~85% of neurons), and concluded that GABA and/or glycine-mediated inhibitory inputs are tonotopically aligned with excitatory inputs onto the same cell. Same-frequency inhibition and excitation (SFIE) is also suggested by morphological labeling studies, which indicate that the likely sources of inhibitory input (dorsal CN and superior olivary complex) are matched in frequency with their post-synaptic targets in the VCN (e.g. Wickesberg and Oertel, 1988).

Other studies have focused on understanding the time courses of inhibitory and excitatory influences in the VCN. Intracellular recordings in the brain-slice preparation

of excitatory and inhibitory post-synaptic potentials (EPSPs and IPSPs) in response to current injections at the stump of the AN have provided details of these time courses that would not be possible to infer using extracellular methods (Oertel, 1983). Specifically, EPSPs of bushy cells in the VCN can be described as alpha functions with a time constant of approximately 0.5 ms, and IPSPs are slower, with time constants on the order of 2 ms. Differences in latencies between excitatory and inhibitory influences in these studies are also informative and consistent with the fact that all AN fibers are excitatory: inhibitory component onsets often begin 1-2 ms after excitation is observed and are therefore assumed to be disynaptic (Oertel, 1983). These observations can be built in to our model of neural processing between AN fibers and VCN neurons.

A similar framework of evidence exists for SFIE circuitry in the IC. Neurons originating in the dorsal nucleus of the lateral lemniscus (DNLL) and projecting to the IC provide one source of GABA-ergic inhibitory input to the central nucleus of the IC (ICc) (Schneiderman *et al.*, 1988). Palombi and Caspary (1996) selectively blocked GABA in the ICc, and showed that rate increases occur mainly near CF in the absence of inhibition, similar to findings in the VCN. Brain slice intracellular recordings have suggested that synaptic modification of incoming spike trains in the IC lasts significantly longer than in the VCN (e.g. Wu *et al.*, 2002). Carney and Yin (1989) used extracellular responses of IC cells to clicks presented binaurally with an interaural time difference (ITD) to infer durations of inhibitory influence. They recorded from cells with a wide range of long-lasting inhibitory components that sometimes suppressed firing for tens of milliseconds following an initial click response.

A.1.3 Modeling

Previous modeling efforts of physiological AM coding have used a coincidence-detection mechanism at the level of the model IC cells to generate band-pass rate tuning for modulation frequency (Langner, 1981; Langner and Schreiner, 1988; Hewitt and Meddis, 1994). The elements of Langner's model, which is focused on pitch encoding instead of pure AM, are only loosely based on physiology. Neurons are hypothesized to perform a cross-correlation analysis between spike trains synchronized to the modulation frequency and the carrier frequency, with small delays between the inputs (Langner, 1981). The model loses its appeal, however, because it is only plausible at low carrier frequencies where phase-locking to the stimulus fine-structure is observed in inputs to the IC (Yin, 2002).

The Hewitt and Meddis model for AM sensitivity is built around a population of intermediate 'chopper' model VCN cells, whose parameters are adjusted to provide a set of envelope-locked inputs to post-synaptic model IC cells. In the framework of their model, AM-synchronized spikes in the VCN tend to fire at intervals related to their chopping period, and stimuli with AM fluctuations that are close in frequency to this inherent chopping elicit a more highly synchronous response. The synchrony-BMF (equivalently, the chopping period) of a model chopper cell is adjusted by varying a potassium-conductance time constant (i.e. Hewitt and Meddis, 1994) in a point-neuron model. One key assumption of this model structure is that the AM rate tuning observed in IC cells is determined by the synchrony tuning properties of VCN choppers that provide inputs to the IC. As a result, the range of rate-BMFs in model IC cells is determined by the range of synchrony-BMFs in their VCN inputs. By allowing the VCN

potassium-conductance time constant to vary from 0.2 ms to 7 ms, Hewitt and Meddis (1994) showed responses from AM-tuned model cells with BMFs between 50 Hz and 400 Hz. One issue for the Hewitt and Meddis model is the disparity in the ranges of VCN chopper synchrony-BMFs (~ 150 Hz – 700 Hz; Rhode and Greenberg, 1994; Frisina *et al.*, 1990) and IC cell rate-BMFs (~ 1 Hz – 150 Hz; Krishna and Semple, 2000).

Physiological work in the auditory system has suggested a possible role for envelope-locked inhibition in AM processing (i.e. Krishna & Semple, 2000; Caspary *et al.*, 2002), but the details of such a mechanism have not been agreed upon or quantitatively tested. We used the results from a modeling study in another modality as a starting point for the development of a new model that incorporates phasic inhibition. Krukowski and Miller (2001) implemented a model of temporal frequency tuning in the visual system in which inhibition dominates over excitation. These model neurons respond only to stimuli with envelope fluctuations because excitation and inhibition can occur out of phase with one another when the stimulus is modulated. Applied to the auditory system, we will show that an implementation of this mechanism can explain the transition from a temporal AM code in the periphery to a rate-based code at the level of the IC. A similar mechanism has recently been used to explain modulation coding at the level of the auditory cortex (Elhilali *et al.*, 2004).

In summary, the existence of physiological “modulation filters” has been experimentally shown. The details of the relationship between behavioral and neural AM responses remain unclear. We propose that one approach towards bridging this gap is to compare the predicted responses of a physiologically-realistic computational model to actual AM responses. The specific model we have chosen to evaluate uses the convergence of long-duration inhibition and short-duration excitation as a mechanism for extracting modulation information from spike trains in lower levels.

A.2 METHODS

A.2.1 Auditory-nerve model

A new version of the physiologically-based auditory-nerve (AN) model developed by Zhang *et al.* (2001) and modified by Heinz *et al.* (2001a) was used to simulate responses at the first level of neural coding. The modifications outlined below were included specifically to improve responses to AM tones, but they did not significantly affect responses to simple stimuli; model responses to tones are similar to those described in Heinz *et al.* (2001a, their model #3).

A.2.1.1 New signal path filters

Effects of basilar membrane compression and two-tone-suppression were omitted in the modified linear version of the Zhang *et al.* (2001) AN model used in the present study. Spontaneous rate was adjusted as in the Heinz *et al.* (2001a) model to match the values described in the physiology, where available. We chose to use linear signal-path filters because of unresolved limitations of the nonlinear AN models at high CFs and high SPLs (discussed later). Tuning of the signal-path gammatone filters was adjusted based on properties of AN fibers in response to AM CF tones. Specifically, CF-dependent bandwidths of the sharp linear AN model (Heinz *et al.*, 2001a, model #3) were changed

to account for the corner frequencies measured in low-pass sync-MTFs. For CFs above 1 kHz, the filter time constant was specified by

$$\tau = \frac{Q_{10}}{2\pi CF}, \quad (1)$$

where the value of Q_{10} is estimated from a linear fit of measured values of Q_{10} (Miller *et al.*, 1997; see Zhang *et al.*, 2001). The value described here, which is an estimate of tuning at intermediate SPLs, is half the τ_{narrow} described in previous versions of the AN model (which was estimated from AN responses at low sound levels). The resulting high-CF model sync-MTFs have significantly higher corner frequencies than previous AN model versions, and are very close to those described in the cat by Joris & Yin (1992). For very low CFs (< 1 kHz), the equation for τ_{narrow} described by Zhang *et al.* (2001, their Eq. (4)) is sufficient to account for the sync-MTF corner frequency in the small amount of data available from low-CF AN fibers.

A.2.1.2 Modified synapse model

Previous AN models have used a simplified implementation of Westerman and Smith's (1988) time-varying three-store diffusion model (Carney, 1993; Zhang *et al.*, 2001; Heinz *et al.*, 2001a) to describe the transformation of a stimulus-driven voltage to an instantaneous firing rate at the output of the IHC-AN synapse. Only the immediate permeability is assumed to depend on the acoustic stimulus (through the inner-hair-cell voltage). The parameters of these models are determined based on desired properties of high-CF PST histograms, including spontaneous rate (SR), steady-state discharge rate, and onset adaptation time constants at high sound levels. Once the initial parameters are set, the immediate permeability described in Westerman and Smith (1988) determines the (level-dependent and time-varying) effective time courses of the two adaptation processes. High-level stimuli result in smaller rapid and short-term time constants.

There is a basic limitation of these prior implementations that directly affects envelope-locked responses to AM tones. Recovery from prior stimulation (offset adaptation) in the models has two components and follows time courses that are set by the rapid and short-term components of onset adaptation. As a result, the final output of the previous synapse model (the product of an immediate permeability and concentration) recovers too quickly after tone stimulation. Physiological AN recordings (Smith, 1977; Smith and Zwislocki, 1975; Harris and Dallos, 1979) suggest that fibers with medium- or high-spontaneous rates recover more slowly than onset time constants would dictate, with a level-dependent dead time and a time constant longer than that of short-term onset adaptation.

Zhang (personal communication) has developed a strategy that effectively avoids these constraints on the synapse model. A SR-dependent shift is added to the desired rate response before implementing the difference equations that describe the diffusion model. After the differential equations are simulated on the up-shifted version of the pre-synaptic response, the resulting output is shifted back down by the same amount and half-wave rectified (so the final synapse output is never negative). By including the shift, offset recovery is effectively slowed because the early offset response is set to zero when the down-shifted output is half-wave rectified. The modification can be physiologically interpreted as a constant neurotransmitter leak in the synaptic cleft (Zhang, personal communication).

For all of the simulations presented here, the shift (which was the same in amplitude for the upward shift as for the downward shift) was equal to twice the desired SR of the fiber. This value was a compromise: larger shifts further increase the modulation gain, but also caused systematic variations in average firing rate with modulation frequency. A SR-independent shift was also considered, but steady-state rates of low-SR fibers at high sound levels were unrealistically high when a high shift value (e.g. shift = 120 sp/sec) was used for all model AN fibers. Figure A-1 illustrates the effect of the 2 x SR shift on a response to a 200-ms (8-ms \cos^2 rise/fall time), 25-dB SPL pure tone presented at CF for an 8-kHz AN model fiber with a SR of 50 sp/sec. The final synapse outputs are shown for the current model version, as well as for the Zhang *et al.* (2001) AN model (their complete nonlinear model with compression and suppression). The effect most relevant to increasing modulation gain can be seen after the tone offset, where the response is suppressed below SR for about 150 ms in this example. If a fully modulated AM tone is thought of as a train of tone pips, the corollary effect would be to suppress firing in the envelope valleys or troughs, which is observed physiologically (Joris & Yin, 1992).

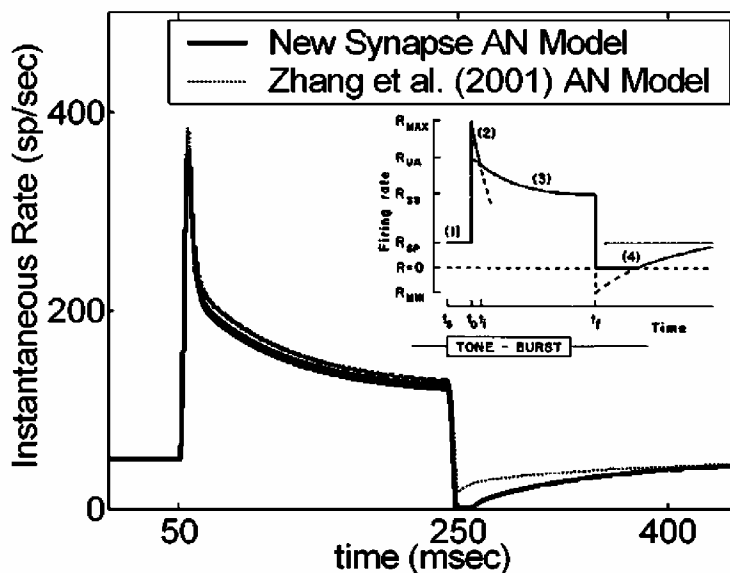


FIG A-1. AN model pure-tone responses (main panel) and schematic of AN responses from Harris and Dallos' (1979) physiological forward masking study (inset; with permission). Modification of the IHC-AN synapse portion of the Zhang *et al.* (2001) AN model resulted in longer (and more physiologically-realistic) offset suppression. Model waveforms are responses of an 8-kHz CF fiber to a 25-dB SPL pure tone at CF.

A.2.1.3 Input and output signals and other details

Inputs to the AN model were the instantaneous pressure waveforms of the stimuli in Pascals. The corresponding output was the time-varying discharge rate of the synapse model, $r_{AN}(t)$, in spikes/second. Characteristic frequency (CF) and spontaneous rate were adjusted and matched to single-fiber physiology (where available) to generate an individual fiber response. Simulation sampling rates varied between 75 kHz and 200 kHz, depending on the CF of the AN fiber and the stimulus frequency. Stimuli had 25-

ms rise-fall times, and both carrier and modulator were started in sine phase, unless explicitly stated otherwise. Detailed responses of a previous version of this phenomenological model have been compared to physiological data describing AN responses to tones, noises, and clicks (Zhang *et al.*, 2001). One goal of the current study was to compare responses of the new AN model to general trends and absolute values of envelope synchrony and average firing rate observed in cat AM physiology (neglecting the initial onset response, as in Joris and Yin, 1992).

A.2.2 Model cochlear nucleus cells

Some basic physiological observations are incorporated into the generation of model responses at the first levels of AN fiber convergence. First, membrane and channel properties of bushy cells in the VCN have the effect of low-pass filtering incoming spike trains (Oertel, 1983). VCN post-synaptic potentials were approximated by alpha functions of the form $P(t) = te^{-t/\tau}$, where values of τ were chosen based on Oertel's intracellular recordings of bushy cells. Inhibitory influences are slower and longer-lasting (i.e. IPSP time constants are longer, on the order of 2 ms) than excitatory input influences, which can be characterized by EPSP time constants of about 0.5 ms. Computationally, smoothed inputs to the model VCN cells were obtained by convolving $r_{AN}(t)$ with the appropriate alpha function. The relative strengths of inhibition and excitation were adjusted by varying the area under the alpha function (or equivalently, the gain of the low-pass filter pass-band). At this first level of convergence, the strength of excitation was assumed to be greater than that of inhibition. Because all AN fibers are excitatory, model inhibitory inputs to VCN cells were also slightly delayed ($D_{CN} = 1$ ms) relative to the excitation, to simulate a disynaptic delay along an intrinsic inhibitory pathway (Oertel, 1983).

Another physiologically-based observation that is included in the simulated VCN cells is that many neurons in the anteroventral cochlear nucleus (AVCN) receive same-frequency inhibitory and excitatory (SFIE) inputs (see Introduction). This mechanism was realized simply by subtracting a membrane-modified, inhibitory AN response from a corresponding smoothed excitatory AN response with the same CF. In terms of the AN input $r_{AN}(t)$ and the approximated post-synaptic potentials, the VCN response is given by:

$$r_{CN}(t) = A[te^{-t/\tau_{exc}} * r_{AN}(t) - S_{CN,inh}(t - D_{CN})e^{-(t-D_{CN})/\tau_{inh}} * r_{AN}(t - D_{CN})], \quad (2)$$

where different time courses of inhibition and excitation ($\tau_{exc}=0.5$ ms, $\tau_{inh}=2$ ms) are represented in the two convolution (*) terms, and same-frequency inhibition and excitation (SFIE) is included by using the same AN instantaneous discharge rate for both inputs. The resulting difference was half-wave rectified so that negative rates caused by the subtraction were represented by zero firing in the resulting model VCN cell instantaneous firing rate. The scalar $S_{CN,inh}$ sets the strength of inhibition relative to the excitation. The scalar A was taken into account to achieve realistic average rates; this value was set to 1.5 for the simulations presented here.

A.2.3 Model inferior colliculus cells

Responses of model VCN cells were used as inputs to higher levels of convergence in a hypothetical processing cascade. Synaptic modifications, input frequency convergence, and relative delays between the inputs were implemented as in the first level of processing. Mathematically, Eq. (2) was used to derive IC model responses, with $r_{AN}(t)$ being replaced by $r_{CN}(t)$, and $A = 1$. One important difference between the two levels of processing was the strength of inhibition: model IC neurons

had stronger inhibition than excitation ($S_{IC,inh} > 1$), while model VCN cells were dominated by their excitatory inputs ($S_{CN,inh} < 1$). As a result, the IC model cell responded strongly only when the timing of the two inputs was such that a peak of excitation coincided with a valley of the envelope-locked inhibition.

Because the time courses of inhibitory influences at the level of the IC are more variable than those observed in the VCN (e.g. Carney and Yin, 1989), the inhibitory time constant in the second level of model cells was systematically varied. Consequences of this variation are described in Results. The single inhibitory time constant of the model should not necessarily be interpreted as a simple time constant of a single IPSP, but rather as an effective time constant describing the overall time-course of the inhibitory influence (which could be, for example, a train of IPSPs rather than a unitary IPSP). A schematic diagram of the two-layer SFIE model is shown in Fig. A-2.

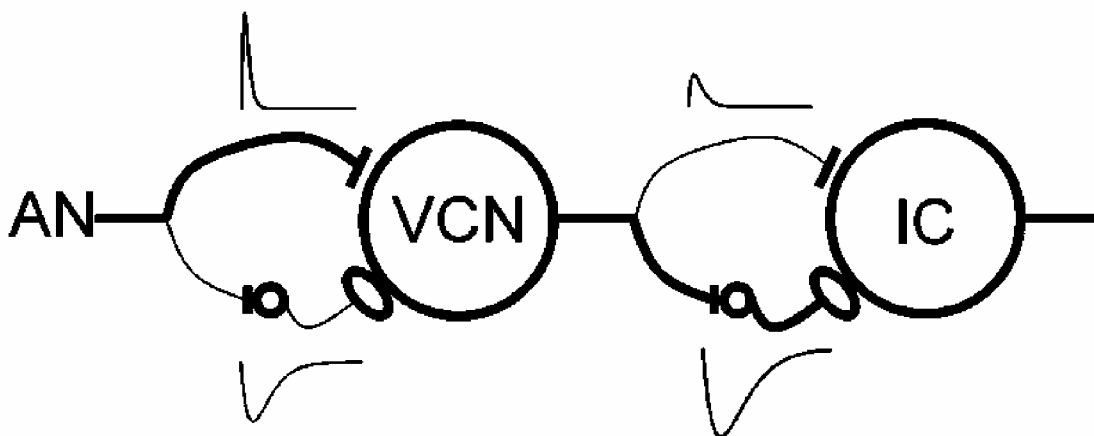


FIG A-2. Schematic diagram of the same-frequency inhibition and excitation (SFIE) model. A single model AN fiber provides the post-synaptic cell with both excitatory and inhibitory input, via an inhibitory interneuron. The thickness of the lines corresponds to the relative strength of the inhibition and excitation at each level. Alpha functions representing the assumed membrane and synaptic properties are also shown.

A.3 RESULTS

A.3.1 AN model responses

Joris and Yin's (1992) systematic study of cat AN responses to SAM tones provides an excellent template for detailed evaluation of AN model responses. Physiological and model AN sync-MTFs and rate-MTFs of high-CF fibers are stereotyped in their shape: sync-MTFs are low-pass at low to moderate signal levels with -3dB cutoffs between 600 and 1000Hz, and rate-MTFs of high-SR AN fibers show little variation in average rate with f_m . Also, both model and actual AN responses at low f_m can be more modulated than a half-wave rectified version of the stimulus (modulation gain > 0).

Physiology and model predictions for a single high-CF AN fiber are illustrated in Fig. A-3 in terms of modulation gain, average firing rate, and PST histogram shape. Model fiber parameters were matched to the physiological fiber description (CF and SR), and stimulus conditions (carrier frequency and relative level) were also duplicated. The main effect of modifying the AN-IHC synapse portion of the AN model was to increase

the envelope synchrony (or equivalently, modulation gain) to physiologically-realistic values. Previous AN model versions had maximum modulation gains of approximately -5 dB at low SPLs (the full nonlinear version of the Zhang *et al.* (2001) model predicts a SC of 0.29 for a fully-modulated stimulus), while the version described here has a peak gain of +2.5 dB (SC = 0.66). The low-pass shape and corner frequency of the sync-MTF are consistent with the data (Fig. A-3A); the flat rate-MTF and steady-state average rate values are also well-predicted (Fig. A-3B). Shapes of the PST histograms are not strictly determined by measurements of synchrony and rate; however, the model responses are also qualitatively similar in PST shape to those measured physiologically (Fig. A-3C).

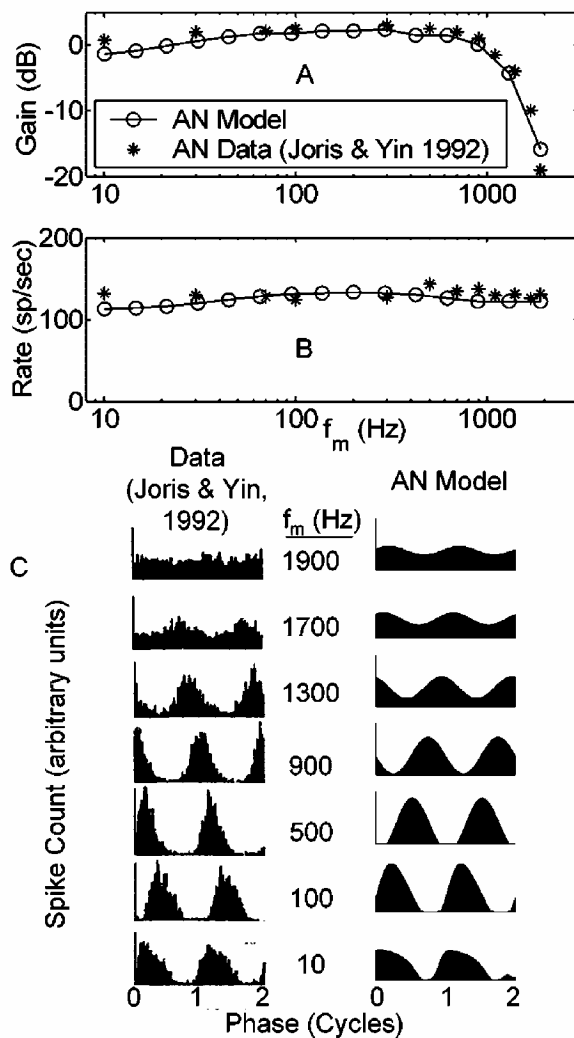


FIG A-3. Comparison of model and actual modulation transfer functions and period histograms. (A) Physiological and model AN sync-MTFs for a high-CF fiber. (B) Rate-MTFs for the same fiber (onset response neglected). (C) Period histograms for the actual AN fiber (left column) and model AN fiber with matched parameters. Two cycles of the response are shown starting at a zero sine phase point of the modulating waveform. CF = 21 kHz; SR = 61 sp/sec. Stimulus carrier frequency = CF; SPL set at ~20 dB above rate threshold. Cat data from Joris and Yin (1992, their Fig. 10; with permission).

Another characteristic of AN fiber responses to AM can be observed by varying the modulation depth (m) at a given f_m . The effects of changes in m are important to test with the model for our long-term purposes because psychophysical data is often reported in terms of a threshold m that is always less than $m = 1$ (i.e. less than 100% stimulus modulation). Physiological and modeled response synchrony and modulation gain are plotted against stimulus modulation depth for a high-CF, high-SR AN fiber in Fig. A-4. Predictions based on the modified synapse AN model fit the data quite well, and are a significant improvement over the previous model version. In addition, the model responses (not shown) are very similar in shape to the PSTs illustrated in Joris and Yin (1992, their Fig. 1). For direct evaluation of the contribution of the new version of the AN-IHC synapse model, predictions are also shown using the same pre-synaptic model structure in combination with the Zhang *et al.* (2001) synapse model. Synchrony using the new “constant neurotransmitter-leak” model is significantly enhanced over the previous version for all modulation depths.

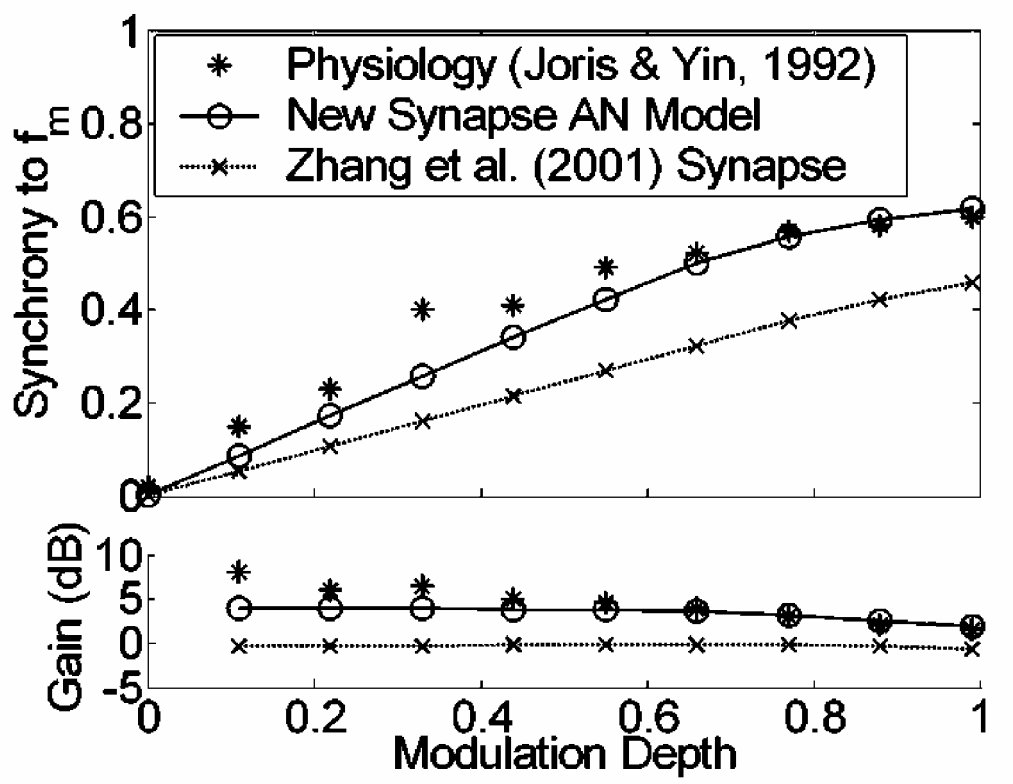


FIG A-4. Changes in synchronization to the modulation period with stimulus modulation depth for model and actual AN fibers (physiological responses from Joris & Yin, 1992, their Fig. 1(b); with permission). Predictions from two AN model versions are shown; the version with the modified synapse model will be used for all subsequent simulations. Fiber parameters: CF = 20.2 kHz; SR = 53 sp/sec. Stimulus parameters: carrier frequency = CF; f_m = 100 Hz; SPL set to elicit the maximum synchrony value (20 dB SPL for the model fiber; 49 dB SPL for the actual fiber).

In addition to comparing measured and simulated response characteristics for different modulation frequencies and depths, we have also evaluated the level-dependence of the AN-model synchrony and rate. Physiologically, modulated synchrony-level functions (synchrony vs. level for a given f_m) are non-monotonic, with a peak at some best modulation level (BML, usually 10-20 dB above rate threshold) over a wide range of modulation frequencies (Joris and Yin, 1992). The AN model used in this study is characterized by the same level-dependent synchrony trends. Figure A-5 compares physiology (unconnected symbols) and simulations (connected symbols) of responses to four combinations of fiber CF, SR, and stimulus f_m in terms of average rate (*) and synchrony (o) to the modulation period. The general shapes of both rate-level (monotonic and saturating) and sync-level functions (non-monotonic) are accurately predicted by the AN model. Thresholds in the model fibers are set near 0 dB SPL, regardless of SR; as a result, the simulated low-SR rate-level and sync-level functions (lower two panels) are shifted toward lower sound levels. Peak firing rates are higher in the AN model low-SR and low-CF groups in comparison to Joris & Yin's (1992) physiological results (panels A, C, and D of Fig. A-5). This is an inherent property of the model, and it is one that we chose not to focus on in our efforts to match the AM response properties. However, the rates are within a reasonable range when population data from both Joris & Yin (1992) and Liberman (1978) are considered. Low-SR model fibers exhibit slightly lower peak envelope synchrony and broader synchrony-level functions than the actual AN fibers (panels C and D). The high-SR, high-CF model fiber shown in panel B of Fig. A-5 is accurate in terms of its rate threshold, dynamic range, maximum rate, and synchrony-level function. It is this class of AN fibers that is described most thoroughly in Joris & Yin's study, and it is responses from these fibers that were used to test and validate the AN model.

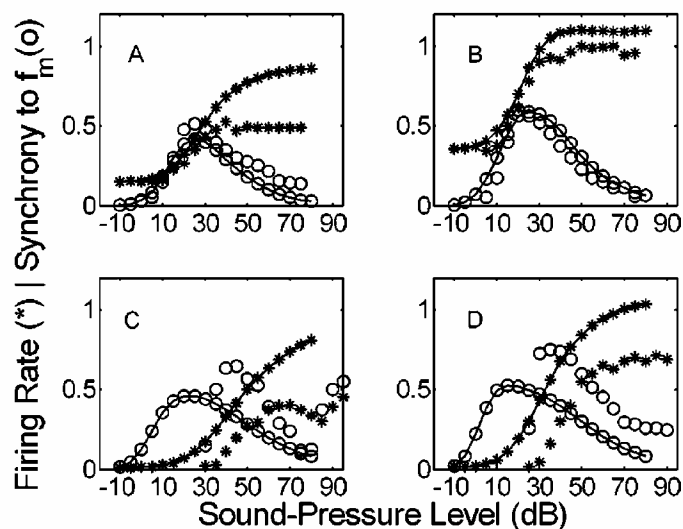


FIG A-5. Comparison of actual (unconnected symbols) and model (connected symbols) rate-level (*) and synchrony-level (o) functions for 4 combinations of AN parameters. Firing rates are normalized by 200 sp/sec. CF, SR, and f_m ; A: 510 Hz, 30 sp/sec, 10 Hz, B: 12 kHz, 71 sp/sec, 100 Hz, C: 710 Hz, 2 sp/sec, 10 Hz, D: 8.1 kHz, 2.6 sp/sec, 100 Hz. The carrier frequency was set to the fiber's CF. Physiological responses re-plotted from Joris & Yin (1992, their Fig. 5; with permission).

Joris and Yin (1992) also quantified the relationship between a fiber's sync-MTF corner frequency and its tuning-curve bandwidth. High-frequency AN fibers have broader tuning than low-CF fibers, which allows more AM sideband energy to pass for relatively higher stimulus f_m . Accordingly, -3dB cutoffs increase with CF. Interestingly, at very high CFs (> 15 kHz in the cat) the positive relationship between CF and cutoff frequency saturates (the corner frequency becomes independent of CF). The AN model predicts similar trends, although the absolute values of the model cutoff frequencies are slightly lower than the average physiologically measured values (but model values lie within the scatter of the data; Joris and Yin, 1992, their Fig. 14). Figure A-6 shows examples of model sync-MTFs at three CFs along with those from comparable AN fibers (from Joris & Yin, 1992).

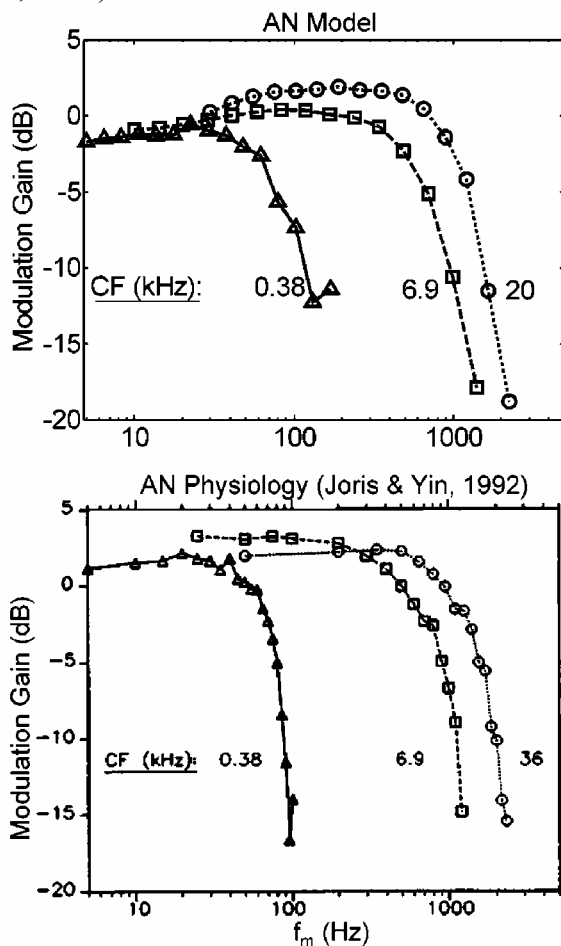


FIG A-6. Sync-MTFs ($m = 0.99$) for three fibers with different CFs: AN model and actual data (from Joris & Yin, 1992, their Fig. 13a; with permission). SR of each fiber; triangles: 24 sp/sec; squares: 6.3 sp/sec; circles: 39 sp/sec. Model SPL = 24 dB.

A.3.2 Model CN cell responses

Figure A-7 illustrates simulations of the effect of the convergence of slow inhibition and fast excitation with the same CF on a post-synaptic model VCN cell. The model's synchronous response to SAM is affected in two ways that are consistent with physiology (Rhode and Greenberg, 1994; Frisina *et al.*, 1990): (1) synchrony is enhanced

with respect to AN inputs at low modulation frequencies, and (2) the upper frequency limit of phase-locking to the envelope is lower in the model VCN cells. The model parameter that most directly contributes to these response characteristics is the strength of inhibition (relative to excitation). MTFs for model VCN cells with a range of inhibitory to excitatory strength ratios between 0 and 0.6, along with the input AN MTFs, are shown in Fig. A-7. Average rates decrease for model cells with stronger inhibition (for all f_m), and rate-MTFs can exhibit a shallow peak for the model cells receiving the strongest inhibitory inputs (lower panel). Rates in VCN model cells with no inhibition are higher than the AN inputs because the excitatory alpha-function area (or low-pass filter gain) was greater than unity (area = 1.5). Also, as inhibition increases, VCN sync-MTFs systematically have higher corner frequencies, and maximum modulation gain values increase (upper panel).

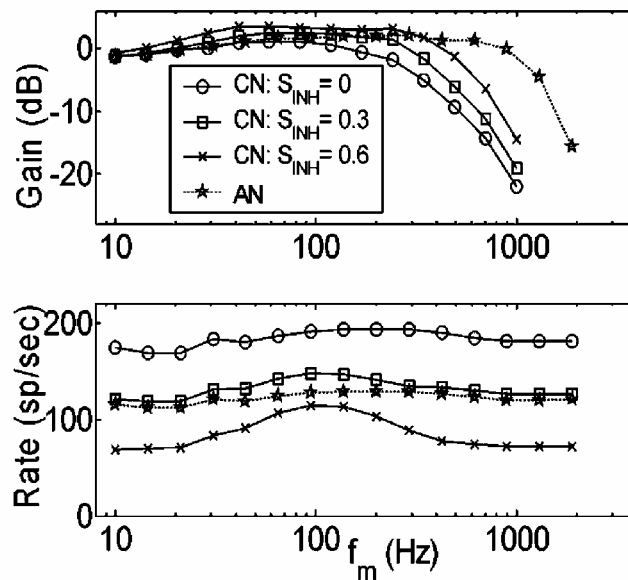


FIG A-7. Sync-MTFs (top panel) and rate-MTFs (bottom panel) for an 8-kHz CF AN model fiber, and three model VCN cells with varying strengths of inhibition (re: strength of excitation). Model AN fiber SR = 50 sp/sec. Model VCN $\tau_{exc} = 0.5$ ms; $\tau_{inh} = 2$ ms; $D_{CN} = 1$ ms. Stimuli were presented at 24 dB SPL.

What is the specific mechanism that gives rise to enhanced envelope synchrony in the model VCN cells? The effect can be understood by considering the differences in assumed membrane and/or channel properties for inhibition and excitation. Stronger and faster excitatory inputs interact with weaker and slower inhibitory inputs in ways that depend on the stimulus f_m to produce the final model VCN cell response. For low and intermediate modulation frequencies (i.e. $10 \text{ Hz} < f_m < 200 \text{ Hz}$), the more sluggish inhibitory inputs ($\tau_{inh} = 2$ ms) are not able to follow the fast fluctuations in the AN responses as faithfully as the excitatory inputs ($\tau_{exc} = 0.5$ ms). Higher synchrony results because excitation dominates for a more focused time interval (near the modulation period onset). For $f_m > 200$ Hz, the slow inhibitory component acts to effectively remove a DC offset from the excitatory component. This causes higher modulation gain values

and higher sync-MTF cutoff frequencies for model cells that receive stronger inhibition. VCN sync-MTF corner frequencies are lower than those in the model AN fibers (regardless of the relative strength of inhibition) because they are limited by the excitatory synaptic properties that modify the ascending VCN model cell inputs.

For nearly equal-strength inhibition and excitation (CN $S_{inh} > 0.6$, not shown), model VCN rate-MTFs are more band-pass in shape than has been reported in most physiological studies of bushy cells. When the overall strengths of the cell's two inputs are nearly the same, there is a narrow range of f_m for which the temporal interactions between inhibition and excitation cause high firing rates in the postsynaptic cell. At very low f_m and very high f_m , the two inputs are similar in their ability to keep up with the stimulus modulations, resulting in a low discharge rate since inhibition is nearly as strong as excitation. To avoid band-pass rate-MTFs, the strength of inhibition in the model VCN bushy cells was set to 0.6 when they were used to provide inputs to model IC cells.

A.3.3 Model IC cell responses

A.3.3.1 MTFs & effect of varying time constants and delays

A second layer of model cells receiving inhibition-dominated SFIE inputs is hypothesized to represent IC units that integrate information from many convergent inputs. The most basic results observed in the model responses are that IC cells fire only over some narrow range of f_m (i.e., they are rate-tuned to f_m), and their AM responses are highly synchronized to the modulation period. This is consistent with physiological responses in the gerbil (Krishna and Semple, 2000) and cat (Langner and Schreiner, 1988). The BMF of a given IC model cell is determined mainly by the time constants of the inhibitory and excitatory influences: fast-acting inputs give rise to high BMFs; slower time constants result in lower BMFs. We constrain the inhibitory τ to always be equal to or longer than the excitatory τ when generating model responses. The range of BMFs that can be obtained by varying these parameters over a physiologically-realistic range is illustrated in Fig. A-8. Each rate-MTF describes the responses of a model cell with a given combination of $\tau_{IC,exc}$ and $\tau_{IC,inh}$. The same model AN fiber and VCN cell provided inputs to each of these model IC cells (i.e. rate-tuning in the IC is not determined by synchrony-tuning in the VCN). Absolute rate (top panel) and normalized rate-MTFs (bottom panel) are shown. Note that a wide range of rate-BMFs (~40 Hz – 120 Hz) can be obtained with a fixed $\tau_{IC,exc}$, and variation of $\tau_{IC,inh}$ from 1 ms to 7 ms. Tuning to even slower envelope frequencies is achieved by assuming longer time constants (Cell A in Fig. A-8, with $\tau_{IC,exc} = 5$ ms and $\tau_{IC,inh} = 10$ ms, is tuned to ~20 Hz). The upper limit of model BMFs (~120 Hz) is consistent with that observed in the gerbil IC (Krishna and Semple, 2000). This boundary is determined in the model by restricting the time constants of the excitatory and inhibitory inputs to be longer than 1 ms (Wu *et al.*, 2002; Carney and Yin, 1989).

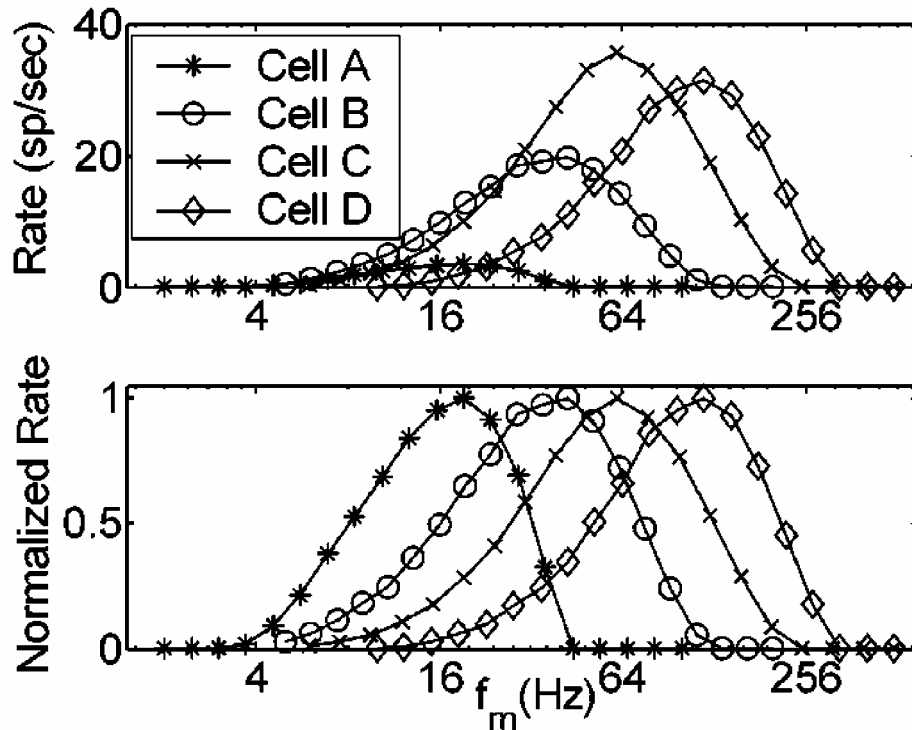


FIG A-8. Absolute and normalized rate-MTFs of four model IC cells with different combinations of excitatory and inhibitory time constants. $\tau_{IC,exc}$ and $\tau_{IC,inh}$; Cell A: 5 ms & 10 ms; Cell B: 2 ms & 6 ms; Cell C: 1 ms & 3 ms; Cell D: 1 ms & 1 ms. Common parameters to all four cells: AN CF = 8 kHz; AN SR = 50 sp/sec; VCN τ_{exc} = 0.5 ms; VCN τ_{inh} = 2 ms; D_{CN} = 1 ms; $S_{CN,INH}$ = 0.6; D_{IC} = 2 ms; $S_{IC,INH}$ = 1.5. Stimulus carrier frequency = AN CF; m = 1; SPL = 24 dB.

The exact f_m that elicits the largest rate-based model response can be further adjusted by changing the delay between excitation and inhibition. When the inhibitory delay is kept within a physiologically realistic range (< 10 ms), rate-based BMF shifts at low modulation frequencies are relatively small but systematic. Longer delays result in a shift of rate-MTF peaks to lower f_m and increases in rate at BMF. Short delays between inhibition and excitation cause maximal overlap in the envelope-locked inputs, and thus decrease the overall firing rate and increase the cell's BMF. Grothe (1994) has proposed a scheme for AM tuning in the bat medial superior olive (MSO) that uses a pure delay between inhibition and excitation to set the model cell's BMF. We chose not to rely on such a mechanism because of one specific consequence that is not observed in the physiology: a multi-peaked rate-MTF results, with rebounds in rate at envelope frequencies that are multiples of the "fundamental" f_m . This problem is avoided in the current model by the use of strong inhibition that is de-synchronized at high f_m .

A.3.3.2 Effect of varying stimulus modulation depth

Responses of model IC cells change for stimuli with different modulation depths in a way that is consistent with physiological observations (Krishna and Semple, 2000). Figure A-9 shows an example of physiological MTFs (top panels, from Krishna and Semple, 2000) and model MTFs (bottom panels); the parameter in the figure is stimulus m . Firing rate increases with modulation depth, most significantly near BMF. Synchrony

saturates rapidly as m is increased. The possible shapes of the model IC sync-MTFs are limited because a narrow range of f_m s elicits a high enough rate to compute a significant SC. Envelope locking persists in the response at modulation rates well above the high-frequency rate-MTF cutoff frequency. Each of these effects of varying m has been observed in physiological recordings (e.g. Krishna and Semple, 2000). One clear discrepancy between model and data is in the lack of a synchrony roll-off in the model at high f_m . This is due to the fact that the model is deterministic, and any response will be synchronized, even at very low rates. The addition of some amount of noise (i.e. jitter in the discharge times) to the model IC cell output would avoid the artificially high synchrony values for conditions with very low average rates (e.g. at high modulation frequencies for the model cell in Fig. A-9). Figure A-9 also illustrates the stereotyped shapes of model cell rate-MTFs. The high-frequency roll-off is more abrupt than the low-frequency transition on a logarithmic frequency axis, and rates go to zero at very high f_m . These features have also been described in gerbil IC recordings, and have been hypothesized (but not tested) to be caused by inhibitory inputs (Krishna and Semple, 2000).

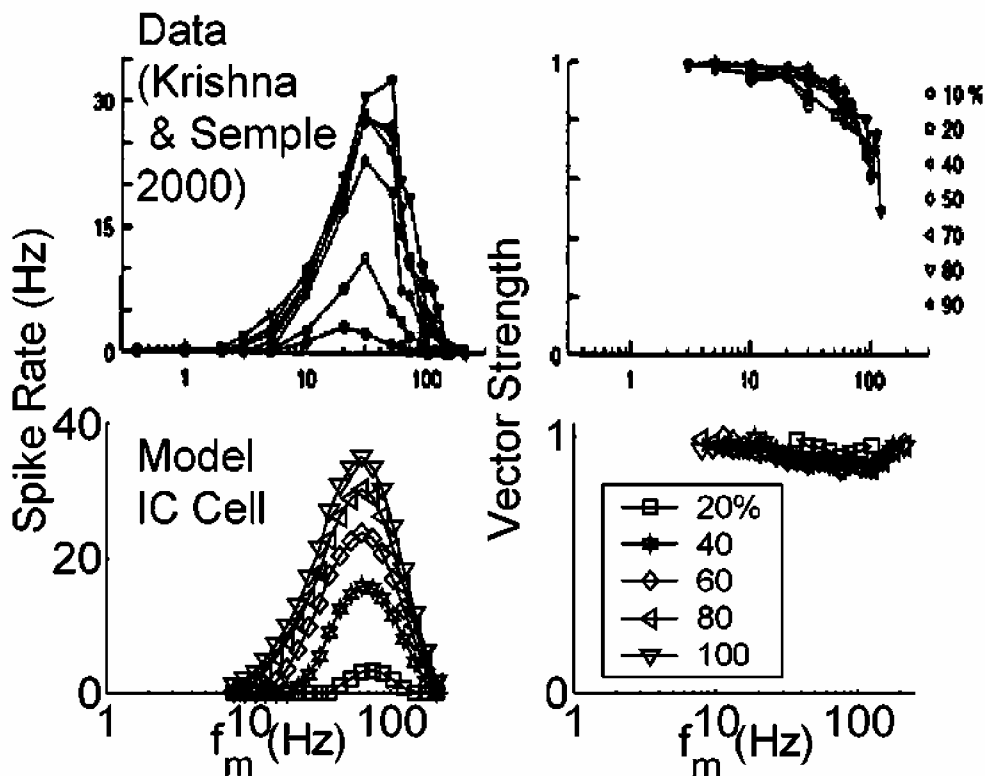


FIG A-9. Effect of varying stimulus m on actual (upper panels) and model (lower panels) IC unit MTFs. Model and stimulus parameters were the same as those describing Cell C in Fig. A-8. Rate-MTFs are shown on the left; sync-MTFs on the right. Gerbil physiological data from Krishna and Semple (2000, their Fig. 2C; with permission).

A.3.3.3 Effect of varying stimulus SPL

IC cell rate-MTFs can change with sound level (e.g. Krishna and Semple, 2000). This point has important implications for relating physiological responses to

psychophysical measures, which are almost always robust across a wide range of SPLs. Predicted rate-MTFs based on the IC model cells also depend on SPL in ways broadly consistent with some of the single unit physiology (although the physiological level-dependence can be very different from cell to cell; Krishna and Semple, 2000). Figure A-10 compares responses from one cell in Krishna and Semple's (2000) study to those of the model IC cell illustrated in Fig. A-9. Several trends are predicted by the model. (1) The largest peak rate response is elicited by a medium-SPL stimulus (40 dB in the physiology; 30 dB in the model). (2) The rate-BMF shifts to lower f_m with increasing sound level. (3) The bandwidth of the rate-MTF tends to increase with sound level. (4) Synchrony remains relatively high despite these changes in the rate-based characterization.

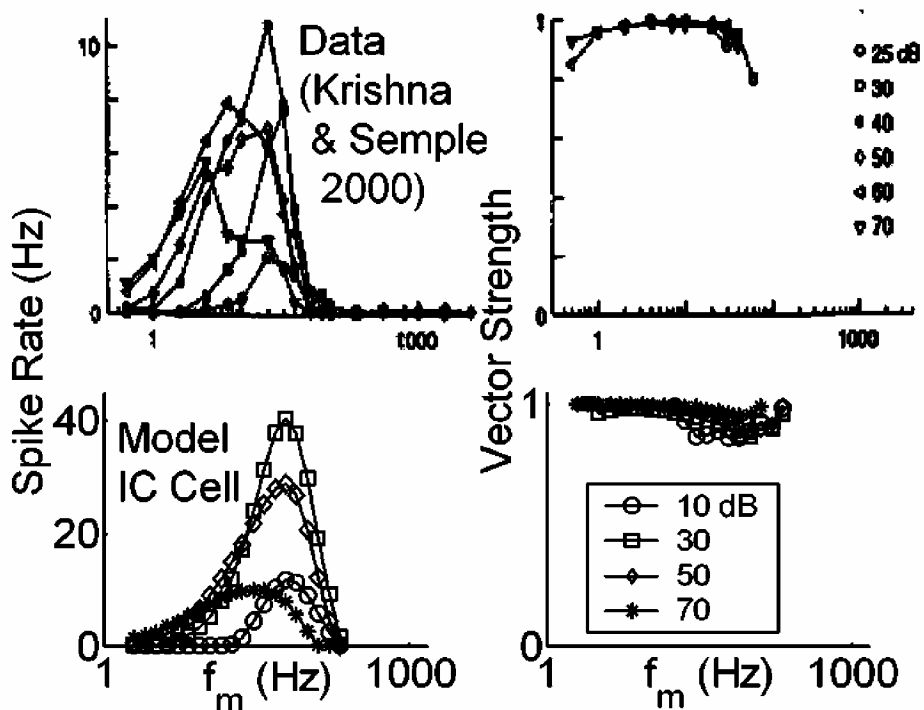


FIG A-10. Effect of varying stimulus SPL on actual (upper panels) and model (lower panels) IC unit MTFs. Model and stimulus parameters were identical to those described in Fig. A-9. (with $m = 1$). Data from Krishna and Semple (2000, their Fig. 5D; with permission).

A.3.3.4 Model mechanisms and PSTs

In the present model, mechanisms underlying specific features of model rate-MTFs and sync-MTFs can be determined directly by considering the intermediate waveforms that shape the model IC cell response. At low f_m , neither the inhibitory nor the excitatory inputs are significantly modified by their low-pass membrane and synaptic properties. That is, inputs to the model IC cell are able to closely follow the VCN response. As a result, excitation and inhibition are overlapping for most of the stimulus cycle (depending on their relative delays), and since these cells are inhibition-dominated, the average firing rate is very low. For stimulus frequencies near the cell's BMF, interactions between inhibition and excitation become more interesting. The slower

inhibitory inputs start to lose much of their AC response, and the weaker excitation will cause higher post-synaptic rates when its stimulus-locked peaks line up in time with the stronger inhibitory ‘valleys’. At high f_m , the inhibitory inputs are nearly tonic, and the excitatory components are also smoothed by their effective synaptic and membrane filtering properties. In the extreme, at very high f_m , both inputs to the model IC cell are time-independent. The result is little postsynaptic activity because inhibition dominates, and temporal interactions can no longer provide stimulus-locked firing in the final response.

Despite the good agreement between available physiology and model IC cell responses in terms of their rate- and sync-MTFs, the predictive power of the model structure in describing PST histogram shapes has not yet been demonstrated. Examples of physiological (from Langner and Schreiner, 1988) and modeled temporal responses for a single IC unit at various modulation frequencies (with a stimulus rise-fall time of 5 ms) are shown in Figure A-11. The model cell was chosen such that the rate-BMF was similar to that of the physiological cell (BMF = 60 Hz). A strong onset is present in both data and model responses, regardless of the stimulus f_m . This feature is not part of any of the model quantifications presented up to this point (all rate and synchrony calculations were made in the steady-state portion of the response). In the model, the onset response is caused by the relative sluggishness of the inhibitory input with respect to excitation. These model cells would be classified as onset units in terms of their pure-tone responses.

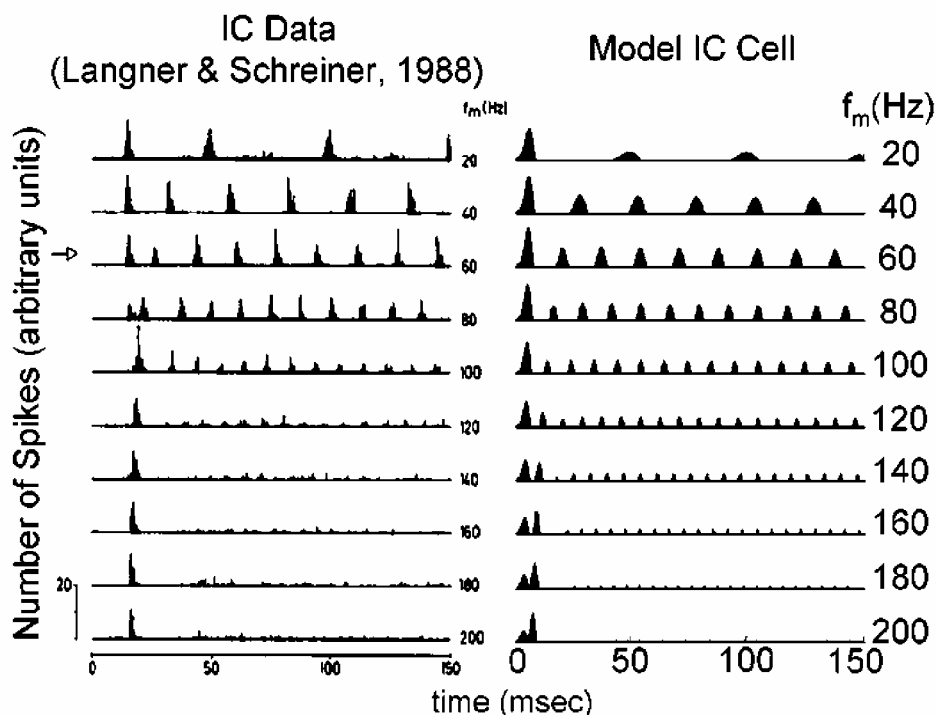


FIG A-11. Comparison of actual (left) and model (right) IC cell PST histograms for a range of stimulus f_m . Model parameters were the same as the cell illustrated in Fig. A-10 (with SPL = 20 dB). Physiological PSTs from Langner and Schreiner (1988, their Fig. 2, unit IC115; with permission).

Other subtle similarities are present that are not reflected in rate- or sync-MTFs. The slow build-up at high f_m after onset that is observed in IC physiology (e.g. at $f_m = 120 - 200$ Hz) is captured by the model. This phenomenon is a result of the slow inhibitory input in the model cell, which determines how quickly the cell response reaches steady-state. Also, two peaks in the onset response at certain modulation frequencies are present in both data and model (caused by fixing the starting phase of the signal modulation). In general, the qualitative similarities in the PST histograms suggest that the simple mechanism proposed here can predict more than gross rate and synchrony changes with modulation frequency.

A.3.4 Summary

The apparent transition from a temporal AM code in the periphery to a rate-based scheme at the level of the IC is predicted by a simple neural encoding hypothesis. Specifically, two levels of convergent slow inhibition and fast excitation reproduce several established response characteristics of VCN and IC neurons, at least qualitatively. The most fundamental difference between the processing strategies at the two levels is the strength of the inhibitory input relative to the excitatory component. A summary of the responses of two-stage SFIE is shown in Fig. A-12.

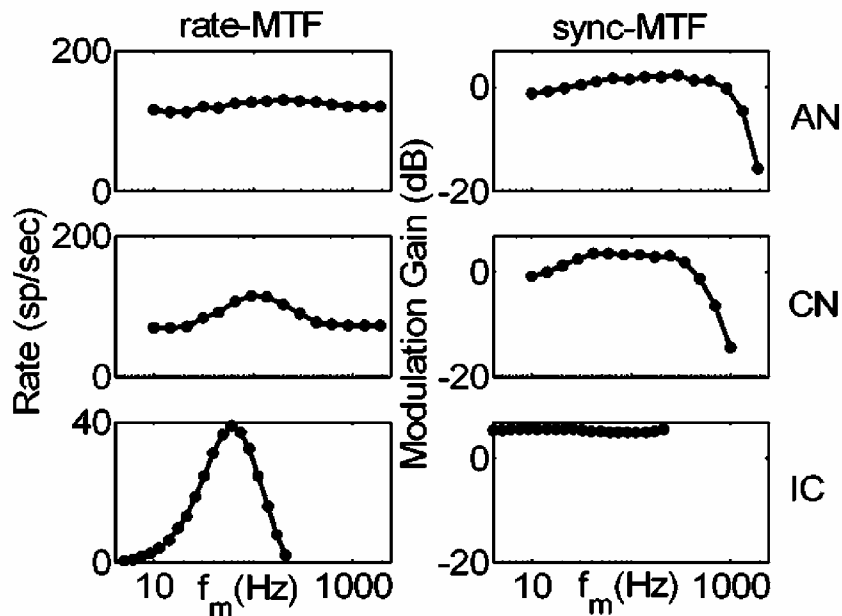


FIG A-12. Summary of the effects of SFIE on rate- and sync-MTFs for two model layers of convergent excitation and inhibition. Model and stimulus parameters match those of the previous three figures (with SPL = 24 dB and $m = 1$).

A.4 DISCUSSION

Several important features of neural AM encoding in the auditory brainstem and midbrain were predicted by the cascaded convergence of same-frequency inhibitory and excitatory inputs with the physiologically realistic time courses described in this study. A modified version of the Zhang *et al.* (2001) AN model that is consistent with physiological recordings in the cat provided inputs to this processing mechanism. At the level of the AN, changes in envelope synchrony and rate with modulation frequency and depth were accurately captured by the new AN model. A single layer of convergent

inhibition and excitation yielded model responses with increased synchrony relative to AN firing patterns, similar to trends seen in VCN bushy cell physiology. In a second level of convergent SFIE, model cell responses were rate-tuned to stimulus modulation frequency by allowing the inhibitory inputs to be stronger than the corresponding excitatory inputs. Rate tuning to f_m has been reported in both gerbil and cat IC (i.e. Krishna and Semple, 2000; Langner and Schreiner, 1988).

A.4.1 Limitations of the AN model

A significant shortcoming of the AN model used in the present study becomes apparent in its response to high-SPL, high- f_m AM stimuli, which elicit unrealistically high synchrony in the model. This results in strongly band-pass model AN sync-MTFs at high sound levels. Evaluation of the model in this study was limited to low and moderate SPLs, to avoid using physiologically inconsistent AN responses as inputs to higher processing centers. A possible source for this discrepancy is the speed and strength with which the onset properties of the AN model act on a cycle-by-cycle basis. At high levels and low f_m , there is a clear onset component with each AM stimulus cycle, but synchrony is reduced compared to lower-SPL stimuli because the remainder of the cycle histogram (after the onset) is saturated. In contrast, high-SPL, high- f_m SAM stimuli cause a highly synchronized response that is completely dominated by the strong onset, with no saturated portion of the cycle histogram. It should be noted that at very high f_m (i.e. $f_m > 1000$ Hz), envelope synchrony still rolls off at high SPLs; it is for stimulus f_m between 20 and 500 Hz that the time course of onset adaptation causes the most significant increase in synchrony (that has not been observed physiologically). A clear direction of future work is to understand and possibly modify the role that the synapse model onset components play in shaping the high-SPL AM cycle histograms.

This inconsistency between the model and data at high levels is tempered by the fact that the AN model fibers always have a pure-tone rate threshold of approximately 0 dB SPL, while real AN fibers generally have thresholds that vary between fibers. Much of the AN physiology has been collected at levels that are on the steepest part of the rate-level function, usually 10-15 dB above rate threshold. The AN model can therefore be considered to provide physiologically realistic AM responses at levels below pure-tone rate saturation.

A.4.2 Model VCN cells: alternative mechanisms

Synchronous AM response features in the VCN are diverse and correlated with PST classification type (Rhode and Greenberg, 1994). Part of the diversity is a result of the fact that VCN neurons receive different numbers of inputs that act with various strengths and latencies. Despite the additional complications, some broad observations are clear: (1) VCN sync-MTFs can have low-pass or band-pass shapes, (2) sync-MTFs have lower corner frequencies than high-CF AN fibers, and (3) rate-MTFs remain relatively flat (with the exception of onset choppers; Rhode and Greenberg, 1994). We have shown that a simple two-input model bushy cell that receives membrane and synapse-modified inhibition and excitation with the same CF can exhibit low-pass or weakly band-pass sync-MTFs with corner frequencies lower than the model AN fiber sync-MTFs, depending on the relative strengths of excitation and inhibition. Also, the firing rate of the model VCN neurons is nearly independent of modulation frequency

when the inhibitory strength is less than 50-60% of the excitatory strength. Allowing the inhibition to become stronger than this threshold results in tuned rate-MTFs. Interestingly, Joris and Yin (1998) show that many of the globular bushy cells in their population exhibit systematic variation in average rate with f_m (their Fig. 13C).

The observation that model cells could be rate-tuned in the first layer of convergent SFIE leads to an important question: is the cascade of two post-synaptic model cells necessary to explain the transition from a synchrony code to a rate code? While the two-layer structure is not strictly required to understand rate tuning, it is useful as a tool to understand the flow of information along the auditory pathway. Our physiologically-based approach incorporates empirical data into the details of the model at each level of processing. By using data obtained from bushy cells in the VCN, we have described an alternative envelope information pathway that does not require populations of VCN choppers to understand the synchrony-to-rate transformation. However, interpretation of the current model structure is complicated by the fact that bushy cells in the VCN do not send axons directly to the IC; they project to the superior olivary complex (e.g. Oliver and Huerta, 1992). Neurons in the lateral superior olive (LSO) have low-pass sync-MTFs, similar to those of AN fibers and VCN bushy cells (Joris and Yin, 1998). The stereotypical rate-MTF in the LSO is also low-pass in shape, which is in contrast to the flat (all-pass) rate-MTFs measured in LSO afferents (Joris and Yin, 1998). However, the corner frequencies of monaural LSO rate-MTFs are significantly higher than those found in the IC; this suggests that the transformation that occurs at the level of the LSO is not crucial to generating the details of rate tuning in the IC. Physiological data in response to monaural AM tones in the medial superior olive (MSO) of cat and gerbil are not available.

Is the SFIE convergence mechanism for generating increased synchrony and band-pass sync-MTFs a unique solution? No; it is only a possible explanation for the physiological observations. While our model is physiologically plausible, it is also phenomenological in nature. An alternative hypothesis with significant physical support is that a coincidence-detection mechanism could also enhance envelope synchrony. A recent modeling study has suggested that VCN neurons could act as cross-frequency coincidence detectors on a shorter relative time scale (Heinz *et al.*, 2001b; Carney *et al.* 2002), and work continues to investigate the possibility that slower fluctuations could also be modified by such a mechanism in a manner consistent with physiology.

A.4.3 Model IC cells

The SFIE model described in this study is attractive for several reasons. Its primary appeal lies in its simplicity: a population of model IC cells, each receiving only two inputs, can span a wide range of rate-based BMFs. Individual cell rate tuning is determined primarily by differences in synaptic dynamics between excitatory and inhibitory inputs, and rate-MTFs can be further adjusted by varying the inhibitory delay. For the basic model with a single excitatory input and a single inhibitory input, overall rates are lower in cells with low BMFs, and peak rates drop again at high BMFs; this trend could be adjusted simply by changing the number of inputs to each cell. In other words, there is no inherent relationship between maximum rate and BMF. The high-synchrony and low-pass or all-pass sync-MTF features of the model cells are also

consistent with physiological measurements in units that respond strongly at the onset of pure-tone stimulation.

It is worthwhile to consider some details of the shapes of simulated IC cell rate-MTFs, and the model features that underlie them. One such detail is the sharp roll-off on the high-frequency side of the rate-MTF, and the more gradual rate increase with f_m on the low-frequency end. Both stereotyped traits of the model cells are attributable to the strength and speed of inhibition relative to excitation. For f_m well above BMF, both inhibition and excitation tend towards constant (tonic) values because membrane and synaptic time constants limit the extent to which envelope locking can occur. Since inhibition is stronger in this level of the model, zero firing is observed on the output for the high- f_m stimulus condition. For very slow amplitude fluctuations, the overall rate is low because both inhibitory and excitatory components can accurately follow the modulations of the VCN cell response, resulting in a high degree of overlap between the two inputs. As f_m is increased, the temporal interactions gradually change until an optimal frequency causes phase-locked excitatory peaks to line up in time with inhibitory valleys (at BMF).

A more quantitative way to compare rate-MTF shapes between model cells is to measure their Q value ($Q = \text{BMF} / \text{bandwidth at } 1/2 \text{ of the cell's peak rate}$). For the parameter spaces investigated, model cells were all observed to have rate-tuning Q values ≤ 1.2 at low to medium SPLs, with most measurements near 1. Although the metric was not reported in Krishna and Semple (2000), approximate calculations using their published rate-MTFs suggest similar physiological values. In the present model, effectively broader tuning can be caused by allowing inhibitory synaptic filtering properties to be significantly different than those of excitatory inputs ($\tau_{\text{IC,inh}} \gg \tau_{\text{IC,exc}}$). If we assume that the excitatory time constant is relatively fast and consistent across cells, this mechanism would predict that cells with lower BMFs would also have broader tuning.

Krishna and Semple (2000) found that a large proportion (~45%) of IC neurons exhibited a *suppressive* region in the rate-MTF. That is, the firing rate was depressed over some narrow range of f_m relative to rates at higher and lower stimulus modulation frequencies. The model presented here does not directly predict such a trend, but a simple scheme can be imagined in which suppressive regions can be created. Rate-tuned collaterals or ascending inputs (possibly from the VNLL; Batra, 2004) converging on a unit in the IC could provide inhibitory inputs to a post-synaptic cell. Krishna and Semple (2000) point out that for such a scheme to work, the net inhibitory effect would simply be proportional to the mean spike rate of the inputs. Another population of cells in the IC is characterized by a low-pass to band-pass sync-MTF shape transition with increasing SPL that is similar to some onset and chopper responders in the VCN (i.e. Krishna and Semple, 2000). The model presented here does not predict such a transformation; all model IC cell responses are highly synchronized to the modulation period, even at high SPLs.

Sinex *et al.* (2002) described recordings in the chinchilla IC in response to mistuned complex tones that have beating periodicities which are not present in the stimulus envelope. They put forth a simplified processing model that could account for some of these periodicities by combining information across different CFs. The model presented here would not predict such interactions, because model inputs are all from a

single CF. Given the complex connections (both ascending and descending) at the level of the IC, it is reasonable to assume that information is integrated across different frequencies. We have restricted excitatory and inhibitory inputs on a postsynaptic cell to have the same CF in an effort to keep the model as simple as possible, but this approach could be extended to include model IC cells that receive inhibition and excitation with different CFs.

A.4.4 Future directions

This study was a necessary first step in an effort to quantitatively relate physiological responses to human psychophysical performance in AM detection tasks. Recent psychophysical models of AM perception assume that a population of hard-wired filters tuned to f_m provides information about a signal's temporal envelope to higher processing centers (e.g. Dau *et al.*, 1997; Ewert *et al.*, 2002). This 'modulation filterbank' model structure is fundamentally different than that of previous models which assume that the net effect of central processing is to low-pass filter, or smooth, the envelope (e.g. Viemeister, 1979). The model IC cells presented here can be thought of as a physiological implementation of a modulation filter, and work continues to understand the relationship between model responses and perceptual modulation tuning.

Long-duration inhibitory influences in the model IC cells may also provide a physiologically-realistic mechanism for understanding psychophysical forward-masking phenomena. Offset suppression measured in AN fibers is too short to account for the time course of forward masking in humans (Smith, 1977; Harris and Dallos, 1979; Plack and Oxenham, 1998). Preliminary work suggests that the same f_m -tuned model IC cells presented here may effectively suppress responses to a probe signal that is presented after a masking stimulus for significantly longer masker-probe intervals. A physiological framework that integrates our understanding of modulation detection and forward masking is appealing.

An important test of the AN model will be to implement the newly-derived sharp human filter bandwidths (Shera *et al.*, 2002; Oxenham and Shera, 2003), and observe the effects on AM response properties. Sharper filters will result in lower envelope synchrony for a given carrier frequency, as AM sidebands will be more strongly attenuated. In this respect, some of the modulation encoding ability seems to be lost with narrower-band filters. From a different perspective, a sharp peripheral filter will result in significant modulation of wideband stimuli, emphasizing fluctuations in the filter's frequency pass band. Taken together, these observations illustrate the importance of using many stimuli and response quantifications when evaluating the performance of any model (especially one that simulates a highly nonlinear system such as the auditory system).

The main focus of future work on model VCN neurons will be to see if alternative processing mechanisms (i.e. coincidence detection) could also increase synchrony and underlie band-pass sync-MTFs. Coincidence detection can be approximated computationally using the simplified approach described in this work by taking the product of multiple input instantaneous firing rates as the response of a postsynaptic cell. Integrate-and-fire model neurons based on dynamic channel properties would provide a higher-order approximation of information integration at the level of VCN cells, which receive multiple excitatory AN inputs, as well as slower inhibitory inputs. Physiological

observations that estimate the frequency range of a cell's inputs, as well as the number of inputs, would be very useful in building a realistic VCN model cell.

Our model of processing in the IC should be extended to include cells that receive and integrate excitatory and inhibitory inputs with different CFs. This may allow the model described here to predict responses in the IC to mistuned complex tones (as in Sinex *et al.*, 2002) by combining resolved spectral information from different frequency bands. We should also consider whether long-lasting inhibition can be modeled simply with a single long IPSP, as opposed to a train of shorter and weaker IPSPs. This could be accomplished by directly comparing responses to the two IPSP configurations. Since the computational cost of convolution is quite high, carrying out this experiment should also motivate the development of a faster, frequency-domain algorithm that performs the same function of convolution in the time domain. Another possible direction for studying the fundamental mechanism of an additive combination of envelope-locked excitation and inhibition with different time courses and relative strengths would be to simplify the structure of the model presented here (as in Krukowski and Miller, 2001); a reduced parameter space might allow for more systematic variation of the key components of the 2nd-order (IC) model cells.

The techniques we have used to simulate the effects of membrane and synaptic filtering provide a first approximation of how information in the central auditory system may be modified as it ascends. While the computational methods described here are physiologically-based, they are by no means exhaustive. Perhaps most importantly, we assume that both inhibitory and excitatory contributions to a postsynaptic model cell are simply proportional to the instantaneous firing rate of a lower-level input. One way to refine the model would be to take into account the effects of membrane and synapse properties on a spike-by-spike basis. Implementing a conductance-based integrate-and-fire model neuron would allow us to include other realistic properties of cells in the auditory brainstem and midbrain. For example, some neurons may act as coincidence detectors; low-threshold potassium channels have been implicated as candidate mechanisms that could allow for such interactions between sequential inputs (Svirskis & Rinzel, 2003). Due to computational demands, it was not possible to include these effects in the present version of the model, but this is another clear direction for future work.

ACKNOWLEDGMENTS

We thank Xuedong Zhang for his help in implementing modifications to the AN model. Ray Meddis and an anonymous reviewer provided useful comments on an earlier version of this paper. This research was supported by NIH-NIDCD R01-01641 and the Jerome R. and Arlene L. Gerber Fund.

BIBLIOGRAPHY

- Aitkin, L. M., and Phillips, S. C. (1984). "Is the inferior colliculus an obligatory relay in the cat auditory system?" *Neurosci. Lett.* **44**, 259-264.
- Aitkin, L. (1991). "Rate-level functions of neurons in the inferior colliculus of cats measured with the use of free-field sound stimuli," *J. Neurophysiol.* **65**, 383-392.
- Anderson, D. J., Rose, J. E., Hind, J. E., and Brugge, J. F. (1971). "Temporal position of discharges in single auditory nerve fibers within the cycle of a sine-wave stimulus: frequency and intensity effects," *J. Acoust. Soc. Am.* **49**, 1131-1139.
- Atlas, L. E., and Shamma, S. A. (2003). "Joint acoustic and modulation frequency," *EURASIP J. Applied Signal Processing* **7**, 668-675.
- Bacon, S. P., and Grantham, D. W. (1989). "Modulation masking: effects of modulation frequency, depth, and phase," *J. Acoust. Soc. Am.* **85**, 2575-2580.
- Bacon, S. P., and Viemeister, N. F. (1985). "The temporal course of simultaneous tone-on-tone masking," *J. Acoust. Soc. Am.* **78**, 1231-1235.
- Bacon, S. P., and Viemeister, N. F. (1994). "Intensity discrimination and increment detection at 16 kHz," *J. Acoust. Soc. Am.* **95**, 2616-2621.
- Bartlett, E. L., and Wang, X. (2005). "Long-lasting modulation by stimulus context in primate auditory cortex," *J. Neurophysiol.* **94**, 83-104.
- Batra, R., Kuwada, S., and Stanford, T. R. (1989). "Temporal coding of envelopes and their interaural delays in the inferior colliculus of the unanesthetized rabbit," *J. Neurophysiol.* **61**, 257-268.
- Batra, R. (2004). "Responses to amplitude-modulated tones of neurons in the ventral nucleus of the lateral lemniscus of the unanesthetized rabbit," *Association for Research in Otolaryngology Abstracts*, **27**:914.
- Bibikov, N. G. (2002). "Addition of noise enhances neural synchrony to amplitude-modulated sounds in the frog's midbrain," *Hear. Res.* **173**, 21-38.
- Blackburn, C. C., and Sachs, M. B. (1989). "Classification of unit types in the anteroventral cochlear nucleus: PST histograms and regularity analysis," *J. Neurophysiol.* **62**, 1303-1329.
- Bregman, A. S. (1990). *Auditory Scene Analysis*. Cambridge, MA: MIT Press.

- Bullock, D. C., Palmer, A. R., and Rees, A. (1988). "Compact and easy-to-use tungsten-in-glass microelectrode manufacturing work station," *Med. & Biol. Eng. & Comput.* **26**, 669-672.
- Burger, R. M., and Pollak, G. D. (1998). "Analysis of the role of inhibition in shaping responses to sinusoidally amplitude-modulated signals in the inferior colliculus," *J. Neurophysiol.* **80**, 1686-1701.
- Brosch, M., and Schreiner, C. E. (1997). "Time course of forward masking tuning curves in cat primary auditory cortex," *J. Neurophysiol.* **77**, 923-943.
- Buus, S. (1985). "Release from masking caused by envelope fluctuations," *J. Acoust. Soc. Am.* **78**, 1958-1965.
- Calford, M. B., and Semple, M. N. (1995). "Monaural inhibition in cat auditory cortex," *J. Neurophysiol.* **73**, 1876-1891.
- Campbell, R. A., and Lasky, E. Z. (1967). "Masker level and sinusoidal-signal detection," *J. Acoust. Soc. Am.* **5**, 972-976.
- Carney, L. H., and Yin, T. C. T. (1989). "Responses of low-frequency cells in the inferior colliculus to interaural time differences of clicks: Excitatory and inhibitory components," *J. Neurophysiol.* **62**, 144-161.
- 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., Heinz, M. G., Evilsizer, M. E., Gilkey, R. H., and Colburn, H. S. (2002). "Auditory phase opponency: A temporal model for masked detection at low frequencies," *Acta Acustica United with Acustica.* **88**, 334-347.
- Casparly, D. M., Backoff, P. M., Finlayson, P. G., and Palombi, P. S. (1994). "Inhibitory inputs modulate discharge rate within frequency receptive fields of anteroventral cochlear nucleus neurons," *J. Neurophysiol.* **72**, 2124-2132.
- Casparly, D. M., Palombi, P. S., and Hughes, L. F. (2002). "GABAergic inputs shape responses to amplitude modulated stimuli in the inferior colliculus," *Hear. Res.* **168**, 163-173.
- Chatterjee, M., and Robert, M. E. (2001). "Noise enhances modulation sensitivity in cochlear implant listeners: stochastic resonance in a prosthetic sensory system?," *J. Assoc. Res. Otolaryngol.* **2**, 159-171.

- Colburn, H. S., Carney, L. H., and Heinz, M. G. (2003). "Quantifying the information in auditory-nerve responses for level discrimination," *J. Assoc. Res. Otolaryngol.* **4**, 294-311.
- Cooper, N. P., Robertson, D., and Yates, G. K. (1993). "Cochlear nerve fiber responses to amplitude-modulated stimuli: variations with spontaneous rate and other response characteristics," *J. Neurophysiol.* **70**, 370-386.
- Creutzfeldt, O. D., Hellweg, F. C., and Schreiner, C. E. (1980). "Thalamocortical transformation of responses to complex auditory stimuli," *Exp. Brain Res.* **39**, 87-104.
- Dai, H. (1995). "On measuring psychometric functions: A comparison of the constant-stimulus and adaptive up-down methods," *J. Acoust. Soc. Am.* **98**, 3135-3139.
- Dau, T., Kollmeier B., and Kohlrausch, A. (1997). "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. (1997). "Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration," *J. Acoust. Soc. Am.* **102**, 2906-2619.
- Dau, T., Verhey, J., and Kohlrausch, A. (1999). "Intrinsic envelope fluctuations and modulation-detection thresholds for narrow-band noise carriers," *J. Acoust. Soc. Am.* **106**, 2752-2760.
- Delgutte, B. (1987). "Peripheral auditory processing of speech information: implications from a physiological study of intensity discrimination," in *The Psychophysics of Speech Perception*, ed. M. E. H. Schouten (Dordrecht, Nijhoff), pp. 333-353.
- Derleth, R. P., and Dau, T. (2000). "On the role of envelope fluctuation processing in spectral masking," *J. Acoust. Soc. Am.* **108**, 285-296.
- Drullman, R. (1995). "Temporal envelope and fine structure cues for speech intelligibility," *J. Acoust. Soc. Am.* **97**, 585-592.
- Durlach, N. I., and Braida, L. D. (1969). "Preliminary theory of intensity resolution," *J. Acoust. Soc. Am.* **46**, 372-383.
- Eggermont, J. J. (1994). "Temporal modulation transfer functions for AM and FM stimuli in cat auditory cortex. Effects of carrier type, modulating waveform, and intensity," *Hear. Res.* **74**, 51-66.
- Elhilali M., Fritz, J. B., Klein, D. J., Simon, J. Z., and Shamma, S. A. (2004). "Dynamics of precise spike timing in primary auditory cortex," *J. Neurosci.* **24**, 1159-1172.

- Ewert, S. D., and Dau, T. (2000). "Characterizing frequency selectivity for envelope fluctuations," *J. Acoust. Soc. Am.* **108**, 1181-1196.
- Ewert, S. D., Verhey, J.L., and Dau, T. (2002). "Spectro-temporal processing in the envelope-frequency domain," *J. Acoust. Soc. Am.* **112**, 2921-2931.
- Ewert, S. D., and Dau, T. (2004). "External and internal limitations in amplitude-modulation processing," *J. Acoust. Soc. Am.* **116**, 478-490.
- Faure, P. A., Fremouw, T., Casseday, J. H., and Covey, E. (2003). "Temporal masking reveals properties of sound-evoked inhibition in duration-tuned neurons of the inferior colliculus," *J. Neurosci.* **24**, 3052-3065.
- Fleischer, H. (1982). "Modulationsschwellen von Schmalbandrauschen," *Acustica* **51**, 154-161.
- 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.
- Frisina, R. D., Smith, R. L., and Chamberlain, S. C. (1985). "Differential encoding of rapid changes in sound amplitude by second-order auditory neurons," *Exp. Brain Res.* **60**, 417-422.
- Frisina, R. D., Smith, R. L. and Chamberlain, S. C. (1990). "Encoding of amplitude modulation in the gerbil cochlear nucleus. I. A hierarchy of enhancement," *Hear. Res.* **44**, 99-122.
- Füllgrabe, C., and Lorenzi, C. (2003). "The role of envelope beat cues in the detection and discrimination of second-order amplitude modulation (L)," *J. Acoust. Soc. Am.* **113**, 49-52.
- Füllgrabe, C., Moore, B. C. J., Demany, L., Ewert, S. D., Sheft, S., and Lorenzi, C. (2005). "Modulation masking produced by second-order modulators," *J. Acoust. Soc. Am.* **117**, 2158-2168.
- Füllgrabe, C., and Lorenzi, C. (2005). "Perception of the envelope-beat frequency of inharmonic complex temporal envelopes," *J. Acoust. Soc. Am.* **118**, 3757-3765.
- Gallun, E., and Hafter, E. R. (**in review**). *J. Acoust. Soc. Am.*
- Green, D. M. (1969). "Masking with continuous and pulsed sinusoids," *J. Acoust. Soc. Am.* **46**, 939-946.
- Green, D.M. (1983). "Profile analysis. A different view of auditory intensity discrimination," *Am. Psychol.* **38**, 133-142.

- Greenberg, S., and Arai, T. (2001). "The relation between speech intelligibility and the complex modulation spectrum," Proc. 7th Eur. Conf. on Speech Comm. pp. 473-476.
- Goldberg, J. M., and Brown, P. B. (1969). "Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization," J. Neurophysiol. **32**, 613-636.
- Grothe, B. (1994). "Interaction of excitation and inhibition in processing of pure tone and amplitude-modulated stimuli in the medial superior olive of the mustached bat," J. Neurophysiol. **71**, 706-721.
- Haftner, E. R., Bonnel, A. M., and Gallun, E. (1997). "A role for memory in divided attention between two independent stimuli," in *Psychophysical and Physiological Advances in Hearing*, ed. A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London), pp. 228-237.
- Hall, J. W., Haggard, M. P. and Fernandes, M. A. (1984). "Detection in noise by spectro-temporal pattern analysis," J. Acoust. Soc. Am. **76**, 50-56.
- Hall, J. W., Buss, E., and Grose, J. H. (2003). "Modulation rate discrimination for unresolved components: temporal cues related to fine structure and envelope," J. Acoust. Soc. Am. **113**, 986-993.
- Harris, J. D. (1963). "Loudness discrimination," J. Speech Hear. Disord. Monogr. Suppl. **11**, 1-59.
- Harris, D. M., and Dallos, P. (1979). "Forward masking of auditory nerve fiber responses," J. Neurophysiol. **42**, 1083-1107.
- Heil, P., Schulze, H., and Langner, G. (1995). "Ontogenetic development of periodicity coding in the inferior colliculus of the Mongolian gerbil," *Audit. Neurosci.* **1**, 363-383.
- Heinz, M. G., Zhang, X., Bruce, I. C., and Carney, L. H. (2001). "Auditory-nerve model for predicting performance limits of normal and impaired listeners," J. Assoc. Res. Otolaryngol. **2**, 91-96.
- Heinz, M. G., Colburn, H. S., and Carney, L. H. (2001). "Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve," *Neural Comput.* **13**, 2273-2316.
- Heinz, M. G., Colburn, H. S., and Carney, L. H. (2001). "Rate and timing cues associated with the cochlear amplifier: Level discrimination based on monaural cross-frequency coincidence detection," J. Acoust. Soc. Am. **110**, 2065-2084.

- Hewitt, M. J., and Meddis, R. (1994). "A computer model of amplitude-modulation sensitivity of single units in the inferior colliculus," *J. Acoust. Soc. Am.* **95**, 2145-2159.
- Houtgast, T. (1989). "Frequency selectivity in amplitude-modulation detection," *J. Acoust. Soc. Am.* **85**, 1676-1680.
- Irwin, R. J., and Purdy, S. C. (1982). "The minimum detectable duration of auditory signals for normal and hearing-impaired listeners," *J. Acoust. Soc. Am.* **71**, 967-974.
- Joris, P. X., and Yin, T. C. T. (1992). "Responses to amplitude-modulated tones in the auditory nerve of the cat," *J. Acoust. Soc. Am.* **91**, 215-232.
- Joris, P. X., and Yin, T. C. T. (1998). "Envelope coding in the lateral superior olive. III. Comparison with afferent pathways," *J. Neurophysiol.* **79**, 253-269.
- Joris, P. X., Schreiner, C. E., and Rees, A. (2004). "Neural processing of amplitude-modulated sounds," *Physiol. Rev.* **84**, 541-577.
- Kay, R. H., and Matthews, D. R. (1972). "On the existence in human auditory pathways of channels selectively tuned to the modulation present in frequency-modulated tones," *J. Physiol.* **225**, 657-677.
- Khanna, S. M., and Teich, M. C. (1989). "Spectral characteristics of the responses of primary auditory-nerve fibers to amplitude-modulated signals," *Hear. Res.* **39**, 143-157.
- 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," *Research Monographs No. 35* (MIT, Cambridge, MA).
- 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**, 1340-1354.
- Kohlrausch, A., Fassel, R., and Dau, T. (2000). "The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers," *J. Acoust. Soc. Am.* **108**, 723-734.
- Krishna, B. S., and Semple, M. N. (2000). "Auditory temporal processing: Responses to sinusoidally amplitude-modulated tones in the inferior colliculus," *J. Neurophysiol.* **84**, 255-273.

- Krukowski, A.E., and Miller, K.D. (2001). "Thalamocortical NMDA conductances and intracortical inhibition can explain cortical temporal tuning," *Nat. Neuro.* **4**, 424-430.
- Kuwada, S., Stanford, T. R., and Batra, R. (1987). "Interaural phase-sensitive units in the inferior colliculus of the unanesthetized rabbit: effects of changing frequency," *J. Neurophysiol.* **57**, 1338-1360.
- Kuwada, S., and Batra, R. (1999). "Coding of sound envelopes by inhibitory rebound in neurons of the superior olivary complex in the unanesthetized rabbit," *J. Neurosci.* **19**, 2273-2287.
- Langner, G. (1981). "Neuronal mechanisms for pitch analysis in the time domain," *Exp. Brain. Res.* **44**, 450-454.
- Langner, G., and Schreiner, C. E. (1988). "Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms," *J. Neurophysiol.* **60**, 1799-1822.
- Langner, G., Schreiner, C. E., and Merzenich, M. M. (1987). "Covariation of latency and temporal resolution in the inferior colliculus of the cat," *Hear. Res.* **31**, 197-201.
- Le Beau, F. E., Rees, A., and Malmierca, M. S. (1996). "Contribution of GABA- and glycine-mediate inhibition to the monaural temporal response properties of neurons in the inferior colliculus," *J. Neurophysiol.* **75**, 902-919.
- Lee, J., and Bacon, S. P. (1997). "Amplitude modulation depth discrimination of a sinusoidal carrier: Effect of stimulus duration," *J. Acoust. Soc. Am.* **101**, 3688-3693.
- Leshowitz, B., and Cudahy, E. (1972). "Masking with continuous and gated sinusoids," *J. Acoust. Soc. Am.* **58**, 235-242.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467-477.
- Liang, L., Lu, T., and Wang, X. (2002). "Neural representations of sinusoidal amplitude and frequency modulations in the primary auditory cortex of awake primates," *J. Neurophysiol.* **87**, 2237-2261.
- Liberman, M. C. (1978). "Auditory-nerve response from cats raised in a low-noise chamber," *J. Acoust. Soc. Am.* **63**, 442-455.
- Lorenzi, C., Berthommier, F., and Demany, L. (1999). "Discrimination of amplitude-modulation phase spectrum," *J. Acoust. Soc. Am.* **105**, 2987-2990.

- Lorenzi, C., Soares, C., Vonner, T. (2001). "Second-order temporal modulation transfer functions," *J. Acoust. Soc. Am.* **110**, 1030-1038.
- Lorenzi, C., Simpson, M. I. G., Millman, R. E., Griffiths, T. D., Woods, W. P., Rees, A., and Green, G. G. (2001). "Second-order modulation detection thresholds for pure-tone and narrow-band noise carriers," *J. Acoust. Soc. Am.* **110**, 2470-2478.
- Luscher, E., and Zwislocki, J. (1949). "Adaptation of the ear to sound stimuli," *J. Acoust. Soc. Am.* **21**, 135-139.
- Macmillan, N. A. (1971). "Detection and recognition of increments and decrements in auditory intensity," *Percept. Psychophys.* **10**, 233-238.
- Malmierca, M. S., Merchan, M. A., Henkel, C. K., and Oliver, D. L. (2002). "Direct projections from cochlear nucleus complex to auditory thalamus in the rat," *J. Neurosci.* **22**, 10891-10897.
- Miller, R. L., Schilling, J. R., Franck, K. R., and Young, E. D. (1997). "Effects of acoustic trauma on the representation of the vowel /ε/ in cat auditory nerve fibers," *J. Acoust. Soc. Am.* **101**, 3602-3616.
- Moore, B. C. J., Alcantara, J. I., and Dau, T. (1998). "Masking patterns for sinusoidal and narrowband noise maskers," *J. Acoust. Soc. Am.* **104**, 1023-1038.
- Moore, B. C. J., Sek, A., and Glasberg, B. R. (1999). "Modulation masking produced by beating modulators," *J. Acoust. Soc. Am.* **106**, 938-945.
- Moore, B. C. J., and Sek, A. (2000). "Effects of relative phase and frequency spacing on the detection of three-component amplitude modulation," *J. Acoust. Soc. Am.* **108**, 2337-2344.
- Mueller-Preuss, P., Flachskamm, C., and Bieser, A. (1994). "Neural encoding of amplitude modulation within the auditory midbrain of squirrel monkeys," *Hear. Res.* **80**, 197-208.
- Nagarajan, S. S., Cheung, S. W., Bedenbaugh, P., Beitel, R. E., Schreiner, C. E., and Merzenich, M. M. (2002). "Representation of spectral and temporal envelope of twitter vocalizations in common marmoset primary auditory cortex," *J. Neurophysiol.* **87**, 1723-1737.
- Nelson, P. B., Jin, S. H., Carney, A. E., and Nelson, D. A. (2003). "Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners," *J. Acoust. Soc. Am.* **113**, 961-968.

- Nelson, P. B., and Jin, S. H. (2004). "Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners," *J. Acoust. Soc. Am.* **115**, 2286-2294.
- Nelson, P. G., Erulkar, S. D., and Bryan, J. S. (1966). "Responses of units in the inferior colliculus to time-varying acoustic stimuli," *J. Neurophysiol.* **29**, 834-860.
- Nelson, P. C., and Carney, L. H. (2004). "A phenomenological model of peripheral and central neural responses to amplitude-modulated tones," *J. Acoust. Soc. Am.* **116**, 2173-2186.
- Nelson, P. C., and Carney, L. H. (submitted). "Cues for masked amplitude-modulation detection," *J. Acoust. Soc. Am.*
- Oertel, D. (1983). "Synaptic responses and electrical properties of cells in brain slices of the mouse anteroventral cochlear nucleus," *J. Neurosci.* **3**, 2043-2053.
- Oliver, D. L., and Huerta, M. F. (1992). "Inferior and superior colliculi," in *The Mammalian Auditory Pathway: Neuroanatomy*, edited by D. R. Webster, A. N. Popper, and R. R. Fay (Springer Verlag, New York), pp. 168-221.
- Oxenham, A. J., and Shera, C. A. (2003). "Estimates of human cochlear tuning at low levels using forward and simultaneous masking," *JARO* **4**, 541-554.
- Palombi, P. S., Caspary, D. M. (1996). "GABA inputs control discharge rate primarily within frequency receptive fields of inferior colliculus neurons," *J. Neurophysiol.* **75**, 2211-2219.
- Plack, C. J., and Oxenham, A. J. (1998). "Basilar-membrane nonlinearity and the growth of forward masking," *J. Acoust. Soc. Am.* **103**, 1598-1608.
- Ramachandran, R., Davis, K. A., and May, B. J. (1999). "Rate representation of tones in noise in the inferior colliculus of decerebrate cats," *J. Assoc. Res. Otolaryngol.* **1**, 144-160.
- Ramon y Cajal, S. (1904). "Textura del Sistema Nervioso del Hombre y de los Vertebrados," Nicolas Moya, Madrid.
- Rees, A. and Moller, A. R. (1983). "Responses of neurons in the inferior colliculus of the rat to AM and FM tones," *Hear. Res.* **10**, 301-330.
- Rhode, W. S. (1994). "Temporal coding of 200% amplitude modulated signals in the ventral cochlear nucleus of the cat," *Hear. Res.* **77**, 43-68.
- Rhode, W. S., and Greenberg, S. (1994). "Encoding of amplitude modulation in the cochlear nucleus of the cat," *J. Neurophysiol.* **71**, 1797-1825.

- Richards, V. M., and Nekrich R. D. (1993). "The incorporation of level and level-invariant cues for the detection of a tone added to noise," *J. Acoust. Soc. Am.* **94**, 2560-2574.
- Ryugo, D. K., and Parks, T. N. (2003). "Primary innervation of the avian and mammalian cochlear nucleus," *Brain Res. Bull.* **60**, 435-456.
- Sachs, M. B., Voigt, H. F., and Young, E. D. (1983). "Auditory nerve representation of vowels in background noise," *J. Neurophysiol.* **50**, 27-45.
- Saldana, E., and Merchan, M. A. (2004). "Intrinsic and commissural connections of the inferior colliculus," in *The Inferior Colliculus*, eds. J. A. Winer and C. E. Schreiner (Springer, New York), pp. 155-181.
- Salvi, R. J., Giraudi, D. M., Henderson, D., and Hamernik, R. P. (1982). "Detection of sinusoidally amplitude modulated noise by the chinchilla," *J. Acoust. Soc. Am.* **71**, 424-429.
- Schneiderman, A., Oliver, D. L., and Henkel, C. K. (1988). "Connections of the dorsal nucleus of the lateral lemniscus: an inhibitory parallel pathway in the ascending auditory system?," *J. Comp. Neurol.* **276**, 188-208.
- Schofield, B. R., and Cant, N. B. (1996). "Projections from the ventral cochlear nucleus to the inferior colliculus and the contralateral cochlear nucleus in guinea pigs," *Hear. Res.* **102**, 1-14.
- Schooneveldt, G. P., and Moore, B. C. J. (1989). "Comodulation masking release (CMR) as a function of masker bandwidth, modulator bandwidth, and signal duration," *J. Acoust. Soc. Am.* **85**, 273-281.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J. and Elelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**, 303-304.
- Sheft, S., and Yost, W. A. (1990). "Temporal integration in amplitude-modulation detection," *J. Acoust. Soc. Am.* **88**, 796-805.
- Shera, C. A., Guinan, J. J. Jr., and Oxenham, A. J. (2002). "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements," *Proc. Natl. Acad. Sci.* **99**, 3318-3323.
- Shofner, W. P., Sheft, S., and Guzman, S. J. (1996). "Responses of ventral cochlear nucleus units in the chinchilla to amplitude modulation by low-frequency, two-tone complexes," *J. Acoust. Soc. Am.* **99**, 3592-3605.

- Siebert, W. M. (1965). "Some implications of the stochastic behavior of primary auditory neurons," *Kybernetik* **2**, 206-215.
- Siebert, W. M. (1968). "Stimulus transformations in the peripheral auditory system," in *Recognizing Patterns* ed. P. A. Kolars (MIT Press, Cambridge, MA), pp. 104-133.
- Sinex, D. G., Sabes, J. H., and Li, H. (2002). "Responses of inferior colliculus neurons to harmonic and mistuned complex tones," *Hear. Res.* **168**, 150-162.
- Smith, R. L., and Zwislocki, J. J. (1975). "Short-term adaptation and incremental responses in single auditory-nerve fibers," *Biol. Cybern.* **17**, 169-182.
- Smith, R. L. (1977). "Short-term adaptation in single auditory nerve fibers: Some post-stimulatory effects," *J. Neurophysiol.* **40**, 1098-1111.
- Smith, R. L. (1979). "Adaptation, saturation, and physiological masking in single auditory-nerve fibers," *J. Acoust. Soc. Am.* **65**, 166-178.
- Smith, R. L., and Brachman, M. L. (1980). "Response modulation of auditory-nerve fibers by AM stimuli: Effects of average intensity," *Hear. Res.* **2**, 123-133.
- Smith, R. L., and Brachman, M. L. (1982). "Adaptation in auditory-nerve fibers: A revised model," *Biol. Cybern.* **44**, 107-120.
- Smith, R. L., Brachman, M. L., and Frisina, R. D. (1985). "Sensitivity of auditory-nerve fibers to changes in intensity: A dichotomy between decrements and increments," *J. Acoust. Soc. Am.* **78**, 1310-1316.
- Spinks, R. L., Baker, S. N., Jackson, A., Khaw, P. T., and Lemon, R. N. (2003). "Problem of dural scarring in recording from awake, behaving monkeys: a solution using 5-Fluorouracil," *J. Neurophysiol.* **90**, 1324-1332.
- Strickland, E. A., and Viemeister, N. F. (1996). "Cues for discrimination of envelopes," *J. Acoust. Soc. Am.* **99**, 3638-3646.
- Sullivan, W. E. 3rd (1982). "Possible neural mechanisms of target distance coding in the auditory system of the echolocating bat *Myotis lucifugus*," *J. Neurophysiol.* **48**, 1033-1047.
- Svirskis, G., and Rinzel, J. (2003). "Influence of subthreshold nonlinearities on signal-to-noise ratio and timing precision for small signals in neurons: Minimal model analysis," *Network: Comput. Neural Syst.* **14**, 137-150.
- Tansley, B. W., and Suffield, J. B. (1983). "Time course of adaptation and recovery of channels selectively sensitive to frequency and amplitude modulation," *J. Acoust. Soc. Am.* **74**, 765-775.

- Theunissen, F. E., and Doupe, A. J. (1998). "Temporal and spectral sensitivity of complex auditory neurons in the nucleus HVC of male zebra finches," *J. Neurosci.* **18**, 3786-3802.
- Tougaard, J. (2000). "Stochastic resonance and signal detection in an energy detector – implications for biological receptor systems," *Biol. Cybern.* **83**, 471-480.
- Ulanovsky, N., Las, L., Farkas, D., and Nelken, I. (2004). "Multiple time scales of adaptation in auditory cortex neurons," *J. Neurosci.* **24**, 10440-10453.
- Verhey, J. L., Dau, T., and Kollmeier, B. (1999). "Within-channel cues in comodulation masking release (CMR): Experiments and model predictions using a modulation-filterbank model," *J. Acoust. Soc. Am.* **106**, 2733-2745.
- Viemeister, N. F. (1972). "Intensity discrimination of pulsed sinusoids: The effects of filtered noise," *J. Acoust. Soc. Am.* **51**, 1265-1269.
- Viemeister, N. F. (1979). "Temporal modulation transfer functions based upon modulation thresholds," *J. Acoust. Soc. Am.* **66**, 1364-1380.
- Viemeister, N. F. (1988). "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.
- Viemeister, N. F., and Wakefield, G. H. (1991). "Temporal integration and multiple looks," *J. Acoust. Soc. Am.* **90**, 858-865.
- Wakefield, G. H., and Viemeister, N. F. (1990). "Discrimination of modulation depth of sinusoidal amplitude modulation (SAM) noise," *J. Acoust. Soc. Am.* **88**, 1367-1373.
- Ward, L. M., Neiman, A., and Moss, F. (2002). "Stochastic resonance in psychophysics and animal behavior," *Biol. Cybern.* **87**, 91-101.
- Warr, W. B. (1982). "Parallel ascending pathways from the cochlear nucleus: neuroanatomical evidence of functional specialization," In *Contributions to Sensory Physiology*, ed. W. D. Neff (Academic Press, New York), pp. 1-38.
- Wegel, R. L., and Lane, C. E. (1924). "The auditory masking of one sound by another and its probable relation to the dynamics of the inner ear," *Phys. Rev.* **23**, 266-285.

- 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.
- Wickesberg, R. E., and Oertel, D. (1988). "Tonotopic projection from the dorsal to the anteroventral cochlear nucleus of mice," *J. Comp. Neurol.* **268**, 389-299.
- Wiesenfeld, K., and Jaramillo, F. (1998). "Minireview of stochastic resonance," *Chaos* **8**, 539-548.
- Winer, J. A., Larue, D. T., and Pollak, G. D. (1995). "GABA and glycine in the central auditory system of the mustache bat: structural substrates for inhibitory neuronal organization," *J. Comp. Neurol.* **355**, 317-353.
- Winer, J. A. (2004). "Three systems of descending projections to the inferior colliculus," in *The Inferior Colliculus*, eds. J. A. Winer and C. E. Schreiner (Springer, New York), pp. 231-247.
- 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**, 165-189.
- 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.
- Wojtczak, M., and Viemeister, N. F. (1999). "Intensity discrimination and detection of amplitude modulation," *J. Acoust. Soc. Am.* **106**, 1917-1924.
- Wojtczak, M., and Viemeister, N. F. (2005). "Forward masking of amplitude modulation: Basic characteristics," *J. Acoust. Soc. Am.* **118**, 3198-3210.
- Wu, S. H., Ma, C. L., Sivaramakrishnan, S., Oliver, D. L. (2002). "Synaptic modification in neurons of the central nucleus of the inferior colliculus," *Hear. Res.* **168**, 43-54.
- Yin, T. C. T. (2002). "Neural mechanisms of encoding binaural localization cues in the auditory brainstem," in *Integrative Functions in the Mammalian Auditory Pathway*, edited by D. Oertel, A. N. Popper, and R. R. Fay (Springer Verlag, New York), pp. 99-159.
- Yost, W. A., and Sheft, S. (1997). "Temporal modulation transfer functions for tonal stimuli: Gated versus continuous conditions," *Aud. Neurosci.* **3**, 401-414.
- Young, E. D. and Barta, P. E. (1986). "Rate responses of auditory nerve fibers to tones in noise near masked threshold," *J. Acoust. Soc. Am.* **79**, 426-442.

Zhang, H., and Kelly, J. B. (2003). "Glutamatergic and GABAergic regulation of neural responses in the inferior colliculus to amplitude modulated sounds," *J. Neurophysiol.* **90**, 477-490.

Zhang, X., Heinz, M. G., Bruce, I. C., and Carney, L. H. (2001). "A phenomenological model for the responses of auditory-nerve fibers: I. Non-linear tuning with compression and suppression," *J. Acoust. Soc. Am.* **109**, 648-670.

Zhang, X. (2004). Personal communication.

Zwicker, E. (1952). "Die Grenzen der Hörbarkeit der Amplitudenmodulation und der Frequenzmodulation eines Tones," *Acustica, Akust. Beih.* **2**, AB125-AB133.

VITA

NAME OF AUTHOR: Paul Christian Nelson

PLACE OF BIRTH: Saint Cloud, Minnesota

DATE OF BIRTH: February 22, 1979

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

Syracuse University, Syracuse, New York

Boston University, Boston, Massachusetts

DEGREES AWARDED:

Master of Science in Bioengineering, 2003, Syracuse University

Bachelor of Science in Biomedical Engineering, 2001, Boston University

AWARDS AND HONORS:

Ruth L. Kirchstein National Research Service Award, 2005-2006

PROFESSIONAL EXPERIENCE:

Research Assistant, Department of Biomedical Engineering,
Boston University, 2001

Research Assistant, Department of Bioengineering and Neuroscience,
Syracuse University, 2001-2006