Detection of Tones in Reproducible Noises:
Prediction of Listeners' Performance
in Diotic and Dichotic Conditions

by

Junwen Mao

Submitted in Partial Fulfillment of the
Requirements for the Degree
Doctor of Philosophy

Supervised by Professor Laurel H. Carney

Department of Electrical and Computer Engineering
Arts, Sciences and Engineering
Edmund A. Hajim School of Engineering and Applied Sciences

University of Rochester
Rochester, New York
2013

Biographical Sketch

Junwen Mao was born in China. She attended Zhejiang University, and graduated with the Bachelor of Science degree in Measurement and Control Techniques & Instruments in 2007. She began her doctoral studies in Electrical and Computer Engineering at the University of Rochester in 2007. She received the Master of Science degree in Electrical and Computer Engineering from the University of Rochester in 2009. She pursued her research in understanding how listeners detect tones in noisy environments under the direction of Dr. Laurel H. Carney.

The following publications were a result of work conducted during doctoral study:

Journal Publications:

"Predictions of Diotic Tone-in-Noise Detection Based on a Nonlinear Optimal Combination of Energy, Envelope, and Fine-Structure Cues," Junwen Mao, Azadeh Vosoughi, and Laurel H. Carney, *Journal of the Acoustical Society of America*, vol. 134, No. 1, pp. 396-406, July 2013.

Conference Publications:

"Effects of Sensorineural Hearing Loss on Roving-Level Tone-in-Noise Detection," Junwen Mao, Karen A. Doherty, Kelly-Jo Kock, and Laurel H. Carney, *Conference of the American Auditory Society*, Scottsdale, AZ, March 2013.

“Physiologically-based Envelope Cues for Diotic and Dichotic Tone-in-Noise Detection,” Junwen Mao and Laurel H. Carney, *Conference of Association for Research in Otolaryngology*, Abstract: pp. 87, Baltimore, MD, February 2013.

“Modeling Detection of 500-Hz Tones in Reproducible Noise for Listeners with Sensorineural Hearing Loss,” Laurel H. Carney, Junwen Mao, Kelly-Jo Koch, and Karen A. Doherty, *Proceedings of Meetings on Acoustics*, vol. 19, Montreal, Canada, June 2013.

“The Budgerigar as a Model for Human Detection of Tones in Noise,” Laurel H. Carney, Kristina S. Abrams, Junwen Mao, Douglas M. Schwarz, and Fabio Idrobo, *Conference of the Society for Neurosicence*, New Orleans, LA, October 2012.

“Stimulus-based Diotic and Dichotic Models that Combine Cues for Detection of Tones in Reproducible Noise,” Junwen Mao, Azadeh Vosoughi, and Laurel H. Carney, *Conference of the Acoustical Society of America*, Abstract: *Journal of the Acoustical Society of America*, vol. 129, pp. 2489, Seattle, WA, May 2011.

“Detection of Tones in Reproducible Noises: Combining Information across Epochs and across Cues,” Junwen Mao and Laurel H. Carney, Conference of *of Association for Research in Otolaryngology*, Abstract: pp. 114, Anaheim, CA, February 2010.

Acknowledgments

First and foremost, I would like to thank my advisor, Prof. Laurel Carney, for her guidance, support and help during my graduate study. She has spent countless hours to advise all aspects of my thesis, starting from identifying research directions, searching for the best solutions, and all the way through technical writing. Laurel is not only an excellent advisor for my thesis, but also a good mentor and friend for my personal growth. I am deeply indebted to her for my great experience of graduate study. Her passion for work, enthusiasm for life, and kindness for people will always inspire me.

I am also very grateful to have four wonderful committee members. My thesis work started with Prof. Mark Bocko's help in introducing me to Prof. Laurel Carney. Prof. Azadeh Vosoughi provided insightful ideas during our collaboration, which led to a part of this thesis. Prof. Jack Mottley and Prof. William O’Neill have also always been available for matters about my thesis. I'm very thankful to my committee members for the feedback and suggestions to improve my thesis.

During my graduate study, I am fortunate to have many great labmates. I learned a lot from Dr. Muhammad Zilany. Without the help from Kelly-Jo Kock, I would not have finished the experiments so smoothly. Dr. Tianhao Li, not only a good friend, but also inspires me by her passion for research. Kristina Abrams and Akshay Rao are both good labmates and friends. I would like also to thank current and former labmates, Douglas Schwarz, Dr. Kenneth Henry, and Nicholas Huang.

Last, but not least, I thank my parents and husband for their unconditional love and support in my life. Without them, I would not have finished this long journey.

Abstract

Detection of tones in reproducible noises, a set of pre-generated random noises, has been studied for decades. These studies help us to understand how people detect signals in noise in everyday life. However, it is not clear what cues or combination of cues are used by listeners in these tasks. Previous studies have shown that energy and temporal cues could predict a significant amount of the variance in listeners' detection performance in the diotic condition, in which identical noise-alone and tone-plus-noise stimuli were presented at both ears. For the dichotic condition, in which identical noise and out-of-phase tones were presented, interaural level and time difference cues, and combinations of these two cues partially explain listeners' performance.

In this thesis, an optimal cue-combination model was proposed to explain listeners' performance in the diotic condition. This model combined energy and temporal cues nonlinearly, based on the logarithmic likelihood-ratio test. Predictions from this model explained a substantial amount of the variance in listeners' performance from three different sets of reproducible noises.

For the dichotic condition, two different models were proposed: one based on a linear combination of interaural level and time difference cues that included the relation between these two cues, and the other using a binaural envelope cue (slope of the interaural envelope difference). For the wideband noise condition, both models explained significant amounts of the variance in listeners' performance. In particular, predictions from the binaural envelope cue were significantly better than predictions from any available model. For the narrowband noise condition, it is likely that different listeners used envelope information from different frequency channels to detect tones in noise.

Finally, given the robustness of envelope cues in diotic and dichotic conditions, we investigated the reliability of physiological envelope cues in predicting listeners' performance. Responses from model inferior colliculus cells were analyzed in terms of average rate and response fluctuations. For diotic and dichotic conditions, predictions from the physiological envelope cues can explain a similar or larger amount of the variance in listeners' performance than stimulus-based envelope cues. Similar to results from the stimulus-based envelope cue in the dichotic narrowband condition, it was shown from physiological models that different listeners might use different frequency channels to detect tones in noise.

Contributors and Funding Sources

This work was supervised by a dissertation committee consisting of Professor Laurel H. Carney (advisor) of the Department of Biomedical Engineering and Neurobiology & Anatomy, Professors Mark Bocko and Jack Mottley of the Department of Electrical and Computer Engineering, and Professor William O'Neill of the Department of Neurobiology & Anatomy at the University of Rochester, and Professor Azadeh Vosoughi of the Department of Electrical Engineering and Computer Science at the University of Central Florida. The work described in Chapter 2 was done in collaboration with Professor Azaden Vosoughi. All other work conducted for the dissertation was completed by the student independently, under the supervision of Prof. Laurel H. Carney. Graduate study was supported by the National Institute of Health and National Institute on Deafness and Other Communication Disorders under grant number R01-010813.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Detection of signals in noise is common in everyday life. Listeners need to focus on one auditory stimulus while ignoring all other distracting stimuli. The healthy auditory system can tune into a target, or signal, and filter out all other noise sources, a phenomenon often referred to as the "cocktail party effect". The underlying mechanism is related to the availability of binaural cues, listeners' attention, and their ability to detect signals in noise, although it is unknown exactly how these factors combine to affect listeners' detection. Listeners with hearing loss generally have difficulty communicating in noisy backgrounds even when using hearing aids. Identification of reliable cues for detection and recovery of these cues could improve signal processing techniques used in current hearing aids. As a first step toward understanding how listeners with normal hearing detect complex stimuli (*e.g.*, speech) in the presence of competing noise sources, we will start with the basic pure tone stimulus. The focus of this thesis is the identification of possible cues or combinations of cues that can predict listeners' tone-in-noise detection performance.

1.1 Background: Diotic and Dichotic Detection

In early studies of tone-in-noise detection (*e.g.*, Blodgett *et al.*, 1958, 1962; Dolan and Robinson, 1967), random noises were used in each trial and listeners' thresholds were obtained under different signal and noise conditions (*e.g.*, duration, level, tone center frequency, noise bandwidth, etc.). Two different binaural listening conditions have been commonly tested: diotic (N_OS_O) and dichotic ($N_OS\pi$). Identical noise stimuli were presented at the two ears in both conditions; tones were added in-phase in the N_OS_O condition and out-of-phase in the $N_OS\pi$ condition (Fig. 1.1). The key difference in these two conditions is the phase relation of the tones presented at the two ears.

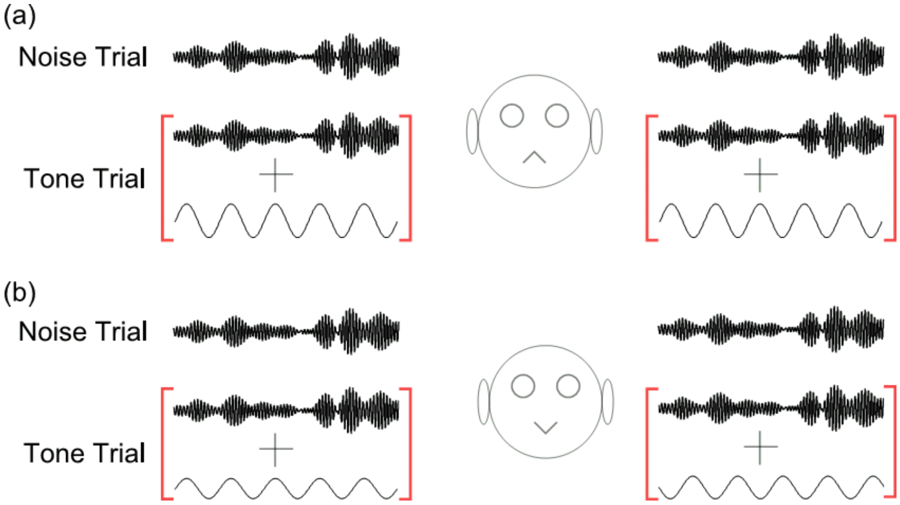

Figure 1.1: A schematic diagram illustrates two basic listening conditions: (a) diotic condition, N_OS_O and (b) dichotic condition, $N_OS\pi$. Note that the tone is inverted in the lower right panel (after Moore, 2003).

The difference between listeners' detection thresholds in the diotic and dichotic conditions is referred to as the binaural masking level difference (BMLD), which is largely due to the availability of interaural difference cues in the dichotic condition (Moore, 2003). A typical BMLD value for N_OS_O and $N_OS\pi$ conditions with tone frequency at 500 Hz in wideband noise is approximately 15 dB. The BLMD decreases as for higher tone frequencies (van der Par and Kolhrausch, 1999), with lower detection thresholds observed in the dichotic condition as compared to the diotic condition. The presence of the BMLD indicates that listeners' ability to detect signals in noise can be improved by the instantaneous phase and level differences introduced from the combination of in-phase noise and out-of-phase tones in the dichotic condition.

Listeners' thresholds can be obtained from studies using random noise maskers; however, a detailed description of the performance with respect to each noise stimulus cannot be achieved in that case. Several studies (Pfafflin and Mathews, 1966; Gilkey *et al.*, 1985; Siegel and Colburn, 1983; Isabelle and Colburn, 1991) tested listeners' detection with reproducible noises, a pre-generated set of random noises, and these noises were used repetitively during the testing process. In each trial, one noise was randomly picked from the set of noises. Listeners' detection responses varied for the same stimulus presented at different times, due to unknown factors that are referred to as internal noise. Because each reproducible noise was tested multiple times, the internal noise that affected listeners' performance was likely to be reduced by averaging, assuming that the internal noise was uncorrelated across trials. Listeners' detection performance was described in terms of hit rate (proportion of responses "tone present" for tone-plus-noise stimuli) and false-alarm rate (FA, proportion of responses "tone present" for noise-alone stimuli) for each tone-plus-noise and noise-alone sample. The set of hit and FA rates for all reproducible noises is referred to as a detection pattern (Davidson *et al.*, 2006). Gilkey et al. (1985) showed in their experiment that listeners' performance differed across noise waveforms.

Individual listeners' detection patterns have been shown to be highly consistent over the course of the experiment in several studies (Gilkey *et al.*, 1985; Isabelle and Colburn, 1991; Evilsizer *et al.*, 2001; Davidson *et al.*, 2006). In addition, detection patterns across different listeners are highly correlated in many listening conditions. Thus, it is likely that in these conditions, listeners use a similar strategy (or cues) for detecting tones in noise. Different models have been proposed to predict listeners' performance in several studies using specific cues. In each model, a decision variable (DV) is computed based on a certain feature of the set of stimuli, and DVs from the set of reproducible noises are compared to listeners' detection patterns.

1.2 Models for Diotic Detection

For the diotic condition, energy and temporal cues have been investigated in most modeling studies. The most common energy-based model is the critical-band (CB) model, and its DV is computed as the root-mean-square (rms) value of the energy in the critical band (Fletcher, 1940). The multiple-detector (MD) model is an extension of the CB model, and uses energies in several frequency bands (Gilkey *et al.*, 1986). Both energy models can predict significant amounts of the variance in listeners' detection performance. However, the CB model fails in roving-level conditions, in which stimulus levels are randomized in each trial of the experiment. Although the MD model is robust in roving-level conditions, the computation of the DV incorporates fitting the model parameters for each listener. In this thesis, the goal is to obtain a general model without fitting to the data.

For temporal models, envelope (slow fluctuations of the amplitude of the time waveform) and fine-structure (fast fluctuations of the time waveform) cues have been proposed to predict listeners' detection patterns. Envelope fluctuation is an indicator of tone presence because adding a tone to a narrowband noise results in a flatter envelope (Richards, 1992). The DV of the envelope cue is computed as the averaged absolute value of the envelope-slope (ES) from a fourth-order gamma-tone filtered stimulus (Richards, 1992; Zhang, 2004; Davidson *et al.*, 2006). Richards (1992) showed that the ES model is robust in the roving-level condition. For the fine-structure cue, the phase-opponency (PO) model generates a DV by computing the response from a coincidence-detector that receives inputs from two model auditory-nerve (AN) fibers. The two model AN fibers' center frequencies are located symmetrically about the tone frequency and have a 180-degree phase difference at the tone frequency (Carney *et al.*, 2002). The DV for the PO model decreases for a tone-plus-noise stimulus compared to a noise-alone stimulus because adding the tone results in out-of-phase responses of the two AN fibers, making the coincidence-detector less likely to fire a spike. Other temporal physiological models have also been used for predicting tone-in-noise detection, such as the Dau *et al.* (1996) and Breebaart *et al*. (2001) models which calculate the correlation between test stimulus and a stored template signal.

Predictions based on the diotic models described above cannot explain the predictable variance, which captures the variance in the detection patterns that is common across all listeners. The predictable variance is computed from the squared average correlation between detection patterns of individual listeners and detection patterns of the averaged listener, and is used as a benchmark for evaluating model predictions (Mao *et al.*, 2013).

1.3 Models for Dichotic Detection

For the dichotic condition, the combination of in-phase noises and out-of-phase tones introduces binaural differences at the two ears. Listeners have lower thresholds for detecting tones in noises in the dichotic condition than in the diotic condition. However, unlike the diotic condition, in which correlations among different listeners are highly consistent in both narrowband and wideband conditions, listeners' detection patterns are less correlated for the narrowband waveforms in the dichotic condition while correlations in the wideband condition are high.

Interaural difference cues have been commonly used to predict listeners' detection patterns. Two basic cues were interaural level difference (ILD) and interaural time difference (ITD). Several studies have tried different forms of ILD and ITD cues, such as the standard deviation or variance of the ILD and ITD (Isabelle, 1995), different forms of linear combinations of ILD and ITD (Isabelle and Colburn, 1987; Isabelle, 1995; Goupell and Hartmann, 2007), and peak ITD values (Isabelle, 1995). Energy models with DVs computed from equalization cancellation (EC) and normalized cross-correlation (NCC) models have also been tested. However, DVs from these two models are highly correlated (Isabelle, 1995), and do not predict a significant amount of the detection patterns. In general, none of the existing dichotic models' predictions can explain the predictable variance.

1.4 Overview of the thesis

Different model have been proposed to explain listeners' detection patterns, however, none of the diotic or dichotic model predictions can explain a substantial portion of the predictable variance. The goal of this thesis is to identify the cues that could explain a satisfactory amount of the detail in listeners' detection patterns. In this thesis, an optimal nonlinear combination of energy and temporal cues is proposed to explain listeners' diotic detection patterns. The nonlinear combination model was based on a likelihood-ratio test, a fundamental two-alternative detection method (Van Trees, 1968). It was shown that the nonlinear model predictions approached the predictable variance.

For the dichotic condition, two different models are proposed here. One model was based on a modified linear combination of ILD and ITD cues. The modification aimed to account for the correlation between ITD and ILD cues (Zurek, 1991; Isabelle, 1995), which has typically been ignored in previous linear models. The other model proposed here was the Slope of the Interaural Envelope Difference model (SIED), which is an extension of the ES model. Predictions obtained from these two models were significantly higher than those of previous models, especially for the SIED model in the wideband condition.

In addition, the hypothesis that both diotic and dichotic envelope cues can be represented by simple neural mechanisms along the auditory pathway was tested. A similar amount of listeners' performance can be predicted using envelope cues derived from physiological models as using the stimulus-based envelope cues. The physiological envelope cue was computed from responses of models for neurons in the auditory midbrain, or inferior colliculus (IC). The auditory midbrain is an ideal place to study because it is the first place along the ascending auditory pathway where tuning to modulation frequency is observed.

This thesis consists of three sections (Chapters 2, 3, and 4), and each is presented in the form of a journal paper. Each chapter has its own abstract, introduction, methods, results, and discussion sections. Chapter 2 describes a nonlinear optimal cue-combination model for the diotic condition. This model was motivated by two-alternative signal detection theory (Van Trees, 1968), and showed significant improvements of predictions for listeners' detection performance compared with earlier single-cue models. This chapter has been published in the Journal of the Acoustical Society of America (Mao *et al.*, 2013). Chapter 3 introduces two models for the dichotic condition. One is a modified model that combines ILD and ITD cues and accounts for the correlation between them, and the other is based on the binaural envelope cue. Both models predicted a significant amount of listeners' detection performance. This chapter has been submitted for publication. In Chapter 4, a physiological model using basic neural mechanisms of responses at the model IC outputs was able to predict similar amounts of the variance in listeners' performance as those from stimulus-based envelope cues. Finally, in Chapter 5, a summary and discussion of all three projects are presented. Ideas for future work are discussed at the end of the thesis.

Bibliography

Blodgett, H. C., Jeffress, L. A., and Taylor, R. W. (1958). “Relation of masked threshold to signal-duration for interaural phase combination,” Am. J. Psychol. 71, 283-290.

Blodgett, H. C., Jeffress, L. A., and Whitworth, R. H. (1962). “Effect of noise at one ear on the masked threshold for tone at the other,” J. Acoust. Soc. Am. 34, 979-981.

Breebaart, J., van der Par, S., and Kohlrausch, A. (2001a). “Binaural processing model based on contralateral inhibition I. Model structure,” J. Acoust. Soc. Am. 110, 1074–1088.

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. Acust. Acust. 88, 334–347.

Davidson, S. A., Gilkey, R. H., Colburn, H. S., and Carney, L. H. (2006). “Binaural detection with narrowband and wideband reproducible noise maskers. III. Monaural and diotic detection and model results,” J. Acoust. Soc. Am. 119, 2258-2275.

Dau, T., Püschel, D., and Kohlrausch, A. (1996). “A quantitative model of the “effective” signal processing in the auditory system. I. Model structure,” J. Acoust. Soc. Am. 99, 3615–3622.

Dolan, T. R., and Robinson, D. E. (1967). “Explanation of masking-level difference that result from interaural intensive disparities of noise,” J. Acoust. Soc. Am. 42, 977-981.

Evilsizer, M. E., Gilkey, R. H., Mason, C. R., Colburn, H. S., and Carney, L. H. (2002). “Binaural detection with narrowband and wideband reproducible maskers: I. Results for human,” J. Acoust. Soc. Am. 111, 333-345.

Fletcher, H. (1940). “Auditory patterns,” Rev. Mod. Phys. 12, 47–65.

Gilkey, R. H., Robinson, D. E., and Hanna, T. E. (1985). “Effects of masker waveform and signal-to-masker phase relation on diotic and dichotic masking by reproducible noise,” J. Acoust. Soc. Am. 78, 1207-1219.

Gilkey, R. H., and Robinson, D. E. (1986). “Models of auditory masking: A molecular psychophysical approach,” J. Acoust. Soc. Am. 79, 1499-1510.

Goupell, M. J., and Hartmann, W. M., (2007). “Interaural fluctuations and detection of interaural incoherence. III. Narrowband experiments and binaural models,” J. Acoust. Soc. Am. 122, 1029-1045.

Isabelle, S. K., (1995). “Binaural detection performance using reproducible stimuli,” Ph.D. thesis, Boston University, Boston, MA.

Isabelle, S. K., and Colburn, H. S., (1987). “Effects of target phase in narrowband frozen noise detection data,” J. Acoust. Soc. Am. 82, S109-S109.

Isabelle, S. K., and Colburn, H. S., (1991). “Detection of tones in reproducible narrow-band noise,” J. Acoust. Soc. Am. 89, 352-359.

Mao, J., Vosoughi, A., and Carney, L. H., (2013). “Predictions of diotic tone-in-noise detection based on a nonlinear optimal combination of energy, envelope, and fine-structure cues,” J. Acoust. Soc. Am. 134, 396-406.

Moore, B. C. J., (2003). *An introduction to the psychology of hearing* (Elsevier Science & Technology Books).

Pfafflin, S. M., and Mathews, M. V., (1966). “Detection of auditory signals in reproducible noise,” J. Acoust. Soc. Am. 39, 340-345.

Richards, V. M. (1992). “The delectability of a tone added to narrow bans of equal energy noise,” J. Acoust. Soc. Am. 91, 3424-3435.

Siegel, R. A., and Colburn, H. S., (1983). “Internal and external noise in binaural detection,” Hearing Reaserch, Vol. 11, 117-123.

van der Par. S, and Kolhrausch. A., (1999). “Dependence of binaural masking level differences on center frequency, masker bandwidth, and interaural parameters,” J. Acoust. Soc. Am. 106, 1940-1947.

Van Trees, H. L. (1968). *Detection, estimation, and modulation theory. Part I. Detection, estimation and linear modulation theory* (John Wiley & Sons, New York), Chap. 2, pp. 26-36.

Viemeister, N. F. (1979). “Temporal modulation transfer function based upon modulation thresholds,” J. Acoust. Soc. Am. 66, 1364-1380.

Zhang, X., (2004). “Cross-frequency coincidence detection in the processing of complex sounds,” Ph.D. thesis, Boston University, Boston, MA.

Zurek, P. M., (1991). “Probability distributions of interaural phase and level differences in binaural detection stimuli,” J. Acoust. Soc. Am. 90, 1927-1932.

Chapter 2

Predictions of Diotic Tone-in-Noise Detection Based on a Nonlinear Optimal Combination of Energy, Envelope, and Fine-Structure Cues

2.1 Abstract

Tone-in-noise detection has been studied for decades; however, it is not completely understood what cue or cues are used by listeners for this task. Model predictions based on energy in the critical band are generally more successful than those based on temporal cues, except when the energy cue is not available. Nevertheless, neither energy nor temporal cues can explain the predictable variance for all listeners. In this study, it was hypothesized that better predictions of listeners' detection performance could be obtained using a nonlinear combination of energy and temporal cues, even when the energy cue was not available. The combination of different cues was achieved using the logarithmic likelihood-ratio test (LRT), an optimal detector in signal detection theory. A nonlinear LRT-based combination of cues was proposed, given that the cues have Gaussian distributions and the covariance matrices of cue values from noise-alone and tone-plus-noise conditions are different. Predictions of listeners' detection performance for three different sets of reproducible noises were computed with the proposed model. Results

showed that predictions for hit rates approached the predictable variance for all three datasets, even when an energy cue was not available.

2.2 Introduction

Detecting signals in noise is important for everyday activities, such as detecting speech in background noise and discriminating sounds in noisy environments. People with hearing loss have difficulty communicating in background noise even when using hearing aids. Thus, it is essential to understand how people with normal hearing can detect signals in noise in order to help design more effective hearing-aid devices. Tone-in-noise detection has been studied for decades as a stepping stone to find the cues that listeners use to detect more complex sounds in noise.

In early tone-in-noise detection studies, noise waveforms were generated randomly for each trial such that no waveform was tested twice (Blodgett *et al.*, 1958, 1962; Dolan and Robinson, 1967). Detection performance was averaged across listeners and waveforms. However, Gilkey *et al.* (1985) found that detection performance varied among listeners and waveforms by inspecting the detection performance for a set of pre-generated waveforms. Because these waveforms were stored and could be "reproduced" exactly, they were referred to as reproducible noises. Using reproducible noise waveforms it is possible to compare each listener's detection performance for individual waveforms and to make detailed tests of different model predictions.

In detection tests, listeners' performance is described by the proportion of correct identification of tone presence for tone-plus-noise waveforms (hit rate), and the proportion of “tone present” responses for noise-alone waveforms (false-alarm, FA rate). The set of hit and FA rates for a given ensemble of reproducible noise maskers has been referred to as a detection pattern (Davidson *et al.*, 2006).

In order to identify the cues used by listeners to detect a tone in noise in the diotic condition, several single-cue models based on energy or temporal cues have been used to predict listeners’ detection patterns. In each model, a set of decision variables (DVs) that represent a particular feature of the corresponding reproducible waveforms is compared with the listeners’ detection patterns. A description of several models in the literature is presented below. In particular, several commonly used energy and temporal cues and their performance in predicting listeners’ detection patterns are described.

The critical-band model (CB, Fletcher, 1940) focuses on energy within a critical bandwidth of the tone frequency, whereas the multiple-detector model (MD, Gilkey *et al.*, 1986) considers energy within and outside a critical bandwidth. Although these energy-based models provide satisfactory predictions of the detection patterns, the CB model fails at predicting the roving-level stimulus condition, in which the level of stimulus is randomly varied for each trial (Kidd *et al.*, 1989). Because the CB model predictions are based on the absolute energy within one filter bandwidth and stimulus levels are not fixed in each trial, “tone presence” would be predicted for a high-level noise-alone stimulus. The MD model is robust for roving-level noises and yields significantly better predictions than the CB model for most listeners in the wideband condition; however, the MD model computations involve fitting to the data (Davidson *et* *al.*, 2009). Fitting the data was avoided in this study in order to achieve a generic model for different types of stimuli and to prevent the risk of over-fitting the data, i.e. adjusting the parameters of variables for individual listeners to better match each detection pattern. In addition, the MD model is not applicable for waveforms whose bandwidths are smaller than one critical bandwidth, because this model requires comparison of energy in different frequency bands. Thus, the CB model was used to describe the energy cue in this study.

Two types of temporal cues are robust to the roving-level condition: envelope and fine-structure. The envelope-slope model (ES, Richards, 1992; Zhang, 2004; Davidson *et al.*, 2006) examines the changes in envelope fluctuations. Adding a tone to a narrowband noise results in a decrease in envelope fluctuations, thus lower values of the DV for the ES model indicate a tone-plus-noise waveform. This model can be applied to wideband noises because the output of narrowband cochlear filters is analyzed in the model computation.

The phase-opponency model (PO, Carney *et al.*, 2002), based on fine-structure, i.e. the fast fluctuations in the stimulus, uses responses from a coincidence detector that receives inputs from two model auditory-nerve fibers to predict tone presence. Because the two auditory-nerve fibers are tuned to frequencies symmetrically located around the tone frequency and have phase responses that differ by 180 degrees at the tone frequency, the addition of a tone to a noise waveform yields fewer spike responses from the coincidence detector. Therefore, a lower value of the DV for the PO model indicates a tone-plus-noise waveform. In addition to the ES and PO models, the Dau *et al.* (1996a) and Breebaart *et al.* (2001) template-matching models also use temporal cues. In these models, detection results are based on comparing the internal test waveform representation with the pre-stored waveform representation in the template. However, previous studies have shown that these template-matching models do not yield predictions that were significantly correlated to the detection patterns for the ensemble of reproducible waveforms used in this study (Davidson *et al.*, 2009a). Thus, the ES and PO models were used to evaluate the temporal features of the stimulus waveforms in this study.

Although previous studies have reported that correlations between predictions of some diotic models and listeners' detection patterns are statistically significant, the amounts of variance in the detection patterns that are explained by these models are substantially lower than the predictable variance (Davidson *et al.*, 2009a). The predictable variance is computed as the squared mean of the correlations between detection patterns of individuals and those of the average listener (the mean of the detection patterns from individual listeners). Detection patterns differ for each listener; the predictable variance describes the proportion of the variation in detection patterns that is common among all listeners. Thus, the predictable variance is used as a benchmark for model predictions.

The goal of this study was to test the hypothesis that significantly better predictions for diotic detection could be obtained by using models that *combine* different cues, *i.e.*, multiple-cue models. Given that different cues represent different features of a waveform, it is reasonable to argue that the combination of different cues can capture more information about a waveform than any single cue. Davidson *et al.* (2009b) reported that a multiple-cue model, based on a linear combination of envelope and fine-structure cues, results in poor predictions of listeners' detection patterns. However, energy and temporal cues are correlated, and a simple linear combination of cues is ineffective in characterizing the interaction among cues (Davidson *et al.*, 2009a).

In this study, a *nonlinear* multiple-cue model was proposed to predict listeners' detection patterns, where the model takes into account the statistical correlations among energy and temporal cues in cue combination. The likelihood ratio test (LRT) is an optimal detector for a two-alternative (binary) hypothesis testing (Van Trees, 1968) and is thus a useful tool for tone-in-noise detection data. The LRT-based detection model has previously been used by Siebert (1970), Colburn (1973), and Heinz *et al.* (2001) to predict frequency, interaural time, and level discrimination data, respectively, based on model auditory-nerve responses. In this study, the DV of the nonlinear multiple-cue model was computed as the logarithmic likelihood ratio of cue values given tone-plus-noise and noise-alone waveforms. Distributions of the values of single cues were computed from a set of randomly generated noise-alone and tone-plus-noise waveforms that was different from the reproducible waveforms used for the detection task. Because of the difference between the covariance matrices of cue values for noise-alone and tone-plus-noise waveforms, the expression for the DV is a quadratic function in terms of cue values, implying a nonlinear combination of cues. In addition, the DV also includes cross-products of single cues that characterize the pair-wise interactions between cues.

In summary, a nonlinear cue-combination model which optimally combines energy, envelope, and fine-structure cues is presented in this study. It was shown that model predictions based on the nonlinear multiple-cue model improved significantly compared with those based on single-cue or linear multiple-cue models.

2.3 Description of Data

The diotic detection data was obtained from three previous experiments (Evilsizer *et al.*, 2002; Davidson *et al.*, 2006; Davidson *et al.*, 2009b). Tone frequency was 500 Hz in all three datasets, and listeners were tested at tone levels near their detection threshold (*i.e.*, an overall *d'*=1). In the first two studies, the same set of twenty-five reproducible noise waveforms was used, and eight listeners were tested. The duration of the noise waveforms was 300 ms, and the sound level was 40 dB SPL. Both narrowband (452-552 Hz) and wideband (100-3000 Hz) noises were tested. The spectral content of the narrowband waveform was matched to the corresponding frequency range of the wideband waveform. In the third study, fifty equal-energy reproducible noise waveforms with 100-ms duration, 40 dB SPL, and narrower bandwidth (475-525 Hz) were used (baseline and control stimulus sets as described by Davidson *et al.*, 2009b). Six listeners were tested in that study. In the present study, this dataset based on equal-energy stimuli was useful to test whether model predictions depended more on temporal cues in the absence of the energy cue.

In all studies, listeners responded whether they perceived a tone after each single-interval trial of a noise-alone or tone-plus-noise waveform. Detection patterns were described in terms of hit and FA rates, based on listeners' responses of "tone presence" (details of the experiments can be found in Evilsizer *et al.*, 2002; Davidson *et al.*, 2006; and Davidson *et al.*, 2009b).

Figure 2.1 shows the detection pattern of the average listener (*i.e.*, the average detection pattern across all individual listeners) for the 100-Hz bandwidth waveforms in the Evilsizer *et al.* (2002) and Davidson *et al.* (2006) studies. The detection patterns were consistent over the course of the experiment and were also significantly correlated across listeners. The goal of this study was to predict the variation in the average listener's detection pattern across the set of reproducible noises. Because the detection patterns were significantly correlated among individual listeners, these listeners were assumed to be using similar cues for tone-in-noise detection. Model predictions of the response of the average listener focus on explaining the common variance across listeners' performance while ignoring individual differences, which cannot be accounted for by a single model. The quality of the prediction was described as the proportion of variance in the detection pattern that is explained by a given model.

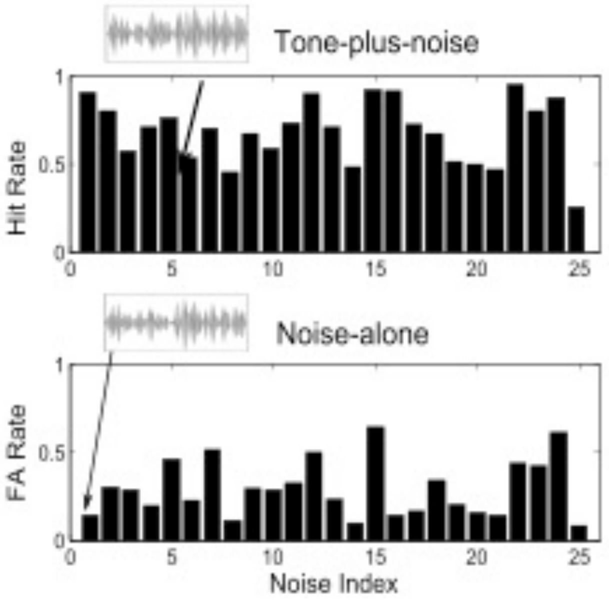

Figure 2.1: The detection pattern of the average listener comprises hit and FA rates for each 100-Hz bandwidth reproducible waveform averaged across eight individual listeners. The x-axis shows the index of the reproducible noise waveforms. The insets show examples of tone-plus-noise (top) and noise-alone (bottom) waveforms (data from Evilsizer *et al.*, 2002; and Davidson *et al.*, 2006).

2.4 Methods

It was hypothesized that better predictions of reproducible-noise detection patterns could be achieved using nonlinear multiple-cue models that consider statistical correlations among different cues. First, the energy, envelope, and fine-structure cues used in the cue combination step will be introduced. Next, the statistical correlations between energy and temporal cues are examined for the three datasets. Last, both the nonlinear LRT-based multiple-cue and the linear multiple-cue models will be described.

2.4.1 Energy and Temporal Cue Models

The CB (Fletcher, 1940) model, which is based on energy within a critical bandwidth of the target frequency, was used in the current study. The DV was computed as the root mean square (RMS) of a fourth-order gamma-tone filtered waveform (centered at 500 Hz) for all three datasets: $CB = \left\{ \frac{\int_T G\left[x(t)\right]^2 dt}{T} \right\}^{\frac{1}{2}}$,where *x(t)* indicates the stimulus waveform, and *G(.)* represents the response of the gammatone filter.

Two temporal models were used: the ES (Richards, 1992; Zhang, 2004; Davidson *et al.*, 2006) and PO (Carney *et al.*, 2002) models. DVs of the ES model were based on changes in envelope fluctuations. The envelope was computed from the Hilbert transform of a fourth-order gamma-tone filtered stimulus (centered at 500 Hz). The DV value is reduced by addition of the tone for the ES model because envelope fluctuation decreases. Figure 2.2 illustrates the averaged distribution of envelope energy for noise-alone (solid lines) and tone-plus-noise (dotted lines) stimuli in the frequency domain. The insets show enlarged views of the circled frequency region that yielded the largest differences in the envelope magnitude between noise-alone and tone-plus-noise stimuli. The ES model was modified in the current study to emphasize this frequency range by substituting the low-pass envelope filter (cutoff frequency at 250 Hz) with a sixth-order bandpass envelope filter centered at 120 Hz (Q=1). The computation of the modified ES cue is:

$$ES = \frac{\int_T \left| H\left[G\left(x(t)\right)\right] - H\left[G\left(x(t+\Delta t)\right)\right] \right| dt}{\left\{ \frac{\int_T H\left[G\left(x(t)\right)\right]^2 dt}{T} \right\}^{\frac{1}{2}}}$$

,where *x(t)* indicates the stimulus waveform, *G(.)* represents the response of the gammatone filter, and *H(.)* is the envelope extracted using the Hilbert transform. The bandpass envelope filter, which is similar to physiological and psychological modulation filters, was applied to extract frequency components in the range illustrated. In addition, this filter attenuated low frequencies, which contain more energy but less information about the presence of the tone. The modified ES model, compared with the original ES model, could predict 20% and 10% more of the variance in hit and FA rates, respectively, for the average listener's narrowband detection patterns; whereas predictions from the modified ES model explained 10% less of the variance for the wideband hit rates than the original ES model, with no change in the FA rates (Davidson *et al.*, 2009a).

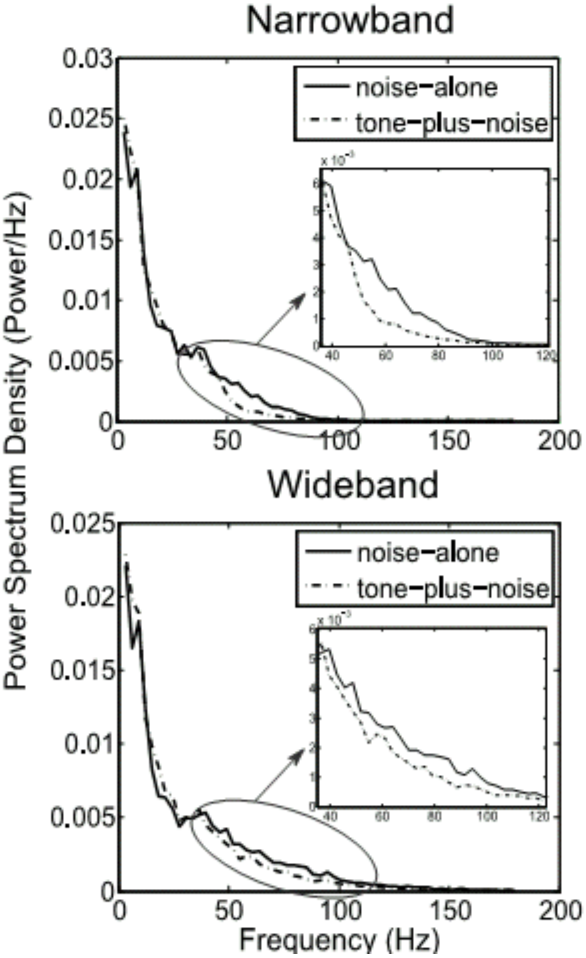

Figure 2.2: Envelope power spectrum density of noise-alone (solid lines) and tone-plus-noise (dotted lines) stimuli in narrowband (top) and wideband (bottom) conditions. Insets show an enlarged view of the circled frequency range where the largest difference of the envelope spectral energy between these two stimuli was observed.

The PO model extracts fine-structure information from the stimuli using a coincidence detector that receives inputs from two model auditory-nerve fiber responses: $PO = \int_T A_{N1}\left[x(t)\right] \cdot A_{N2}\left[x(t)\right] dt$, where *x(t)* indicates the stimulus waveform, and A_{N1} and

A_{N2} denote auditory-nerve models with two different characteristic frequencies. Because tone responses from the two model auditory-nerve fibers differed in phase by 180 degrees, low DV values for the PO model indicate tone-plus-noise waveforms.

Figure 2.3 shows the three models that extract the single cues used in this study: the energy cue (the CB model), envelope cue (the ES model), and fine-structure cue (the PO model).

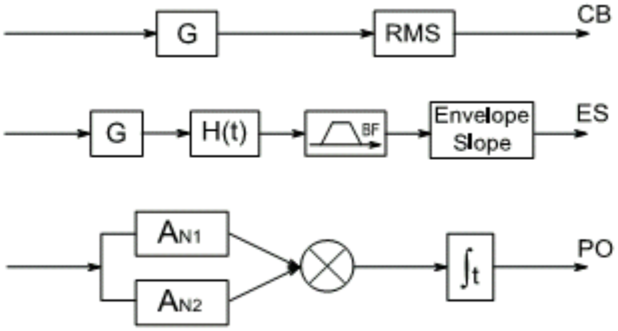

Figure 2.3: A schematic diagram of the CB, ES, and PO models used to extract energy and temporal cues. In the CB model, DV was computed as the root mean square (RMS) of a fourth-order gamma-tone filtered waveform (center frequency 500Hz, bandwidth equaled one critical bandwidth of tone frequency). In the ES model, the envelope of a waveform was computed using a Hilbert transform of a gamma-tone filtered waveform, and the DV was calculated as the slope of a band-pass filtered envelope. In the PO model, responses from two model auditory-nerve fibers that differed in phase by 180 degrees in response to the tone were applied to a coincidence detector, and the DV was computed as the integral of the coincidence detector responses.

2.4.2 Statistical Correlations between Energy and Temporal Cues

In order to investigate the relationship among different cues, the dependencies between pairs of cues were analyzed by computing the Pearson product-moment correlation coefficients between the DVs (Neter *et al.*, 1996). Table 2.1 shows the correlations of DVs for tone-plus-noise and noise-alone reproducible waveforms for the three conditions; bold values indicate DV pairs that are significantly correlated (p<0.05, t-test). For the computations in Table 2.1, the tone level was matched to the average listener's threshold. The two temporal DVs (ES and PO) were correlated in each dataset; the energy (CB) and temporal DVs were also correlated, except for the fine-structure cue in some conditions (Table 2.1). Furthermore, both energy and temporal DVs had distributions that were approximately Gaussian. In Fig. 2.4, the distributions of each DV are shown for large sets (n=500) of randomly generated 100-Hz bandwidth noise-alone and tone-plus-noise waveforms, and the dotted lines show the corresponding Gaussian fits. The correlation between the DV distribution and the fitted Gaussian curve is shown at the top of each panel. The distribution of hits for the ES cue is slightly asymmetric; however, the correlation between the distribution and its Gaussian fit is high (r=0.93). Distributions of cue values for randomly generated 2900-Hz and 50-Hz equal-energy waveforms were also approximately Gaussian (not shown). In addition, further analysis was done to investigate whether the statistical distributions of cue values were Poisson-like. Results showed that the mean values were significantly different from the variance of the distributions for each cue, thus the cues did not have Poisson distributions.

Table 2.1: Correlations between energy and temporal DVs for three datasets. The bold values indicate that two DVs are significantly correlated ($p<0.05$, $r>0.40$ for $n=25$ and $r>0.28$ for $n=50$), and n denotes the number of waveforms in each study.

Name of Cues	**2900-Hz waveforms (n=25)**				**100-Hz waveforms (n=25)**				**50-Hz waveforms (n=50)**			
	Envelope (ES)		Fine-structure (PO)		Envelope (ES)		Fine-structure (PO)		Envelope (ES)		Fine-structure (PO)	
	Hit	FA	Hit	FA	Hit	FA	Hit	FA	Hit	FA	Hit	FA
Energy (CB)	**0.69**	**0.60**	0.36	**0.58**	**0.55**	**0.48**	0.15	0.35	**0.55**	**0.52**	**0.51**	0.19
Envelope (ES)	__	__	**0.48**	**0.74**	__	__	**0.48**	**0.79**	__	__	**0.75**	**0.65**

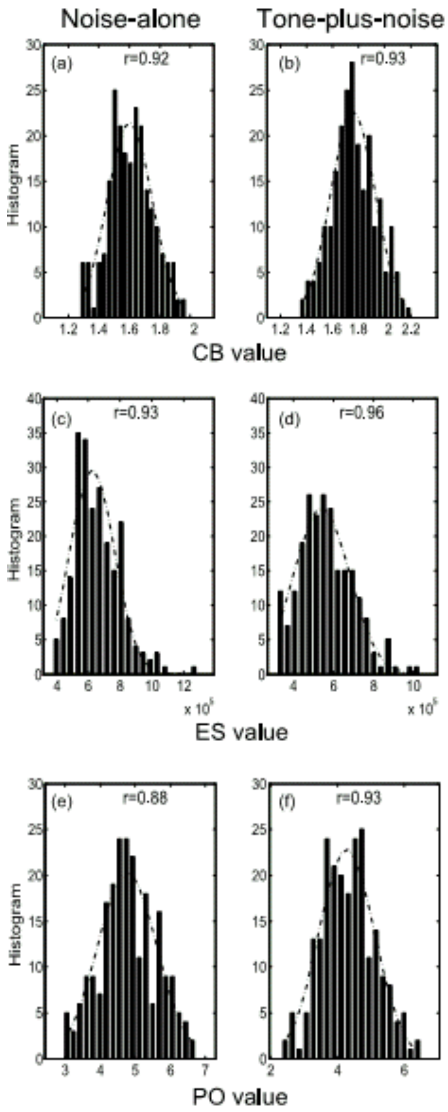

Figure 2.4: DV distributions for 200 randomly generated narrowband noise-alone (left column) and tone-plus-noise (right column) waveforms. The x-axis shows the cue values and the y-axis shows the number of instances in each bin in the histogram (20 bins in total). The label on the x-axis shows the model names. Panels in each row show the distributions of the DVs for the CB (panel a and b), ES (panel c and d), and PO (panel e and f) cues. In each panel, the dotted line represents a Gaussian fit to the DV distribution, and the *r* value at the top indicates the correlation between the DV distribution and the Gaussian fit.

2.4.3 Decision Variable of the Nonlinear LRT-based Multiple-Cue Model

The DV of the test waveform was calculated from the logarithmic LRT of its cue values assuming the test waveform belonged to noise-alone ($x = N$) and tone-plus-noise ($x = S$) categories. Equation 1 shows the nonlinear combination of energy and temporal cues, in which $\mathbf{c} = [c_1, c_2, c_3]^T$ denotes the vector of cue values for the test waveform, c_1 denotes the energy cue (CB), c_2 denotes the envelope cue (ES), and c_3 denotes the fine-structure cue (PO), and n represents the number of cues (n=3 in this study).

$$D(\mathbf{c}) = \log\left(\frac{P(\mathbf{c} \mid S)}{P(\mathbf{c} \mid N)}\right) \quad \textit{and}$$

$$P(\mathbf{c} \mid x) = \frac{1}{\sqrt{(2\pi)^n \det\left(\boldsymbol{\Sigma}_{x,r}\right)}} \exp\left(-\frac{1}{2}\left(\mathbf{c} - \boldsymbol{\mu}_{x,r}\right)^T \boldsymbol{\Sigma}_{x,r}{}^{-1}\left(\mathbf{c} - \boldsymbol{\mu}_{x,r}\right)\right),$$

$$\text{where } x \in \{S, N\}, \text{ and } \boldsymbol{\mu}_{x,r} = E\left[\mathbf{c}_{x,r}\right], \text{ and } \boldsymbol{\Sigma}_{x,r} = E\left[\left(\mathbf{c} - \boldsymbol{\mu}_{x,r}\right)\left(\mathbf{c} - \boldsymbol{\mu}_{x,r}\right)^T\right]. \tag{1}$$

$P(\mathbf{c} \mid x)$ represents the conditional probability of cue values ($\mathbf{c}$) given that the testing waveform belongs to category x *(x=N* or *x=S)*. Because the single-cue DVs were correlated and their values had Gaussian distributions (Fig. 2.4), the conditional probability was computed using a multivariate Gaussian distribution. The term of $\boldsymbol{\mu}_{x,r}$ denotes the expected value of the cue vector ($\mathbf{c}_{x,r}$) for category x computed from the randomly generated waveforms, where r indicates the randomly generated waveforms. The covariance matrix $\boldsymbol{\Sigma}_{x,r}$ characterizes the statistical correlations among different cues; $\boldsymbol{\Sigma}_{S,r}$ and $\boldsymbol{\Sigma}_{N,r}$ are different because the correlations among different cues vary for noise-alone and tone-plus-noise waveforms. Given that $P(\mathbf{c} \mid S)$ and $P(\mathbf{c} \mid N)$ have multivariate Gaussian distributions, the logarithmic LRT in Equation 1 can be described as

$$D(\mathbf{c}) = \frac{1}{2}\log\left(\frac{\det\left(\boldsymbol{\Sigma}_{N,r}\right)}{\det\left(\boldsymbol{\Sigma}_{S,r}\right)}\right) - \frac{1}{2}\left(\mathbf{c} - \boldsymbol{\mu}_{S,r}\right)^{T} \boldsymbol{\Sigma}_{S,r}{}^{-1}\left(\mathbf{c} - \boldsymbol{\mu}_{S,r}\right) + \frac{1}{2}\left(\mathbf{c} - \boldsymbol{\mu}_{N,r}\right)^{T} \boldsymbol{\Sigma}_{N,r}{}^{-1}\left(\mathbf{c} - \boldsymbol{\mu}_{N,r}\right). \qquad (2)$$

On the right-hand side of Equation 2 a quadratic function in terms of the cue values was obtained because $\boldsymbol{\Sigma}_{S,r}$ and $\boldsymbol{\Sigma}_{N,r}$ are different. Thus, the current model is a nonlinear combination of different cues.

The logarithmic likelihood-ratio test is an optimal detector for a two-alternative detection problem (Van Trees, 1968). This test can be interpreted as testing whether the waveform is more likely to contain a tone or not. Specifically, because the prior probabilities of given noise-alone or tone-plus-noise waveforms are equal [$P(N) = P(S)$], a DV with a value greater than zero suggests that the current waveform is a tone-plus-noise stimulus; a DV with a value less than zero suggests that the current waveform is a noise-alone stimulus. The nonlinearity of the LRT model is guaranteed as long as the covariance matrices from noise-alone and tone-plus-noise waveforms are different. Assuming that the two covariance matrices were the same, then the first term in Equation 2 would be zero and the second-order term of cue values would cancel out; thus, this equation would become a linear combination of cue values, as

$$D(\mathbf{c}) = \left(\boldsymbol{\mu}_{S,r}{}^{T} - \boldsymbol{\mu}_{N,r}{}^{T}\right)\boldsymbol{\Sigma}^{-1}\mathbf{c} + \frac{1}{2}\boldsymbol{\mu}_{N,r}{}^{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_{N,r} - \frac{1}{2}\boldsymbol{\mu}_{S,r}{}^{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_{S,r}, \qquad (3)$$

where $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{S,r} = \boldsymbol{\Sigma}_{N,r}$. Furthermore, pair-wise interactions between single cues are guaranteed as long as the cues are correlated. Another case to consider is the assumption that the covariance matrices from noise-alone and tone-plus-noise waveforms are different but single cues are uncorrelated (i.e. the covariance matrices are diagonal). In that case, Equation 2 would reduce to

$$D(\mathbf{c}) = \frac{1}{2}\log\left(\frac{\det(\boldsymbol{\Sigma}_{N,r})}{\det(\boldsymbol{\Sigma}_{S,r})}\right) - \frac{1}{2}\sum_i \frac{\left(c_i - (\boldsymbol{\mu}_{S,r})_i\right)^2}{(\boldsymbol{\Sigma}_{S,r})_{ii}} + \frac{1}{2}\sum_i \frac{\left(c_i - (\boldsymbol{\mu}_{N,r})_i\right)^2}{(\boldsymbol{\Sigma}_{N,r})_{ii}}, \qquad (4)$$

where c_i is the *i-th* cue, $(\boldsymbol{\Sigma}_{S,r})_{ii}$ and $(\boldsymbol{\Sigma}_{N,r})_{ii}$ are the *(i,i)-th* entries of the covariance matrix of the tone-plus-noise and noise-alone waveforms. The DV described by Equation 4 is still nonlinear, but fails to capture the interactions between cues. Equations 3 and 4 serve to illustrate features of the full LRT model, which includes both a nonlinear combination of cues and the interactions between pairs of single cues. Figure 2.5 shows a schematic diagram of the computation of the DV for the nonlinear LRT-based multiple-cue model.

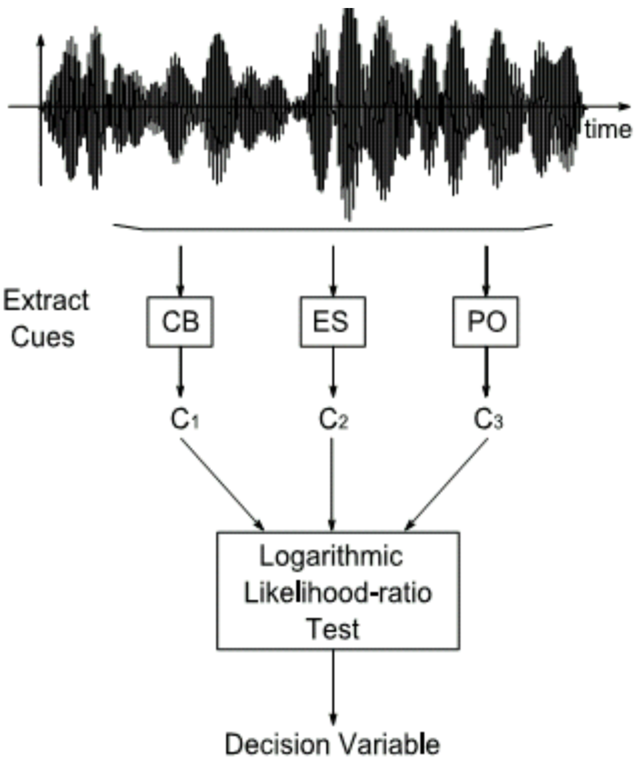

Figure 2.5: This schematic diagram illustrates the strategy for computing the nonlinear combination of cues. The DV is computed by combining energy and temporal cues using the nonlinear LRT-based multiple-cue model. Single cues are computed from the waveform (as in Fig. 2.3), and combined with a logarithmic likelihood-ratio test (shown in Equation 1, where c_1, c_2, and c_3 denote the cue values).

2.4.4 Decision Variable of the Linear Multiple-Cue Model

The DVs for a linear multiple-cue model were also computed using a weighted sum of energy and temporal cues. Performance of the linear and nonlinear cue-combination models was compared. Equation 5 illustrates the linear combination (LC) of energy and temporal cues, in which c_1 denotes the energy cue (CB), c_2 denotes the envelope cue

(ES), and c_3 denotes the fine-structure cue (PO) for the test waveform. The weights corresponding to each cue are designated as $w_{1,x,r}$, $w_{2,x,r}$, and $w_{3,x,r}$; x denotes the waveform category, and any term with the subscript r is computed from a large set of randomly generated waveforms.

$$\begin{aligned} &DV = D_S - D_N, \\ &D_x = w_{1,x,r}c_1 + w_{2,x,r}c_2 + w_{3,x,r}c_3, \\ &\text{where } x \in \{S, N\},\ w_{i,x,r} = \left[\left(\mathbf{\Sigma}_{x,r}\right)_{ii}\right]^{-1},\ \text{ and } i = 1,2,3. \end{aligned} \tag{5}$$

For each cue, the weight equals the inverse of the variance of the cue values, which corresponds to the inverse of the *(i,i)-th* entry in the covariance matrix $\mathbf{\Sigma}_{x,r}$. Assuming that listeners used a combination of energy and temporal cues in the detection task, this linear combination would yield an optimal estimation of the combined cue value if the energy and temporal cues were uncorrelated (Yuille and Bulthoff, 1996); however, energy and temporal cues are typically correlated (Davidson *et al.*, 2009a).

Given that the test waveform category was unknown during the detection task, the DV was computed as the difference between the combined cues for tone-plus-noise and noise-alone conditions. A DV with a value greater than zero suggests that the current waveform is a tone-plus-noise stimulus; a DV with a value less than zero suggests that the current waveform is a noise-alone stimulus.

2.5 Results

It was hypothesized that if a listener used a particular cue-combination rule to detect a tone in noise, then DVs computed from that particular rule would be strongly correlated to the listener's detection pattern. In this section, predictions from single-cue and multiple-cue models were evaluated by computing the squared Pearson product-moment correlation coefficient between DVs and the z-score of listeners' detection patterns. In the following figures, each bar shows the proportion of predicted variance (squared correlation between detection patterns and hit or FA rates) for the average listener. The length of the error bar shows the standard deviation of the predicted proportion of variance across individual listeners.

Figure 2.6a shows predictions based on the energy (CB) and temporal (ES and PO) single-cue models, as well as the linear (LC) and nonlinear (LRT) multiple-cue models for the 2900-Hz bandwidth waveforms. Predictions from the CB model alone were the best among the three single-cue models for both hit and FA rates. For multiple-cue models, predictions based on the LC model were similar to those of the CB model. However, predictions based on the LRT model approached the predictable variance (squared mean of the correlations between detection patterns of individuals and those of the average listener) for both hit and FA rates.

Model predictions based on the energy and temporal single-cue models, as well as the linear (LC) and nonlinear (LRT) multiple-cue models for the 100-Hz bandwidth waveforms are shown in Fig. 2.6b. Similar to the results for the 2900-Hz bandwidth waveforms, predictions based on the CB model alone were the best among the three single-cue models for both hit and FA rates, and predictions based on the LC model were similar to those of the CB model. Furthermore, predictions based on the LRT model met the predictable variance for both hit and FA rates.

For the 50-Hz bandwidth equal-energy waveforms, Fig.2.6c shows model predictions based on the energy and temporal single-cue models, as well as the linear (LC) and nonlinear (LRT) multiple-cue models. In contrast to the previous two datasets, the energies of the noise-alone and tone-plus-noise waveforms in this dataset were equalized, in an effort to remove the energy cue. Model predictions of hit and FA rates based on the ES model were the best among the three single-cue models. Similar to the other two datasets, predictions based on the LC model were close to those of the CB model.

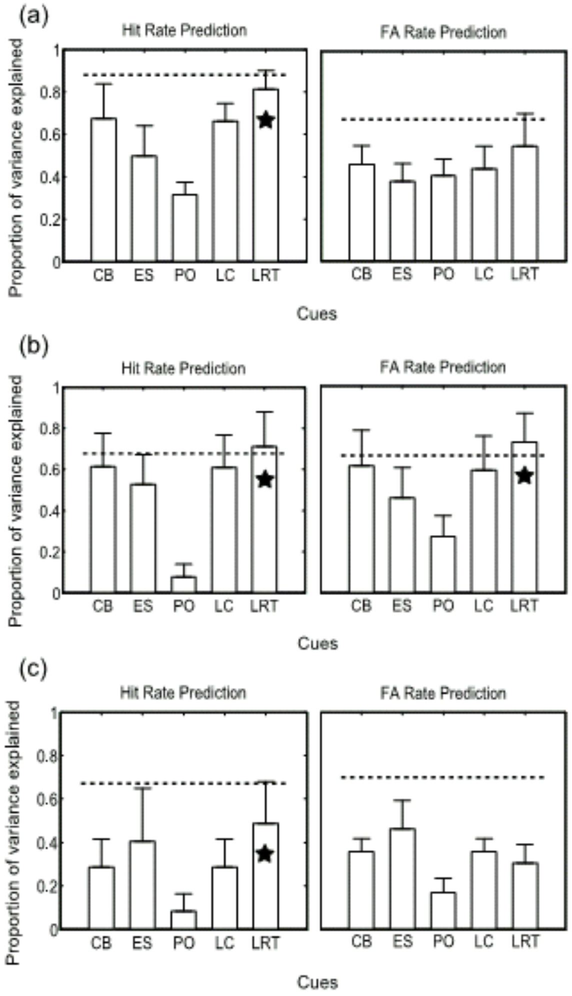

Figure 2.6: The proportion of variance explained by single-cue and multiple-cue models of the average listener for the (a) 2900-Hz bandwidth, (b) 100-Hz bandwidth, and (c) 50-Hz bandwidth waveforms. The x-axis shows the names of different models (CB: energy cue, ES: envelope cue, PO: fine-structure cue, LC: linear combination of three cues, LRT: nonlinear logarithmic likelihood ratio test combination of three cues). The stars indicate that multiple-cue model predictions were significantly improved compared with predictions from any single-cue model (p<0.05, n=25 for 2900-Hz and 100-Hz waveforms, n=50 for 50-Hz equal-energy waveforms). The y-axis shows the proportion of variance explained by different models. The length of the error bar shows the standard deviation of the predicted proportion of variance across individual listeners. The dotted lines indicate the predictable variance for hit and FA rates.

Model predictions for waveforms from the three datasets suggested that for tone-in-noise detection listeners may use a nonlinear combination of energy and temporal cues that takes into account the statistical correlations of the three cues. In order to test whether predictions from the LRT or LC model were significantly better than those of single-cue models, an incremental F-test was carried out to analyze the model predictions. In Fig. 2.6, bars with stars indicate that the nonlinear (LRT) model significantly improved predictions (p<0.05, n=25 for 2900-Hz and 100-Hz waveforms, n=50 for 50-Hz equal-energy waveform). For example, for the 2900-Hz bandwidth waveforms, the single-cue CB, ES and PO models were able to predict 68%, 50%, and 32% of the variance of hit rates, respectively. By combining all three cues with the nonlinear (LRT) model, 81% of the variance in the detection patterns could be predicted, and this amount of predicted variance was significantly greater than that from any of the single-cue models. For the LRT model, the amounts of predicted variance of hit rates for all noise bandwidths were significantly greater than those based on any of the single-cue models. The error bars indicate the standard deviation of model predictions across individual listeners. Although the difference between LRT and ES cue is not as large as in Fig. 2.6a and Fig. 2.6b, 50 waveforms were used in Fig. 2.6c while 25 waveforms were used in Fig. 2.6a and Fig. 2.6b. Thus, the improvement of LRT over ES is statistically significant (p=0.03). In addition, the amount of predicted variance of FA rates for the 100-Hz bandwidth waveform was also significantly greater than those based on any of the single-cue models, whereas amounts of predicted variance of FA rates for the 2900-Hz and the 50-Hz bandwidth equal-energy waveforms were not significantly greater than those based on the best single-cue model. In contrast, the amount of predicted variance of the LC model was not significantly greater than those of single-cue models; LC predictions were similar in quality to the CB predictions across all datasets and for both hits and FAs (Fig. 2.6).

2.6 Discussion

In this study, model predictions for diotic detection based on three different single cues (the CB, ES, and PO models) and combinations of these cues (the LC and LRT models) were tested with detection patterns for three different sets of reproducible noise waveforms. The LRT model provided significantly better predictions of hit rates than any of the single-cue models for all three datasets and of FA rates for the 100-Hz bandwidth waveforms. Using the LRT-based detection model to predict listeners' detection performance is not new. Siebert (1970), Colburn (1973) and Heinz *et al.* (2001) used a similar strategy to predict frequency, interaural time, and level discrimination data from model auditory-nerve fibers. However, these linear models predicted listeners' discrimination thresholds using Possion-distributed model auditory-nerve responses; whereas, in the current study, the Gaussian-distributed cue values yielded a nonlinear cue-combination model to predict listeners' detection patterns.

2.6.1 Alternative Models based on Envelope Cues

For all three datasets studied here, the envelope slope cue was robust in predicting listeners' detection patterns. Wojtczak and Viemeister (1999) showed that the envelope cue was also important for understanding intensity increment discrimination and amplitude-modulation detection experiments. They found that a decision variable based on the ratio between the maximum of the envelope and its minimum could explain the linear relationship between the intensity increment discrimination and amplitude-modulation detection thresholds. A similar max/min statistic was tested on the current datasets; however, this model's predictions were not significantly correlated to listeners' performance. In addition, envelope energy, computed as the sum of the energy in the non-zero frequency components, did not explain a significant amount of listeners' performance. Thus, a detection variable based on envelope fluctuations, such as that used in the ES model (Richards, 1992), outperformed other envelope-based variables for detailed predictions of performance in tone-in-noise detection tasks.

Dau *et al.* (1997) extended their "effective" signal processing model (Dau *et al.*, 1996b) with a modulation filter bank and predicted thresholds for modulation detection and masking with random noises. Results from their study are consistent with auditory tuning to both audio and modulation frequency. They also showed that a bank of bandpass modulation filters predicted the trends of listeners' thresholds across many signal and masking conditions, whereas predictions using lowpass modulation filters (Viemeister, 1979) failed. Consistent with the implications of Dau *et al.*, (1997) that envelope cues are processed in different modulation frequency bands, the ES model with a bandpass modulation filter was used in the current study. However, only one bandpass modulation filter was required here, because lower or higher modulation frequencies did not provide information about the difference between noise-alone and 500-Hz tone-plus-noise stimuli (Fig. 2.2). It was shown that this modified ES model yielded better predictions of listeners' detection results than the original ES model.

In addition, frozen noise stimuli were used in Dau *et al.* (1996b) study of detection in noise. In that study, listeners' thresholds for detecting sinusoids of different durations, onset times, onset phases, or frequencies were predicted by their effective model (without modulation filters) (Dau *et al.*, 1996a). Direct comparisons between their results and the results presented here are difficult. In their three-interval forced-choice test, the same frozen noise was used in all intervals, providing the potential for detailed comparisons across intervals. Their model structure, which utilizes a comparison between noise-alone and tone-plus-noise representations, is appropriate for such a task. However, in the datasets analyzed here, a single frozen noise-alone or tone-plus-noise stimulus was presented in a one-interval force-choice task, and the noise for each trial was selected from an ensemble of waveforms. The models applied here were appropriate for this single-interval task; these models involved comparisons of cues for a single trial to distributions of cue values, but not the cues for a particular waveform. Furthermore, the waveforms studied here consisted of tone and noise waveforms that were gated simultaneously, whereas Dau *et al.*'s (1996b) stimuli were short-duration tones presented at a delay during a longer masking noise, making direct comparisons across the studies difficult.

For single-cue models, the "multiple-look" strategy (Viemeister and Wakefield, 1991) suggests that listeners might extract cues from short durations of the whole waveform in detection and discrimination tests. A similar strategy was tested in the current study by segmenting waveforms into equal-duration epochs. However, predictions based on the multiple-epoch scheme were not significantly different than those based on the single-epoch scheme for either single-cue or multiple-cue models. Thus, results presented above were all based on the single-epoch scheme.

2.6.2 Linear vs. Nonlinear Cue Combination

Davidson *et al.* (2006; 2009a) used different single-cue models to predict listeners' detection performance for the three datasets used in the current study, however, none of the single-cue models could explain the predictable variance. In another study focused on the 50-Hz bandwidth equal-energy waveforms, Davidson *et al.* (2009b) pointed out that a linear combination of the two cues could not explain listeners' detection patterns and suggested the future consideration of models based on nonlinear combinations of cues. Results from these three studies motivated the nonlinear LRT-based multiple-cue model that was tested in this study. Because DVs were computed from a logarithmic likelihood ratio of cue values given noise-alone and tone-plus-noise waveforms, the degree of similarity between the covariance matrices under these conditions determined whether the combination of cues was linear or nonlinear. In the current study, the covariance matrices for noise-alone and tone-plus-noise conditions were different. For the three datasets tested, model predictions of hit rates based on the nonlinear LRT model were significantly better than those based on any of the single-cue models, whereas predictions of FA rates were significantly better for the 100-Hz bandwidth waveform but not for the other two datasets.

In order to understand the difference between the LRT model and the linear cue-combination model, the weights of the different cues in the models (Eq. 2) were inspected (see Appendix A). Recall, that for the linear model the weights are based on the reliability of each single cue (the inverse of the variance), thus higher weights are assigned to more reliable cues. Inspection of weights for the linear cue-combination model showed that CB was the dominant cue and PO had the least significant weight.

Note that for the LRT model the predictions for hit and FA rates were computed with the same model, in which the weights were computed from the distributions of cue values, i.e. the same covariance matrices were used to provide weights for both hits and FAs. For the LRT model, the relationships between different single cues were determined by computing their covariance. Thus, in addition to single cues, pairs of single cues also contributed to the DV in the LRT model. For the 100-Hz bandwidth waveforms, CB, ES, and PO single cues were assigned approximately equal positive weights, whereas the pairs of CB and ES, and ES and PO cues were assigned approximately equal negative weights that were less than the positive weights. For the 2900-Hz bandwidth waveforms, the weight for the CB cue was twice as large as for the ES cue and for the pair of CB and ES cues, and these three weights dominated the weighting matrix. The higher weight for the CB cue was not surprising, because this cue explained more variance than the ES or PO cues for both the 100- Hz and 2900-Hz waveforms (Fig. 2.6). However, for the 50-Hz equal-energy waveforms, even though the CB cue was outperformed by the ES cue in single-cue model predictions, the significantly smaller variance of the CB cue resulting from the equal-energy waveforms yielded a higher weight to the CB cue in the LRT model. Similarly, consistent with the robustness of the ES cue for the single-cue predictions, it was assigned a higher weight than the PO cue. In addition, the weighting matrix of individual listeners was similar to that of the average listener, suggesting that the assumption that listeners used a similar strategy for tone detection in these experiments was reasonable.

2.6.3 Consideration of the Equal-Energy Predictions

Further analysis for the CB cue of the 50-Hz bandwidth equal-energy waveforms showed that small energy differences between waveforms were introduced when the waveforms were passed through the gammatone filter used to calculate DVs of the CB model. Although model predictions from the CB model explained around 30% of the variance in the detection patterns, the absolute size of the energy differences was negligible (Davidson *et al.*, 2009a). Inspection of the DVs from the CB model showed that average sound level difference among fifty tone-plus-noise and noise-alone waveforms was 0.1 dB and 0.2 dB, respectively. Thus, the predictions achieved by the CB model for the narrowband equal-energy condition are likely to be an artifact of the correlation between cues. In addition, the envelope cue was able to explain a significant amount of the variance in the detection pattern, confirming the robustness of the envelope cue, as in previous studies (Kidd *et al.*, 1989; Richards, 1992; Zhang, 2004; Davidson *et al.*, 2009a).

Model predictions based on the LRT model for the 2900-Hz and the 100-Hz bandwidth waveforms were close to the predictable variance; however, predictions for the 50-Hz bandwidth equal-energy waveforms were lower than the predictable variance. Based on the analysis from the weighting strategy above, the CB cue dominated the weighting matrix for the 50-Hz dataset. However, the CB cue was not as effective as the ES cue for the equal-energy waveforms (Fig. 2.6c). Thus, listeners may use alternative strategies to the optimal LRT-based method for the equal-energy narrowband waveforms.

2.6.4 Future Directions

Given that predictions based on the LRT model were most consistent with listeners' detection patterns, it is interesting to ask whether LRT-type processing is observed along the auditory pathway. Because the auditory nerve is the only path from the inner ear to the brain, the nonlinear response of the auditory nerve contains all information available to the central nervous system. Inspection of auditory-nerve (AN) model responses (Zilany *et al.*, 2009) would be a necessary first step. Rate, synchrony and fluctuation of the post-stimulus time histogram (PSTH) computed from model responses could represent the energy, fine-structure, and envelope cues of the stimulus. However, given that both on- and off-frequency AN fibers would respond to the stimuli, it would be interesting to investigate an optimal way to combine these cues.

In addition, responses from higher levels in the brain, such as the cochlear nuclei and inferior colliculus (IC), are also likely to convey information observed from the LRT model. In particular, the IC is a nearly obligatory pathway from the lower brainstem nuclei to higher processing centers. Analysis of IC model responses (Nelson and Carney, 2004) could be tested with responses from the LRT model.

Lastly, internal noise (Spiegel and Green, 1981) was not included in the current signal-processing type model. However, internal noise could be introduced in physiological models as an additive or multiplicative noise to further understand the difference of detection performance among individual listeners.

2.7 Summary

In this study, model predictions for diotic detection based on three different single cues (the CB, ES, and PO models) and combinations of these cues (the LC and LRT models) were tested with detection patterns for three different sets of reproducible noise waveforms. The LRT model, which is an optimal combination of energy and temporal cues, provided significantly better predictions of hit rates than any of the single-cue models or the LC model for all three datasets and of FA rates for the 100-Hz bandwidth waveforms.

ACKNOWLEDGEMENTS

This work was supported by grant NIH-NIDCD R01-DC010813 (LHC) and by NSF CAREER award CCF-1054687 (AV). We would like to thank Kristina Abrams, Kelly-Jo Koch, Dr. Tianhao Li, Douglas Schwartz, and the students in the lab for their helpful suggestions on preparing the manuscript.

Appendix A: Weights for the Nonlinear Cue-Combination Model

The weights for the LRT nonlinear cue-combination model are shown in Tables A1, A2, and A3 for 100-Hz and 2900-Hz bandwidth waveforms and for the 50-Hz bandwidth equal-energy waveforms. In each table, the diagonal entries indicate weights for single cues (e.g. CB, ES, and PO), and the off-diagonal entries indicate weights for two cues (e.g. CB-ES, CB-PO, and ES-PO). Note that the weights are symmetric along the diagonal entries and the weight matrix is normalized to have a sum of one.

Table A1: Weights for 100-Hz bandwidth waveforms.

Weights for Cues	CB	ES	PO
CB	7.30	-6.26	1.41
ES	-6.26	8.40	-8.16
PO	1.41	-8.16	11.34

Table A2: Weights for 2900-Hz bandwidth waveforms.

Weights for Cues	CB	ES	PO
CB	0.43	0.11	-0.00
ES	0.11	0.17	0.07
PO	-0.00	0.07	0.05

Table A3: Weights for 50-Hz bandwidth equal-energy waveforms.

Weights for Cues	CB	ES	PO
CB	1.03	-0.01	0.01
ES	-0.01	-0.11	-0.00
PO	0.01	-0.00	0.05

Bibliography

Blodgett, H. C., Jeffress, L. A., and Taylor, R. W. (1958). “Relation of masked threshold to signal-duration for interaural phase combination,” Am. J. Psychol. 71, 283-290.

Blodgett, H. C., Jeffress, L. A., and Whitworth, R. H. (1962). “Effect of noise at one ear on the masked threshold for tone at the other,” J. Acoust. Soc. Am. 34, 979-981.

Breebaart, J., van der Par, S., and Kohlrausch, A. (2001a). “Binaural processing model based on contralateral inhibition I. Model structure,” J. Acoust. Soc. Am. 110, 1074–1088.

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. Acust. Acust. 88, 334–347.

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

Dau, T., Püschel, D., and Kohlrausch, A. (1996a). “A quantitative model of the “effective” signal processing in the auditory system. I. Model structure,” J. Acoust. Soc. Am. 99, 3615–3622.

Dau, T., Püschel, D., and Kohlrausch, A. (1996b). “A quantitative model of the “effective” signal processing in the auditory system. II. Simulations and measurements,” J. Acoust. Soc. Am. 99, 3623-3631.

Dau. T., Kollmeier, B., and Kohlrausch. A. (1997). “Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers,” J. Acoust. Soc. Am. 102, 2892-2905.

Davidson, S. A., Gilkey, R. H., Colburn, H. S., and Carney, L. H. (2006). “Binaural detection with narrowband and wideband reproducible noise maskers. III. Monaural and diotic detection and model results,” J. Acoust. Soc. Am. 119, 2258-2275.

Davidson, S. A., Gilkey, R. H., Colburn, H. S., and Carney, L. H. (2009a). “An evaluation of models for diotic and dichotic detection in reproducible noises,” J. Acoust. Soc. Am. 126, 1906-1925.

Davidson, S. A., Gilkey, R. H., Colburn, H. S., and Carney, L. H. (2009b). “Diotic and dichotic detection with reproducible chimeric stimuli,” J. Acoust. Soc. Am. 126, 1889-1905.

Dolan, T. R., and Robinson, D. E. (1967). “Explanation of masking-level difference that result from interaural intensive disparities of noise,” J. Acoust. Soc. Am. 42, 977-981.

Evilsizer, M. E., Gilkey, R. H., Mason, C. R., Colburn, H. S., and Carney, L. H. (2002). “Binaural detection with narrowband and wideband reproducible maskers: I. Results for human,” J. Acoust. Soc. Am. 111, 333-345.

Fletcher, H. (1940). “Auditory patterns,” Rev. Mod. Phys. 12, 47–65.

Gilkey, R. H., Robinson, D. E., and Hanna, T. E. (1985). “Effects of masker waveform and signal-to-masker phase relation on diotic and dichotic masking by reproducible noise,” J. Acoust. Soc. Am. 78, 1207-1219.

Gilkey, R. H., and Robinson, D. E. (1986). “Models of auditory masking: A molecular psychophysical approach,” J. Acoust. Soc. Am. 79, 1499-1510.

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.

Kidd, G. Jr., Mason, C. R., Brantley, M. A., and Owen, G. A. (1989). “Roving-level tone-in-noise detection,” J. Acoust. Soc. Am. 86, 1310-1317.

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.

Neter, J., Kutner, M. H., Nachtsheim, C. J., and Wasserman, W. (1996). *Applied linear statistical models.* (WBC McGraw-Hill, Boston), pp. 641.

Richards, V. M. (1992). “The delectability of a tone added to narrow bans of equal energy noise,” J. Acoust. Soc. Am. 91, 3424-3435.

Siebert, W. M. (1970). “Frequency discrimination in the auditory system: place or periodicity mechanisms?” Proc. IEEE. 58, 723-730.

Spiegel, M. F., and Green, D. M. (1981). “Two procedures for estimating internal noise,” J. Acoust. Soc. Am. 70, 69-73.

Van Trees, H. L. (1968). *Detection, estimation, and modulation theory. Part I. Detection, estimation and linear modulation theory* (John Wiley & Sons, New York), Chap. 2, pp. 26-36.

Viemeister, N. F. (1979). “Temporal modulation transfer function based upon modulation thresholds,” J. Acoust. Soc. Am. 66, 1364-1380.

Viemeister, N. F., and Wakefield, G. H. (1991). “Temporal integration and multiple looks,” J. Acoust. Soc. Am. 90, 858-865.

Wojtczak, M., and Viemeister, N. F. (1999). “Intensity discrimination and detection of amplitude modulation,” J. Acoust. Soc. Am. 106, 1917-1924.

Yuille, A. L., and Bulthoff, H. H. (1996). “Bayesian decision theory and psychophysics,” in *Perception as Bayesian Inference*, edited by Knill, D. C., and Richards, W., (Cambridge University Press, London), Part 1, pp. 123-161.

Zhang, X. (2004). “Cross-frequency coincidence detection in the processing of complex sounds,” Ph.D. thesis, Boston University, Boston, MA.

Zilany, M. S., Bruce, I. C., Nelson, P. C., and Carney, L. H. (2009). “A phenomenological model of the synapse between the inner hair cell and auditory nerve: long-term adaptation with power-law dynamics,” J. Acoust. Soc. Am. 126, 2390-2412.

Chapter 3

Prediction of Binaural Detection with Narrowband and Wideband Reproducible Noise Maskers using Interaural Time, Level, and Envelope Differences

3.1 Abstract

The addition of out-of-phase tones to in-phase noises results in dynamic interaural level difference (ILD) and interaural time difference (ITD) cues for the dichotic tone-in-noise detection task. Several models have been used to predict listeners' detection performance based on ILD, ITD, or different combinations of the two cues. The models can be tested using detection performance from an ensemble of reproducible-noise maskers. Previous models cannot predict listeners' detection performance for reproducible-noise maskers without fitting the data. Here, two models were tested for narrowband and wideband reproducible-noise experiments. One model was a linear combination of ILD and ITD that included the generally ignored correlation between the two cues. The other model was based on a newly proposed cue, the slope of the interaural envelope difference (SIED). Predictions from both models explained a significant portion of listeners' performance for detection of a 500-Hz tone in wideband noise. Predictions based on the SIED approached the predictable variance in the wideband condition. The SIED represented a nonlinear combination of ILD and ITD, with the latter cue

dominating. Listeners did not use a common strategy (cue) to detect tones in the narrowband condition and may use different single frequencies or different combinations of frequency channels.

3.2 Introduction

Tone-in-noise detection has been studied for decades; however, it is still not clear which cue or combination of cues can explain listeners' performance. Although model predictions based on a nonlinear combination of cues can explain a substantial amount of listeners' detection patterns in the diotic condition (Mao *et al.*, 2013), no existing model can satisfactorily explain listeners' performance for the dichotic condition. In this study, two binaural models based on combinations of interaural level and time differences are proposed to predict listeners' dichotic performance. This work is part of an ongoing series of experimental and modeling studies of binaural detection (Evilsizer *et al*., 2002; Zheng *et al*., 2002; Davidson *et al*., 2006).

In early studies of binaural detection, random noise waveforms were generated in each trial for each listener (Blodgett et al., 1958; Blodgett et al., 1962; Dolan and Robinson, 1967), and detection performance was averaged across listeners and waveforms, described as 'molar-level' performance by Green (1964). In order to test model predictions and compare the effectiveness of different cues, it is useful to consider detection performance on a waveform-by-waveform basis ('molecular-level') for each listener (e.g., Schönfelder and Wichmann, 2013). Gilkey *et al.* (1985) and Gilkey and Robinson (1986) found that averaging detection performance across masker waveforms obscures the differences across individual waveforms and listeners, suggesting the utility of a more “molecular-level” approach. However, “molecular-level” predictions are difficult to obtain because of the unknown internal noise for each listener and the possible use of different cues by different listeners. The current study analyzed data from Evilsizer *et al.* (2002) and Isabelle and Colburn (1991), plus three additional listeners tested with the same stimuli. In those studies, a ‘quasi-molecular’ method was applied, in which the noise masker for each trial was randomly selected from a set of reproducible-noise waveforms. In the current study, model predictions were computed for the dichotic condition, in which in-phase noise and out-of-phase tone waveforms were presented to both ears.

In each single-interval trial during the task, listeners responded “tone present” or “tone not present” for each binaural noise-alone or tone-plus-noise stimulus. Detection performance was described in terms of hit rates, the proportion of tone-plus-noise trials in which listeners correctly responded “tone present”, and false-alarm (FA) rates, the proportion of noise-alone trials in which listeners incorrectly responded “tone present”. The set of hit and FA rates for an ensemble of reproducible waveforms is referred to as the detection pattern (Fig. 3.1; Davidson *et al.*, 2006).

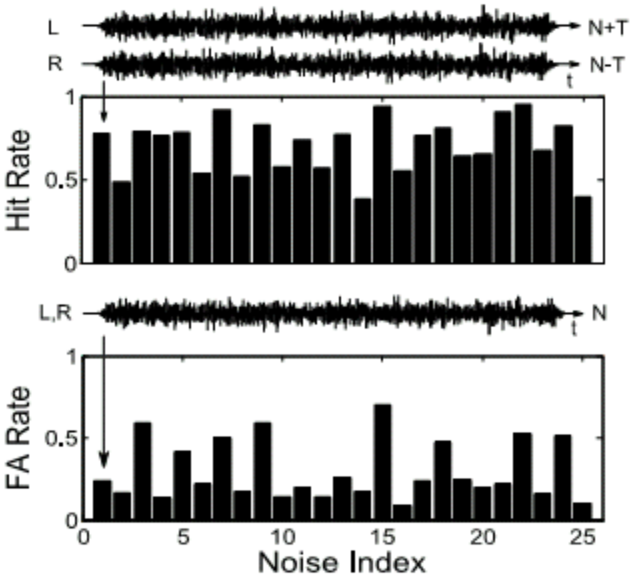

Figure 3.1: A detection pattern for the average listener comprises hit and FA rates for each wideband (2900-Hz bandwidth) dichotic reproducible waveform averaged across six individual listeners. The x-axis shows the index of the reproducible waveform. Insets show examples of the dichotic tone-plus-noise at left (N+T) and right (N-T) ears and the diotically presented noise-alone (N) waveforms for reproducible noise waveform number one (data from Evilsizer *et al.*, 2002 and two listeners tested recently). The tone was added at the average threshold level of the six listeners, and the spectrum level of the noise was 40 dB SPL.

In order to identify which cue or combination of cues listeners use in a dichotic tone-in-noise detection test, different models have been tested to predict detection patterns (Isabelle and Colburn, 1987; Isabelle, 1995; Goupell and Hartmann, 2007; Davidson *et al.*, 2009a). For each model, a set of decision variables (DVs), each derived from a specific feature or combination of features of the waveforms, is compared with the detection patterns. Model predictions can be evaluated based on the amount of variance in the detection pattern that can be explained by calculating the squared correlation between DVs and detection patterns.

Several models based on binaural energy differences, interaural level differences (ILDs), and interaural time differences (ITDs) cues have been tested. Durlach (1963) proposed the equalization and cancellation model (EC), an energy-based model that subtracts the internal stimulus representation in one ear from that in the other ear after equalizing the masking waveforms in both ears. Isabelle (1995) tested the normalized interaural cross-correlation model (NCC) that computes the correlation of the waveforms at the two ears. The NCC model is related to the EC model because the DVs from both models are highly correlated to the energy of the noise-alone waveforms (Colburn *et al.*, 1997). However, Isabelle (1995) showed that neither of these energy-related models could explain a significant amount of the variance in the dichotic detection patterns.

In addition to the energy-based cues, the combination of binaural out-of-phase tones with in-phase noises results in dynamic ILD and ITD cues. DVs computed from the sample standard deviation of the ILD (σ_{ILD}) and ITD (σ_{ITD}) have been used to predict detection patterns (Isabelle, 1995). Isabelle (1995) also calculated the peak deviation of ITD (M_β) by using the rare, large ITD magnitudes. The DV from the M_β model could be interpreted as the proportion of stimulus duration during which the instantaneous ITD magnitude exceeds a certain threshold β. Although some of these ILD- or ITD-based models could predict a significant amount of the variance in a few listeners' detection patterns, none of these models worked for all listeners (Isabelle, 1995; Davidson *et al.,* 2009a).

Given that ILD and ITD represent different features of the waveform, it is reasonable to expect that the combination of these cues could capture more information about the waveform than either one alone. Isabelle and Colburn (1987) combined the two interaural difference cues by using a sum-of-squares model (SS). The DV was computed as a linear combination of the variance of ILD and ITD, and the weights were found by fitting the detection patterns (Isabelle and Colburn, 2004). Predictions from the SS model would be optimal if ILD and ITD were Gaussian-distributed and independent. However, it has been shown that the two cues are correlated (Zurek, 1991; Isabelle, 1995). Isabelle (1995) also combined ILD and ITD cues based on the deviation in lateral position (LP). The LP model was first used by Hafter (1971) to account for time-intensity trading in lateralization tests. The DV from the LP model was calculated as the mean magnitude of the lateralization position, in which ILD and ITD were combined through a trading ratio. The SS and LP models could not explain a significant proportion of variance of the listeners' detection patterns. More recently, Goupell and Hartmann (2007) proposed "independent-center" and "auditory-image" models that linearly combined ILD and ITD to predict listeners' performance for interaural correlation detection; the difference between these two models was the sequence of combining ILD and ITD information and integrating across time. Predictions from Goupell and Hartmann's models were significantly correlated with detection patterns for about half of the listeners (Davidson *et al.*, 2009a). However, Davidson *et al.* (2009a) found by examining data from each listener that either ILD or ITD dominated in Goupell and Hartmann's linear combinations, suggesting that instead of combining ILD and ITD, in fact only the better of the two cues was used by the models.

The goal of the study presented here was to test the hypothesis that significantly better predictions of detection patterns could be obtained from models that combined ILD and ITD cues. Two models were tested for this hypothesis in the current study: a modified ILD-ITD combination model which takes into account the correlation between the two cues and a model based on the slope of the interaural envelope difference (SIED).

In the first model, a modified linear combination of ILD and ITD cues that weighted the two cues based on their covariance matrix (Oruç *et al.*, 2003) was used to compute the DV. By computing weights from the covariance matrix of cue values, it is possible to avoid fitting the detection data as has been done in previous studies (Isabelle and Colburn, 1987; Goupell and Hartmann, 2007; Davidson *et al.*, 2009b). In addition, waveforms were analyzed using multiple epochs, with each epoch weighted separately. Model predictions using this method of combining ILD and ITD cues were significantly better than previous dichotic model predictions.

In the second method, the interaural envelope difference was used to derive the DV. Predictions based on envelope cues from Richards (1992), Zhang (2004), Davidson *et al.* (2009a), and Mao *et al.* (2013) showed that the envelope-slope cue is robust and successful in predicting diotic detection patterns, which motivated the exploration of envelope cues in the dichotic condition. The envelope-slope cue focuses on changes in monaural envelope fluctuations, whereas binaural differences are key for dichotic detection. Thus modification of the diotic envelope-slope cue was required in order to consider the envelopes from both ears. A binaural envelope cue, the SIED, based on the slope of the interaural envelope difference (SIED) was proposed and tested in the second model of this study. Moreover, the SIED was shown to be related to both ILD and ITD information in a nonlinear manner. Predictions of the wideband detection patterns based on the SIED cue were significantly better than predictions using any single cue or any linear combination of ILD and ITD cues. In contrast, none of these cues provided significant predictions of the detection patterns for the narrowband condition, nor did the listeners employ a common strategy in that condition.

Given that there are no interaural differences in the noise-alone trials in a binaural-detection task, the prediction of FA rates in the dichotic condition is not possible with models based on interaural differences. Although internal noise is possibly an important factor to explain the FA rates, the statistical properties of internal noise are unknown. Furthermore, a simple additive noise would not explain the FA rates because such a noise would be averaged out in the "quasi-molecular" data sets analyzed here (because FA rates are computed by averaging multiple noise-alone trials), and a more complex noise model would thus be required. Model predictions for FA rates were not included in the current study.

3.3 Description of Data

The data analyzed in the current study were obtained from two previous experiments (Isabelle and Colburn, 1991; Evilsizer *et al.*, 2002). Three additional listeners were tested with the stimuli from Evilsizer *et al.* (2002), and one of them was also tested with the stimuli from Isabelle and Colburn (1991). A total of six listeners were tested with wideband stimuli, and ten listeners were tested with narrowband stimuli.

In the Evilsizer *et al.* (2002) study, four listeners (S1-S4 in the current study) were tested with a set of twenty-five reproducible noise waveforms. Both narrowband (452-552 Hz) and wideband (100-3000 Hz) noise waveforms of 300-ms duration and a spectrum level of 40 dB SPL (e.g., approximately 75 dB SPL root-mean-square (RMS) level for the wideband condition, and 60 dB SPL RMS level for the narrowband condition) were tested. The spectral content of each narrowband waveform was matched to that of the corresponding frequency range of each wideband waveform. A 500-Hz sinusoidal waveform with 300-ms duration was used, and the tone level was set to equal to the detection threshold of each listener. For the wideband condition, the tone level for the average listener was computed as the mean of the tone levels for all individual listeners. Three additional listeners (S5-S7) were tested with similar techniques, except that a two-down one-up tracking procedure (Levitt, 1971) replaced the fixed-level testing used by Evilsizer *et al.* (2002). Correct-answer feedback was provided after each trial. Listeners' detection thresholds were computed as the mean of the reversals (excluding the first six reversals) in all tracks. In each 100-trial track, trials within a 2-dB range of the detection threshold were used to create the detection patterns. Each listener's patterns were highly consistent over the course of the test.

In Isabelle and Colburn's experiment (1991), three listeners (S8-S10 in the current study) were tested with ten narrowband (445-561 Hz) noises. The duration of the waveform was 300 ms, and the noise spectrum level was 54 dB SPL. The tone frequency was 500 Hz, and its level was matched to listener's detection threshold. Out-of-phase tones (differing by 180 degrees) were added to identical noises for the $N_OS\pi$ tone-plus-noise trials. After identifying each listener's detection threshold in preliminary tests, listeners were tested extensively near their thresholds to estimate their detection patterns. One additional listener (S7) was tested with the same stimuli and similar techniques, except that a two-down one-up tracking procedure (Levitt, 1971) replaced the fixed-level testing used by Isabelle and Colburn (1991). This listener's detection patterns were significantly consistent over the course of the test.

Listeners' detection patterns were described in terms of hit and FA rates, based on the probability that they responded "tone present" for each noise-alone or tone-plus-noise waveform (details of the experiments can be found in Evilsizer *et al.*, 2002, and Isabelle and Colburn, 1991). Figure 3.1 shows the detection pattern of the average listener (i.e., the average detection pattern across six individual listeners who were tested using Evilsizer *et al.*'s stimuli) for the wideband dichotic condition. The detection patterns were reliable, as each listener's detection pattern was highly consistent over the course of the experiment: the average Pearson product-moment correlation of seven listeners between the first-half and second-half of the trials was 0.70 for the Evilsizer *et al*. (2002) narrowband condition, and 0.81 for the six listeners tested with Evilsizer *et al*. (2002) wideband condition. The average Pearson product-moment correlation was not available for the Isabelle and Colburn (1991) stimuli.

Table 3.1a shows that the detection patterns were significantly correlated for all pairs of listeners for the wideband stimuli. Table 3.1b and c show that detection patterns are significantly correlated for six out of twenty-one pairs of listeners for the narrowband stimuli from Evilsizer *et al.* (2002) (r = 0.40, p<0.05 for t-test), and for one out of six pairs of listeners for the stimuli from Isabelle and Colburn (1991) (r=0.63, p<0.05 for t-test). Note that the sign of the correlation varied across listeners for narrowband stimuli from both studies. Note also that the significance criterion differed for the Evilsizer *et al.* (2002) and Isabelle and Colburn (1991) studies due to the different numbers of waveforms used in each study.

There are a few differences between the two narrowband studies that are worth noting. The overall noise level was 15 dB higher for stimuli from Isabelle and Colburn (1991) than for the narrowband stimuli from Evilsizer *et al.* (2002). In addition, for the narrowband waveforms, two out seven listeners tested with the Evilsizer *et al.* (2002) stimuli had thresholds at similar signal-to-noise ratio[1] (SNR) as listeners from Isabelle and Colburn (1991) study, while the remaining listeners tested with the Evilsizer *et al.* (2002) stimuli had higher SNRs. In general, threshold SNRs were more variable across listeners in the narrowband condition, as compared to the wideband condition (Table 3.2).

Table 3.1: correlations between each pair of listeners in narrowband and wideband conditions (bold values indicate significant correlations)

(a) Pair-wise correlations of six listeners' hit rates for wideband stimuli from the Evilsizer *et al.* (2002) study

	S2	S3	S4	S5	S6
S1	**0.56**	**0.63**	**0.68**	**0.42**	**0.48**
S2		**0.62**	**0.51**	**0.58**	**0.64**
S3			**0.62**	**0.58**	**0.53**
S4				**0.71**	**0.69**
S5					**0.70**

(b) Pair-wise correlations of seven listeners' hit rates for narrowband stimuli from the Evilsizer *et al*. (2002) study

	S2	S3	S4	S5	S6	S7
S1	**-0.59**	**0.50**	-0.05	**-0.40**	**-0.54**	0.01
S2		-0.19	-0.32	0.20	0.34	0.16
S3			-0.18	-0.24	**-0.49**	0.08
S4				0.37	0.13	0.09
S5					**0.70**	0.19
S6						-0.04

(c) Pair-wise correlations of four listeners' hit rates for narrowband stimuli from the Isabelle and Colburn (1991) study

	S8	S9	S10
S7	-0.16	-0.22	0.07
S8		**0.69**	0.50
S9			0.54

Table 3.2: Listeners' threshold tone-levels (top, dB SPL) and SNRs[1] (bottom *italic*, dB) for wideband and narrowband conditions. Noise spectrum level in Evilsizer *et al.* (2002) was 40 dB SPL (overall noise level was approximately 75 dB SPL for the wideband condition, and 60 dB SPL for the narrowband condition), and 54 dB SPL (overall noise level was 75 dB SPL) in Isabelle and Colburn (1991).

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Evilsizer *et al.* (2002) Wideband	45.0 *-30*	43.0 *-32*	50.0 *-25*	46.0 *-29*	47.5 *-27.5*	44.7 *-30.0*				
Evilsizer *et al.* (2002) Narrowband	39.0 *-21*	49.0 *-11*	47.0 *-13*	39.0 *-21*	53.6 *-6.4*	49.8 *-10.2*	44.4 *-15.6*			
Isabelle and Colburn (1991) Narrowband							57.8 *-17.2*	55.0 *-20*	55.0 *-20*	55.0 *-20*

For the wideband condition, detection patterns for an "average listener" were computed from the averaged patterns of the six individual listeners, all of whose patterns were significantly correlated (Table 3.1a). For the narrowband stimuli, an average listener was not used because listeners' detections patterns were not significantly correlated, in general, suggesting that they used different strategies in the detection test. Instead, only analyses of individual listeners are presented below for the narrowband condition.

3.4 Methods

In this study, it was hypothesized that significantly better predictions of the dichotic detection patterns could be achieved using DVs that combined ILD and ITD cues. First, single-cue (ILD or ITD) DVs that combined information across time epochs are described. Next, DVs combining ILD and ITD across time epochs are presented. Finally, results for the envelope-related SIED model that includes a nonlinear combination of ILD and ITD information are described.

3.4.1 DVs that Combine Single-Cue Information across Multiple Time Epochs

For both ILD- and ITD-based and SIED DVs, the 300-ms duration waveform was separated into several equal-duration time epochs, and local cue information was obtained from each time epoch. For the wideband condition, listeners were likely to use similar cues or combinations of cues for the detection test given that their detection patterns were significantly correlated. Thus, in order to select the duration of epochs for each cue, DVs for the average listener were computed for different durations of epochs that were divisors of 300 ms, *e.g.*, 300 ms, 150 ms, and 100 ms, *etc*. The Pearson product-moment correlation was calculated to quantitatively compare DVs from different durations of epochs to the detection pattern (hit rates, or percentage of correct identification of tone presence) of the average listener. For each cue, the number of epochs that yielded the highest correlation for the average listener was chosen and used for all listeners. For the narrowband condition, the same multiple-epoch scheme was tested for each individual listener since no average listener was used in this condition.

DVs were computed as the mean of local single-cue information across epochs. Figure 3.2 shows a schematic diagram of the single-cue multiple-epoch model, in which C_i represents the local cue value in the i^{th} epoch and n is the number of epochs. For the ILD or ITD cue, the local cue is the sample standard deviation of ILD or ITD; for the SIED cue, the local cue is the DV from the SIED model. The analytical signal was used to obtain the ILD, ITD, and SIED cues. Both non-overlapping and half-overlapping windows were tested for the multiple-epoch scheme to investigate different ways of computing local cues. No difference in the results was observed between different overlap-window methods. The advantage of applying the multiple-epoch scheme is that a substantial value of the DV could be obtained when there were large variations of local cues only in certain epochs. However, in the single-epoch scheme, these large variations could be lost if the averaged variation of the cue across the entire waveform were small.

Note that DVs for the ILD, ITD, and SIED cues for the wideband condition were computed after applying a gammatone filter with center frequency at 500Hz. For the narrowband condition, the gammatone filter was always used for the SIED cue, which allowed examination of different frequency channels (see below). In the narrowband condition, there were no significant differences between model predictions based on ILD or ITD cues with or without the gammatone filter. In order to match the narrowband ILD and ITD results in Davidson *et al.* (2006) and Isabelle and Colburn (1991), in which no gammatone filter was used, results shown below for the narrowband ILD and ITD cues were computed without the gammatone filter.

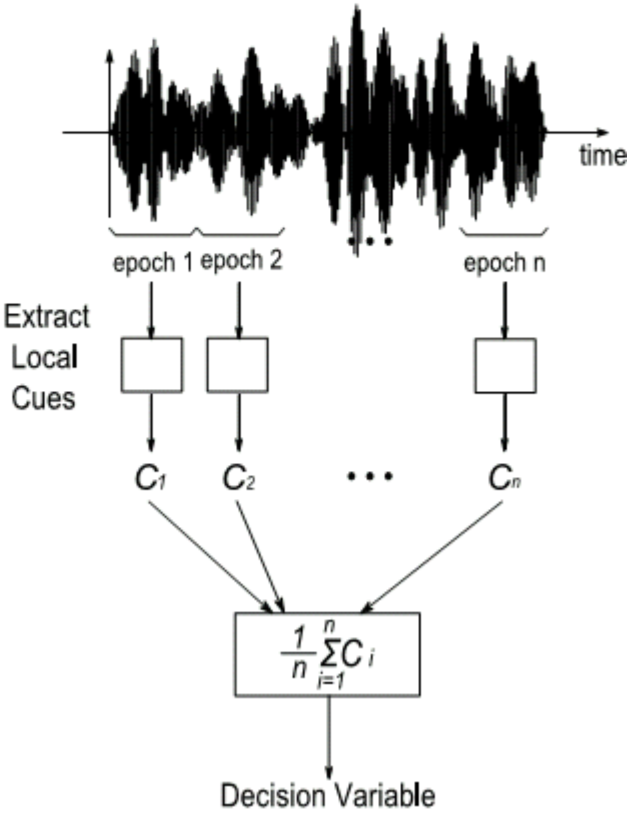

Figure 3.2: A schematic diagram illustrates a DV that was computed by combining local cue information across epochs for a single cue (ILD, ITD, or SIED). The waveform was separated into several equal-duration epochs along the time axis, and the local cues (C_i) were obtained. The DV was the mean value of the cue across epochs.

3.4.2 DVs that Combine ILD and ITD Cues

As Isabelle and Colburn (1987) pointed out in their sum-of-squares model, an optimal linear combination of the ILD and ITD cues could be achieved if these two cues were Gaussian-distributed and independent. In that ideal case, the optimal combination would yield the minimum variance of the combined cues by weighting each cue proportional to the inverse of its variance. Given that a cue with a smaller variance indicates a higher reliability, the optimal cue combination yielded the maximum reliability. However, ILD and ITD cues are correlated (Zurek, 1991; Isabelle 1995). Thus, consideration of the relationship between these two cues is necessary to obtain the optimal cue combination. By assigning weights w_i^{ILD} and w_i^{ITD} to the components in the $\vec{w}$ weight matrix, based on the product of the inverse of the covariance matrix $\mathbf{\Sigma}_{ILD,ITD}$ and a column vector *e* of all ones (Eq. 1), a modified linear cue combination was used that was optimal for correlated cues (Oruç *et al.*, 2003). This combination yields a decision variable, *D*, with the minimum variance of the combined cue, and in turn, the maximum reliability:

$$D = \sum_i \left(w_i^{ILD} S_{ILD}(i) + w_i^{ITD} S_{ITD}(i) \right), \tag{1}$$

where $\vec{w} = [w_i^{ILD}, w_i^{ITD}] \propto \mathbf{\Sigma}_{ILD,ITD}^{-1} e,\ e = [1,1,...,1]^T$.

3.4.3 DVs based on the Slope of the Interaural Envelope Difference (SIED) Cue

In addition to the ILD and ITD cues, an envelope-related cue, the SIED, was tested for its ability to predict listeners' tone-in-noise detection patterns. A binaural envelope cue was investigated in this study because of the success and robustness of a monaural envelope-slope (ES) cue in predicting diotic detection patterns (Richards, 1992; Zhang, 2004; Davidson *et al.*, 2009a; Mao *et al.*, 2013). DVs for the diotic ES model were computed as the integral of the absolute value of the monaural envelope fluctuations across time. When a tone is added to narrowband noise, the envelope flattens and the ES DV decreases (Richards, 1992). This *monaural* ES cue predicts a significant amount of listeners' *dichotic* performance. In addition, the monaural ES model predictions of dichotic performance can be better than predictions based on ILD or ITD cues for most listeners in the narrowband and wideband conditions. Because the envelopes at the two ears are different for the dichotic condition, and the monaural ES cue can only reflect envelope fluctuations at one ear, the SIED model was developed to quantify the fluctuations of the interaural envelope difference.

Figure 3.3 illustrates the SIED cue; the inset figures show the waveforms, monaural envelopes, and the interaural envelope difference. A fourth-order gammatone filter was used here; Johannesma et al. (1971), de Boer and de Jongh (1978), and Carney and Yin (1988) showed that the gammatone filter provides an excellent fit to both amplitude and phase properties of auditory-nerve responses. The SIED model was not intended to fully capture the auditory processing after basilar membrane filtering; rather, it was designed to be a mathematical (signal-processing type) model to test possible cues that listeners use for detection of tones in noise. The interaural envelope difference was calculated by taking the instantaneous difference between the monaural envelopes at the two ears. The instantaneous slope of the time-varying interaural envelope difference was then computed (Eq. 2),

$$y(t) = \frac{d}{dt}\left(E_{vL}(t) - E_{vR}(t)\right), \tag{2}$$

where E_{vL} and E_{vR} were the envelopes at the left and right ear respectively. Finally, the half-wave rectified slope information was integrated over time to yield the DV for the SIED cue. The half-wave rectification was applied in order to better match the SIED model to physiological models; similar model performance was obtained with full-wave rectification (considering both positive and negative slopes) or using negative slopes only. Similar to the ILD and ITD cues, the SIED cue was based on the interaural differences resulting from the binaurally out-of-phase tones. The relationship between the SIED cue and ILD and ITD cues is analyzed in the Results.

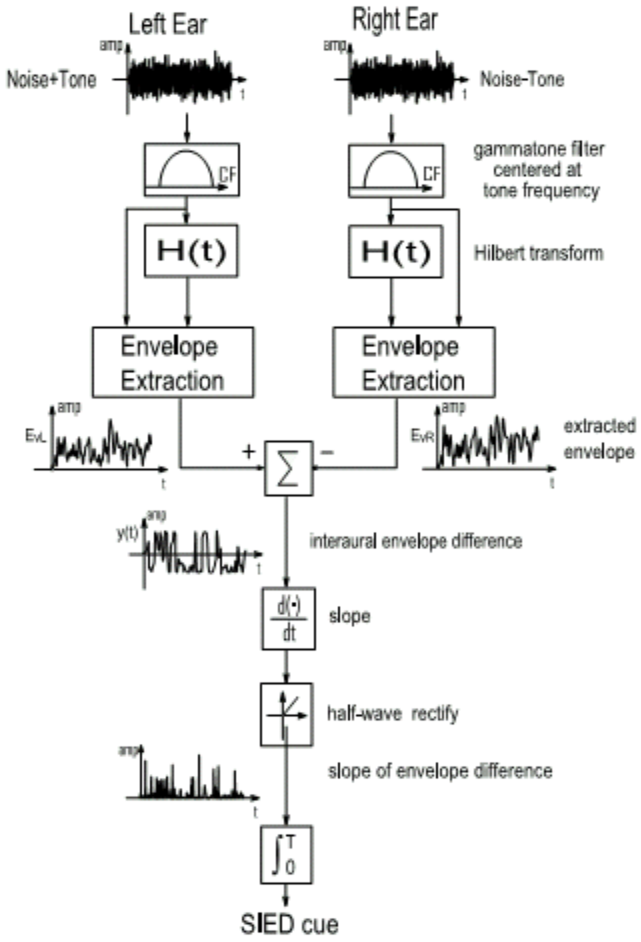

Figure 3.3: A schematic illustration of the calculation of the SIED cue. Envelopes were extracted from the analytic signals, which were obtained using the Hilbert transform of the fourth-order gammatone filtered waveforms. The center frequency of the gammatone filter was set to the tone frequency of 500Hz. The slope of the interaural envelope difference was half-wave rectified and integrated over time to obtain the SIED DV.

Model predictions could be interpreted as the explainable proportion of variance in the listeners' performance across waveforms. In order to evaluate model predictions of listeners' detection patterns, a squared Pearson product-moment correlation (r^2) was calculated between the DVs and the z-score of the detection patterns (Davidson *et al.*, 2009a). The Pearson product-moment correlation (*r*) was compared to the significance level (p<0.05 t-test), to test whether it was different from zero. The r^2 was also compared to the predictable variance to check the effectiveness of model predictions. For the wideband condition, the predictable variance was computed as the squared mean of the correlations between detection patterns of individuals and that of the average listener. Predictions based on the methods described above cannot explain the individual differences among listeners' detection patterns. In other words, a correlation value of one between model DVs and detection patterns for each individual cannot be achieved using a single model unless the listeners have identical detection patterns. However, when listeners' detection patterns are significantly correlated to each other, as in the wideband condition, the predictable variance is high and model predictions could potentially explain a large amount of the variance. Thus, for the wideband condition, the predictable variance was used as a benchmark to evaluate the overall quality of the model predictions. For the narrowband condition, detection patterns were not generally correlated across listeners, and thus neither an average listener nor the predictable variance was useful.

3.5 Results

Model predictions based on the ILD, ITD, and SIED cues are shown in this section. Because dichotic cues rely on interaural differences, which are only available for the tone-plus-noise waveforms, predictions are only shown for hit rates. Model predictions were computed for each stimulus set for individual listeners for wideband and narrowband conditions and the average listener for the wideband condition.

3.5.1 Epoch Duration for Each Cue

The epoch duration used for model predictions was chosen based on the average listener for the wideband condition and the individual listeners for the narrowband condition as described in Methods. Model predictions of the average listener's hit rates in response to wideband stimuli from the Evilsizer *et al*. (2002) study using different epoch durations for each cue are shown in Fig. 3.4. The x-axis shows different epoch durations, and the y-axis shows the proportion of variance in the detection patterns that was explained by the model. The lengths of error bars indicate the standard deviation across the individual listeners. The circles indicate predictions for the wideband conditions; the dotted lines show the predictable variance for the wideband conditions. For all the cues, no significant differences in model predictions were observed using half-overlapping or non-overlapping windows; only the results from the non-overlapping windows are shown. For ILD and ITD cues, no significant differences in model predictions across epoch duration were observed. For the SIED cue, predictions using large epoch durations were significantly more correlated to listeners' detection patterns compared with predictions from small epoch durations (< 75 ms), as expected due to the relatively long time course of envelope cues. In addition, model predictions based on the 75-ms epoch duration approached the predictable variance (squared mean of the correlations between detection patterns of individuals and that of the average listener, see below). Interestingly, the epoch length of 75 ms falls into the range of binaural integration windows (e.g., 50-200 ms) described by several studies of "binaural sluggishness"

(Grantham and Wightman, 1979; Kollmeier and Gilkey, 1990; Culling and Colburn, 2000; Kolarik and Culling, 2009).

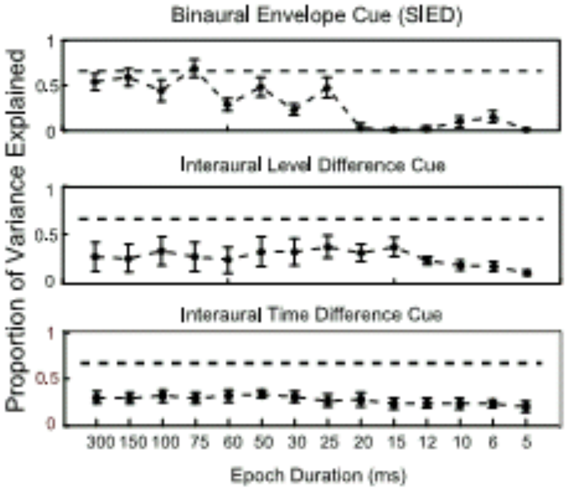

Figure 3.4: Proportion of variance explained by the SIED (upper panel), ILD (middle panel), and ITD (bottom panel) cues for the average listener, based on all responses to the Evilsizer et al. (2002) stimuli, for wideband waveforms using different epoch durations. The dotted line shows the predictable variance for the wideband conditions. The x-axis shows the epoch durations, and different filled circles represent predictions for wideband waveforms. The average listener was computed across six listeners for the wideband condition.

Model predictions of hit rates from the narrowband stimuli in the Evilsizer *et al.* (2002) and Isabelle and Colburn (1991) studies with different epoch durations for the SIED, ILD, and ITD cues were also tested. No significant differences of model predictions among different epoch durations were observed. For consistency, the epoch duration was fixed at 75 ms for ILD, ITD and SIED cues for all datasets. For models that combined ILD and ITD cues, the epoch duration was also fixed at 75 ms for stimuli from both the Evilsizer *et al.* (2002) and Isabelle and Colburn (1991) studies.

3.5.2 Model Predictions

Model predictions of hit rates for individual listeners in response to Evilsizer *et al.*'s stimuli (Fig. 3.5A and Fig. 3.5B) and individual listeners in response to Isabelle and Colburn's stimuli (Fig. 3.5C) are shown. Predictions based on ILD, ITD, the combination of ILD and ITD, and SIED cues are shown in the four groups of symbols in each panel. For each group of symbols, the open symbols indicate the results of the single-epoch model, and the filled symbols show the results of the multiple-epoch model. The dotted line indicates the predictable variance for the wideband condition.

As shown in each panel, model predictions based on the single-epoch and multiple-epoch methods do not differ significantly for the ILD and ITD cues in any condition, except S10 in Isabelle and Colburn's study (1991). For the combination of ILD and ITD cues based on the covariance matrix, multiple-epoch predictions were slightly, though not significantly, better than single-epoch predictions for some listeners. In addition, predictions based on the combination of ILD and ITD cues were also slightly better than predictions based on single ILD and ITD cues for some listeners in response to Evilsizer *et al.*'s stimuli, but not for listeners in Isabelle and Colburn's study. For the SIED cue, single-epoch and multiple-epoch model predictions were not significantly different for most listeners, though predictions using the multiple-epoch model were slightly better than predictions using the single-epoch model for most listeners.

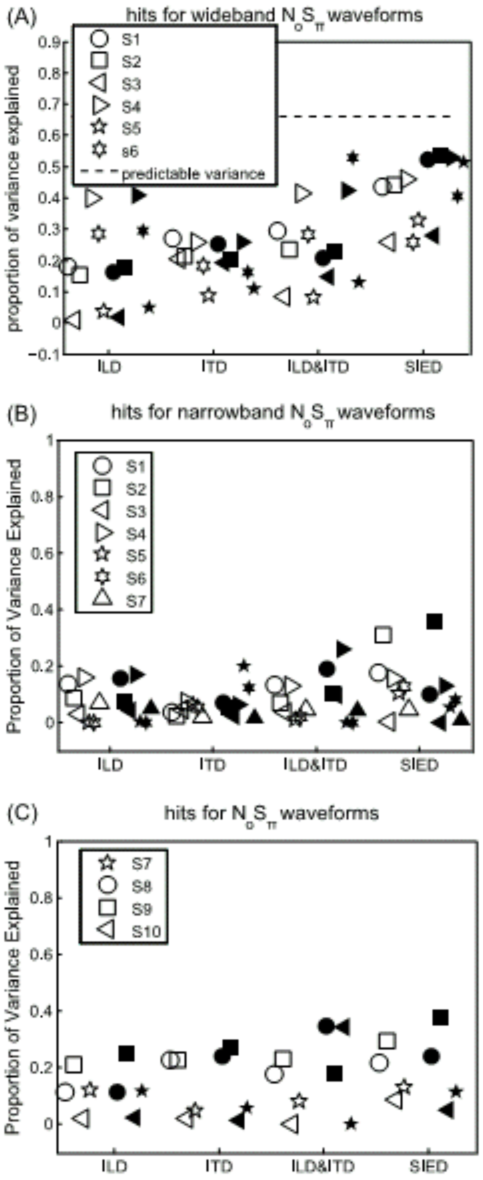

Figure 3.5: The proportion of variance explained by several interaural difference cues (ILD, ITD, combination of ILD and ITD, and SIED) predictions of hit rates for the individual listeners for waveforms of Evilsizer et al. (2002) study (A and B) and waveforms of Isabelle and Colburn (1991) study (C). The epoch duration was 75 ms for the multiple-epoch models (filled symbols). Different listeners were represented by different symbols.

Model predictions of the ILD, ITD, combination of ILD and ITD, and SIED cues for the average listener in the wideband condition are shown in Fig. 3.6. Model predictions using the ILD, ITD, and combination of ILD and ITD cues were similar. The prediction based on the SIED cue was significantly better than the prediction using the other three cues and approached the predictable variance. Note that no average listener was used in the narrowband condition, because listeners' detection patterns were not significantly correlated in general.

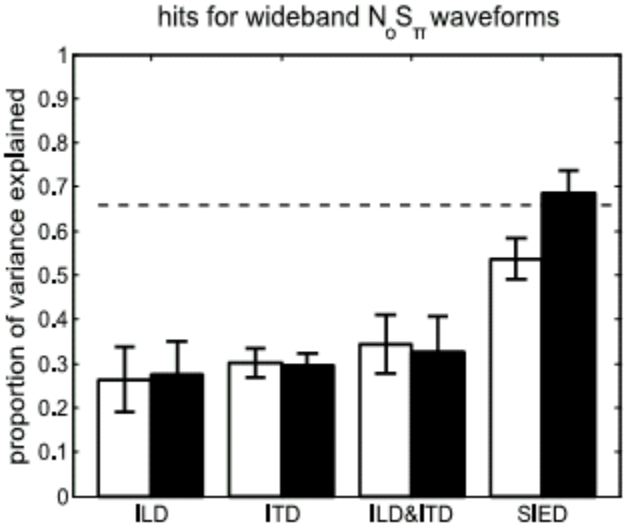

Figure 3.6: The proportion of variance explained by several interaural difference cues (ILD, ITD, combination of ILD and ITD, and SIED) predictions of hit rates for the average listeners for the waveforms of Evilsizer et al.'s study. The white and black bars show the model predictions obtained with single-epoch and multiple-epoch schemes.

Joris *et al.* (2006) suggested that cochlear disparity is potentially important in determining the best delays observed in binaural ITD-sensitive neurons. Additional tests with the SIED cue were carried out using gammatone filters with mismatched center frequencies at the two ears. Predictions of listeners' detection patterns with pairs of filters having different center-frequencies for the two ears (x-axis: left ear, y-axis: right ear) are shown in Fig. 3.7. The grayscale values indicate the predicted variance in the listener's detection patterns.

For the wideband stimuli (Fig. 3.7A), only predictions from the average listener are shown. Trends in the predictions across different frequency channel combinations were similar across individual listeners in the wideband condition. The highest correlation was obtained from models with matched center frequencies at 500 Hz (bottom left corner). Listeners might also use the SIED from frequency channels away from the tone frequency, for example the region of frequency combinations centered on 440 Hz and 550 Hz provides predictions that were significantly correlated to the average listener's detection pattern.

In contrast to the wideband case, for the narrowband stimuli (Figs. 3.7B-F), the center-frequency combinations that provided the best predictions of detection patterns differed qualitatively across listeners. Results from five individual listeners are shown, three from the Evilsizer *et al.* (2002) study and two from the Isabelle and Colburn (1991) study. The across-subject differences in Figs. 3.7B-F may explain the low correlations of detection patterns between pairs of listeners. These results suggest that listeners might use different strategies, including different frequency channels or different combinations of frequency channels, for detecting tones in narrowband noise.

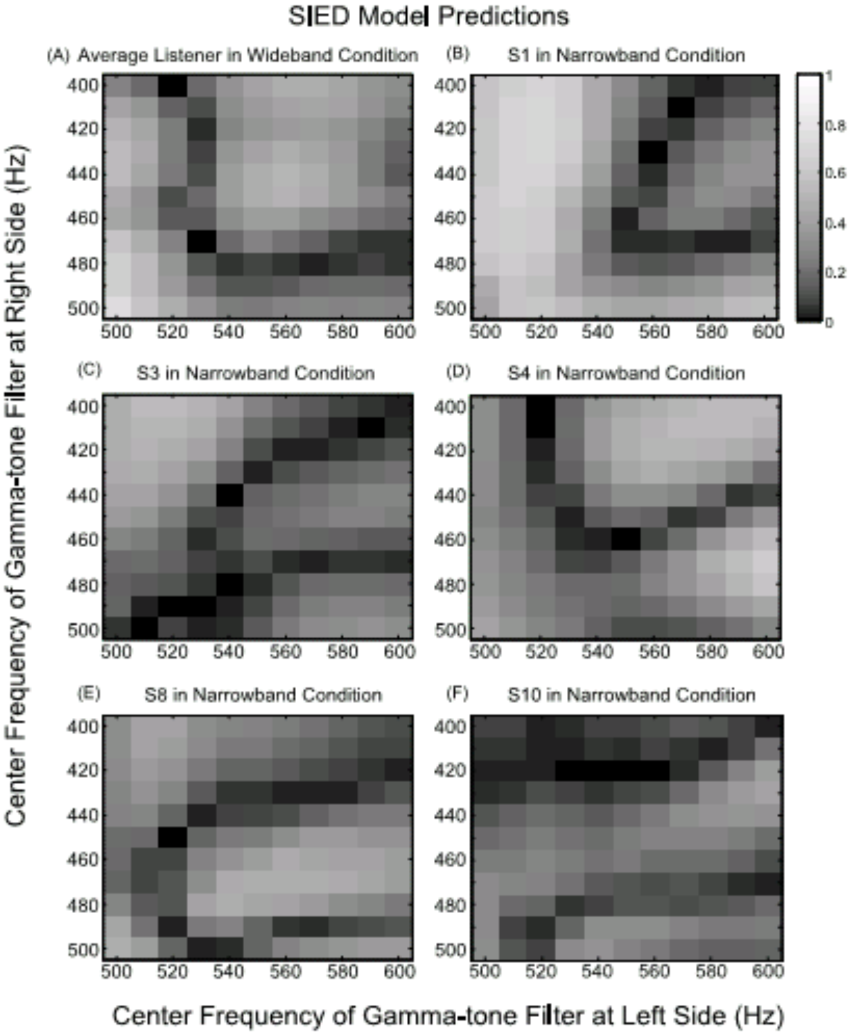

Figure 3.7: Predictions of listeners' detection patterns using mis-matched center-frequency at two ears (x-axis: left ear, y-axis: right ear) for (A) average listener in wideband condition, (B-D) several individual listeners (S1, S3, and S4) in the narrowband condition from Evilsizer *et al.* (2002) and (E-F) several individual listeners (S8 and S10) from Isabelle and Colburn (1991) studies.

3.5.3 Investigation of the SIED Cue using Binaurally Modulated Reproducible Noises

Given the success of the SIED cue in predicting listeners' detection patterns, especially in the wideband condition, it is interesting to investigate how the SIED cue is related to the two classic dichotic cues: ILD and ITD. Van der Heijden and Joris (2010) proposed a method that used binaurally modulated stimuli to degrade ILD, ITD, or both, in order to determine the relative contributions of ILD and ITD cues in a binaural detection test. In the current study, binaural modulation was applied to the reproducible noise stimuli from both the Evilsizer *et al.* (2002) and Isabelle and Colburn (1991) studies to test whether ILD, ITD, or both were related to the SIED cue. Different combinations of amplitude modulation (AM) and quasi-frequency modulation (QFM) were applied to the reproducible noises to introduce new ILDs, ITDs, or both. Then the effects of these manipulations on the SIED DV were examined to determine the contributions of each cue to the SIED.

Figure 3.8 illustrates four different types of binaural modulations, showing the case of modulating a single tone, for simplicity. In each panel, a vector diagram represents the binaural modulations applied to the stimuli at the left and right ear: the solid gray vertical arrows show the carrier (*fc*); the solid black vertical lines indicate the AM component, which is parallel to the carrier; the solid black horizontal lines indicate the QFM component, which is perpendicular to the carrier; and the solid black arrows show the resulting modulated signal. The modulation depth (*m*) is represented by the length of the AM and QFM components. Because the modulation depths of the AM and QFM are equal, the two components have the same length, thus the sum of the two components (solid gray line) always forms an angle of $\pi/4$ radians with respect to the carrier.

For diotic modulation (Fig. 3.8A), identical modulations are applied to the left and right stimuli and no magnitude or phase differences between θ_L and θ_R exist, thus no new ILD or ITD cues are introduced. For mixed modulation (Fig. 3.8B), there is a phase difference of π radians between θ_L and θ_R; both magnitude and phase differences are observed between the solid black arrows for the two ears, thus new ILD and ITD cues are introduced by mixed modulation. For binaural QFM (Fig. 3.8C), there is a phase difference of $3\pi/2$ radians between θ_L and θ_R; only phase differs between the solid black arrows for the two ears, thus a new ITD cue is introduced. For binaural AM (Fig. 3.8D), there is a phase difference of $\pi/2$ radians between θ_L and θ_R; the solid black arrows for the two ears differ primarily in terms of magnitude, with a small difference in phase between φ_L and φ_R, thus a new ILD cue with a small ITD cue is introduced.

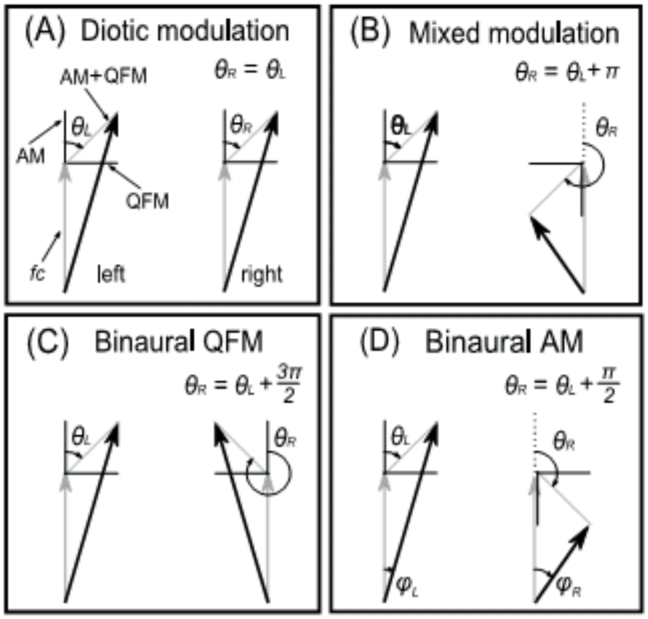

Figure 3.8: Four different binaural modulations used to separate ILD and ITD information: (A) diotic modulation; (B) mixed modulation; (C) binaural QFM; (D) binaural AM (after van der Heijden and Joris, 2010).

In order to apply binaural modulation to reproducible noises, the carrier was the dichotic reproducible waveform (both narrowband and wideband waveforms from the Evilsizer *et al.* (2002) study and narrowband waveforms from the Isabelle and Colburn (1991) study). The modulation frequency, f_m, was 20 Hz (as in van der Heijden and Joris, 2010). Different binaural modulations were applied by varying the phase difference of the combination of AM and QFM at the two ears (θ_L, θ_R), as shown in Fig. 3.8. Given that the AM and QFM components differed by π/2 radians, the complex analytic waveform $Z_L(t)$ or $Z_R(t)$ obtained from the dichotic waveform was used to illustrate the mathematical implementation of binaural modulation (Fig. 3.9). After multiplying $Z_L(t)$ or $Z_R(t)$ with modulators for the two ears, the modulated waveforms were recovered by taking the real part of the complex signal. The effects on the SIED, ILD, and ITD cues after applying the binaural modulation to the reproducible noises are shown for a range of modulation depths, *m* (see Figs. 3.10 and 3.11). The SIED DV was computed as shown in Fig. 3.3, using the binaurally modulated waveforms as inputs.

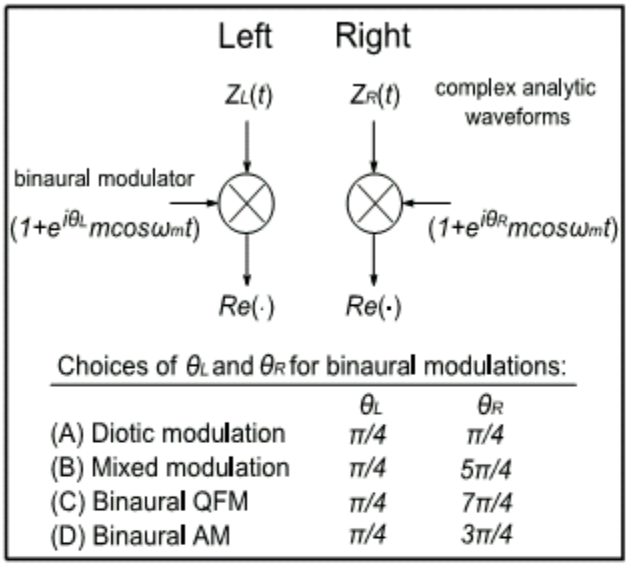
Left
Right
$Z_L(t)$
$Z_R(t)$
complex analytic waveforms
binaural modulator
$(1+e^{i\theta_L} m\cos\omega_m t)$
$(1+e^{i\theta_R} m\cos\omega_m t)$
$Re(\cdot)$
$Re(\cdot)$
Choices of θ_L and θ_R for binaural modulations:
θ_L
θ_R
(A) Diotic modulation $\pi/4$ $\pi/4$
(B) Mixed modulation $\pi/4$ $5\pi/4$
(C) Binaural QFM $\pi/4$ $7\pi/4$
(D) Binaural AM $\pi/4$ $3\pi/4$

Figure 3.9: The mathematical implementation of the binaural modulation of the dichotic waveforms for the left and right ears, where $Z_L(t)$ or $Z_R(t)$ represents the analytic waveform of noise-alone or tone-plus-noise stimuli and $Re(\cdot)$ indicates taking the real part of the complex signal.

In order to verify that the newly introduced ILD and ITD information were separated by the binaural modulation, the RMS values of ILD and ITD cues were computed from the four binaurally modulated dichotic reproducible noise waveforms In Fig. 3.10A, there is no difference in ILD_{RMS} for the diotic and binaural QFM stimuli, or for the mixed and binaural AM stimuli at all modulation depths. In Fig. 3.10B, at small modulation depths

(m≤0.3), no difference in $\mathrm{ITD}_{\mathrm{RMS}}$ was observed for the diotic and binaural AM stimuli, or for the mixed and binaural QFM stimuli. However, when modulation depth increased, the diotic and binaural AM stimuli had different $\mathrm{ITD}_{\mathrm{RMS}}$, whereas $\mathrm{ITD}_{\mathrm{RMS}}$ for the mixed and binaural QFM stimuli remained similar. The reason for the mismatch between $\mathrm{ITD}_{\mathrm{RMS}}$ for the diotic and $\mathrm{ITD}_{\mathrm{RMS}}$ for the binaural AM stimuli at large modulation depths is illustrated in Fig. 3.8D: when the modulation depth increases, the amplitude of the AM and QFM grow, and small phase differences of the solid black arrows between the two ears (φ_L, φ_R) are introduced as a byproduct of the binaural AM. Note that $\mathrm{ITD}_{\mathrm{RMS}}$ and $\mathrm{ILD}_{\mathrm{RMS}}$ are all nonzero because of the binaural differences introduced by the original (un-modulated) dichotic waveforms at both ears. Figure 3.10A, B thus verified that the binaural modulation manipulated the ILD and ITD cues as intended, as least for m≤0.3.

The effects of the binaural modulations on the ILD and ITD cues were verified and interpreted as follows. If the SIED cues computed from the diotic modulation and binaural QFM stimuli were identical, then the ILD cue must dominate the SIED cue, because ILDs are the same for these two types of modulation, but ITDs differ. The similarity of ILD for these conditions is verified by the overlap of the cross and square symbols in Fig. 3.10A. In contrast, this manipulation affects the ITDs, as indicated by the separation of the cross and square symbols in Fig. 3.10B.

Similarly, if the SIED cues obtained from the diotic modulation and binaural AM stimuli were identical, then the ITD cue must dominate the SIED cue, because the ITDs are similar for these two types of modulation, indicated by the overlap of the cross and circle symbols at small modulation depths in Fig. 3.10B. In contrast, new ILDs are introduced by the binaural AM manipulation, as indicated by the separation of the cross and circle symbols in Fig. 3.10A. If neither condition were satisfied, then the SIED would be related to both ILD and ITD.

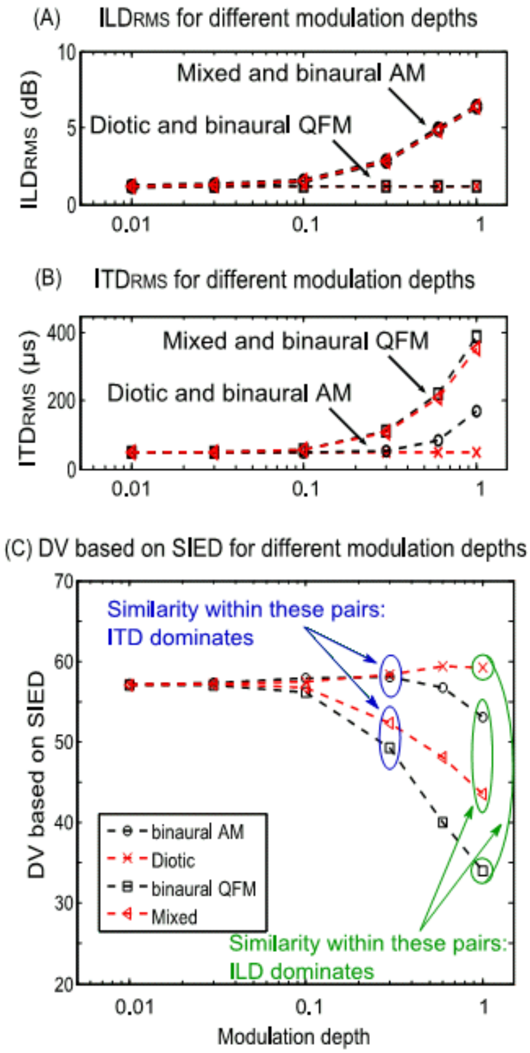

Figure 3.10: ILD_{RMS}, ITD_{RMS}, and DV based on the SIED for binaurally modulated wideband and narrowband stimuli from Evilsizer *et al.* (2002). The x-axis shows the modulation depth of the binaural modulator. Four different symbols are used to represent the four kinds of modulations: black circles for binaural AM, red crosses for diotic, black squares for binaural QFM, and red triangles for mixed modulation. Relations between SIED and ILD, ITD cues are illustrated: if ITD dominates the SIED cue, then the pairs of symbols connected or circled by blue lines should overlap; if ILD dominates the SIED cue, then the pairs of symbols connected or circled by the green lines should overlap.

The effects of the ILD and ITD manipulations on the SIED DV can now be analyzed based on the results shown in Fig. 3.10C, which illustrates the SIED DV for binaurally modulated wideband reproducible noise waveforms. If the SIED DVs were identical for the mixed and binaural AM stimuli, and for the diotic and binaural QFM stimuli (green circled groups), respectively, then the SIED cue would be fully determined by the ILD cue (see Fig. 3.10A). Similarly, if the SIED DVs were the same for the diotic and binaural AM stimuli, and for the mixed and binaural QFM stimuli (blue circled groups), respectively, then ITD would be the dominant cue (see Fig. 3.10B, for $m \leq 0.3$). Also, it is possible that neither ILD nor ITD cue alone completely explains the SIED DV. In that case, both ILD and ITD cues would be related to the SIED cue.

The results of the binaural modulation test of the wideband SIED cue are as follows. At small modulation depths ($m \leq 0.1$), DVs from all four sets of stimuli are similar (Fig. 3.10C), as expected from Figs. 10A and B. When modulation depth increases, DVs from the binaural AM and diotic stimuli, and from the mixed and binaural QFM stimuli, diverge. Thus, neither ILD nor ITD completely dominates the DV associated with the SIED cue. Comparing the trends in the SIED DVs to ILD_{RMS} and to ITD_{RMS}, it is clear that at small modulation depths ($m \leq 0.3$), ITD dominates the SIED cue because the DVs for the diotic and binaural AM stimuli overlap in both Figs. 3.10B and C. However, when modulation depth increases further ($m>0.3$), ILD contributes in addition to ITD, because DVs from both the diotic and binaural AM stimuli, and from the mixed and binaural QFM stimuli, no longer overlap. Thus, the results in Fig. 3.10 suggest that the SIED cue is dominated by ITD, with some contribution from ILD at high binaural modulation depths for the wideband stimuli in the Evilsizer *et al.* (2002) study.

The results of the binaural modulation test of the narrowband SIED cues are shown in Fig. 3.11A for the Evilsizer *et al.* (2002) stimuli and in Fig. 3.11B for the Isabelle and Colburn (1991) stimuli. Figures of $\mathrm{ILD}_{\mathrm{RMS}}$ and $\mathrm{ITD}_{\mathrm{RMS}}$ for these two sets of stimuli are not shown, as these results are the same as in Figs. 3.10A and B. In Figs. 3.11A and B, the SIED DVs from the four sets of binaurally modulated narrowband stimuli start to diverge at small modulation depths (m<0.1), unlike the results seen in Fig. 3.10C for the wideband stimuli. For the narrowband SIED cues from the Evilsizer *et al.* (2002) study, the trends are similar to the trends in Fig. 3.10C at large modulation depths: the SIED DVs fall into two pairs: DVs from the binaural AM stimuli and the diotic stimuli are one pair, and DVs from the binaural QFM stimuli and the mixed stimuli are another pair (Fig. 3.11A). These results indicate that the SIED cue is dominated by ITD for this set of stimuli. However, for SIED cues from the Isabelle and Colburn (1991) study, the trends are different from the trends in Fig. 3.10C. For these stimuli, DVs from all four sets of binaurally modulated stimuli separate at large modulation depths. The interpretation of the relationship between the SIED and ILD and ITD cues, and the different results of the SIED cue observed in Figs. 3.10C, 3.11A and B will be discussed below.

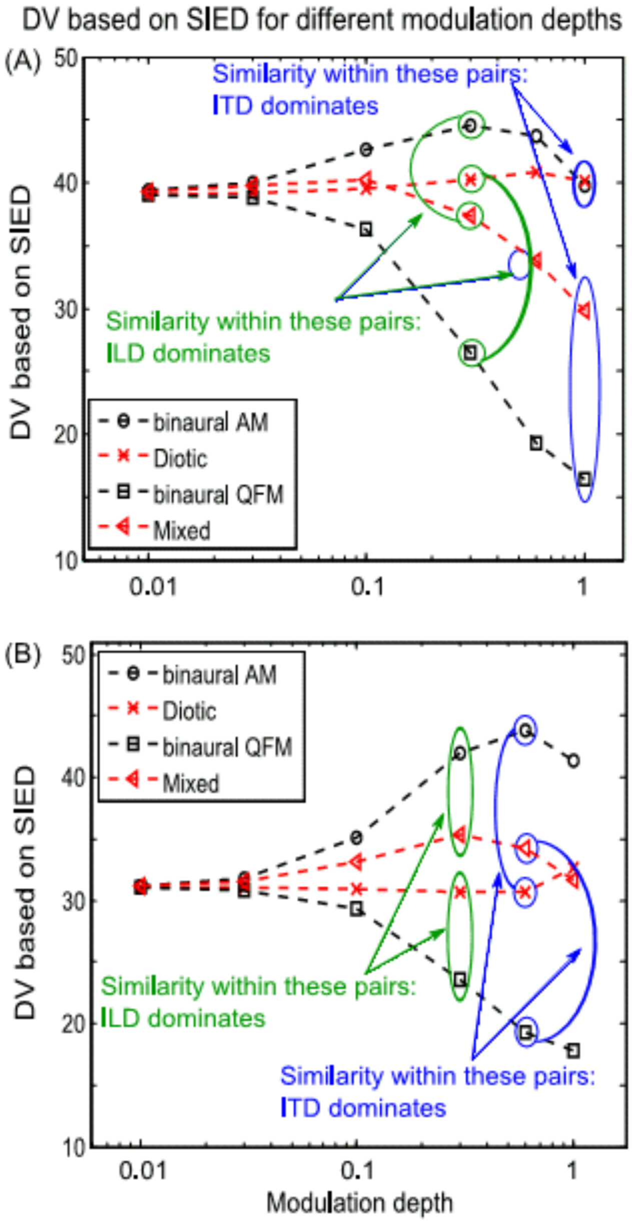

Figure 3.11: The SIED DVs for binaurally modulated (A) narrowband stimuli from Evilsizer *et al.* (2002) and (B) narrowband stimuli from Isabelle and Colburn (1991). The axes and symbols are the same as in Fig. 3.10C.

Although it is difficult to show that the SIED cue is based on a specific nonlinear combination of ILD and ITD cues, these results indicate that a linear combination of these two cues would not yield the SIED cue. As mentioned above, for the stimuli from the Evilsizer *et al.* (2002) study, the SIED DV is mainly determined by ITD at small modulation depths, because the differences of ILD_{RMS} were similar between the diotic and binaural AM stimuli, and between the mixed and binaural QFM stimuli. If the SIED DV were determined by a linear combination of ILD_{RMS} and ITD_{RMS}, then similar changes of the SIED cue would be observed between the diotic and binaural AM stimuli, and between the mixed and binaural QFM stimuli, at large modulation depths. However, at large modulation depths, smaller differences in the SIED DV were observed between the diotic and binaural AM stimuli, as compared to the mixed and binaural QFM stimuli (Fig. 3.10C, Fig. 3.11A). Thus, the SIED cue is related to a nonlinear combination of ILD and ITD, although other unidentified properties of the stimuli might also be related to the SIED cue. For the stimuli from the Isabelle and Colburn (1991) study, it is difficult to identify whether ILD or ITD dominates the SIED cue. As mentioned above, a difference between the two narrowband studies is that both overall noise level and tone levels at listeners' thresholds are higher for the Isabelle and Colburn (1991) stimuli (Table 3.2). This level difference would interact with the binaural modulations. Nevertheless, for both narrowband and wideband stimuli, the SIED cue is a nonlinear combination of ILD and ITD cues.

3.6 Discussion

People with hearing loss find it difficult to discriminate sound sources or communicate in noisy backgrounds (Henry and Heinz, 2012), even when using hearing aids. Thus, it is useful to understand how those with normal hearing detect signals in noise, in order to help design more effective techniques for hearing-aid devices. Understanding tone-in-noise detection is a first step to finding cues that are important for the above goal.

In this study, predictions of hit rates across a set of reproducible noises were computed based on several binaural cues. Comparisons were made between predictions based on ILD, ITD, a linear combination of ILD and ITD, and an envelope-based SIED cue. The combined ILD and ITD model took into account the covariance between these two cues. For listeners tested with the Evilsizer *et al.* (2002) wideband stimuli, the combined ILD and ITD model and the SIED model both yielded significantly better predictions than previous models (Davidson *et al.*, 2009a). In addition, wideband predictions based on the SIED cue approached the predictable variance.

For the narrowband stimuli in both Evilsizer *et al.* (2002) and Isabelle and Colburn (1991), predictions based on the combined ILD and ITD model and the SIED model were not significantly better than the previous models of Isabelle (1995) and Davidson *et al.* (2009a). Further analysis of the correlations between ILD, ITD, and SIED cues for the narrowband and wideband conditions is shown in Table 3.3. All three cues were significantly correlated in these two conditions; however, listeners' detection patterns in these two conditions were not significantly correlated. Thus the observed difference of model predictions in these two conditions might be related to the different strategies used by the listeners across bandwidth conditions.

Table 3.3: Correlation of DVs for narrowband and wideband stimuli in Evilsizer *et al.* (2002). Note that S7 was only tested with the narrowband stimuli and is not listed here.

	S1	S2	S3	S4	S5	S6
ILD	0.43	0.52	048	045	0.62	0.56
ITD	0.39	0.47	0.62	0.42	0.56	0.53
SIED	0.77	0.70	0.67	0.76	0.52	0.68

Similar to previous studies (Isabelle, 1995; Davidson *et al.,* 2009a), model predictions based on a single ILD or ITD cue did not explain a significant amount of the variance in listeners' narrowband detection patterns (Figs. 3.5 and 3.6). Similar to Goupell and Hartmann's method (2007), analysis of the ITD cue was also computed by removing the large instantaneous phase changes when the envelope in either ear was small (Goupell and Hartmann, 2007), but significant model predictions were not observed for this modification of the ITD cue. Moreover, single-cue multiple-epoch methods did not yield significantly better predictions than single-epoch models for most listeners. However, model predictions that combined ILD and ITD across time epochs and took into account their covariance matrix yielded significantly better predictions of hit rates than those using single cue and single epoch for some listeners. Thus, these listeners may use a binaural integration strategy that combines ILD and ITD cues.

The dynamic variation of ILD and ITD cues are interrelated with the changes in the envelopes at both ears; thus the possibility that listeners use envelope cues was examined in this study. The success and robustness of the envelope-slope (ES) cue in predicting diotic detection patterns (Richards, 1992; Zhang, 2004; Davidson *et al.*, 2009a; Mao *et* *al.*, 2013) motivated the examination of a binaural envelope cue for the dichotic condition. The proposed SIED cue yielded better predictions of listeners' detection patterns for the wideband condition than any previous method. Further investigation showed that the SIED cue was related to both ILD and ITD in a nonlinear manner. The SIED cue is a simple description of a nonlinear combination of ITD and ILD cues. In addition, it was shown that the SIED is mainly determined by the ITD cue, though ILD contributes for stimuli with larger amplitude fluctuations. For most complex stimuli with time-varying amplitudes, the modulation depth also changes over time, thus both ILD and ITD cues will contribute to the SIED cue at different points within the stimulus. The dominance of ITD over ILD in predicting binaural detection results is consistent with the studies by van der Heijden and Joris (2010) and Webster (1951).

Further analysis was carried out to examine whether listeners rely on the slope or the energy of the envelope fluctuations to detect tones in noise. Computing the SIED cue using only the sharp slopes (e.g., max values in the slope) did not yield significant correlations to listeners' detection patterns. Instead of using the SIED cue, the energy of envelope fluctuations was also computed based on the energy of non-DC (i.e., non-zero frequency) components in the modulation-frequency domain. The proportion of variance in listeners' detection patterns that was explained based on the envelope-energy model was significantly less than that explained by the SIED cue for all listeners in both the narrowband and wideband conditions. Thus, it was confirmed that the slope rather than the envelope energy yields a DV that is more consistent with observed detection patterns.

The binaural envelope cue has not previously been used to explain binaural detection and discrimination tests, and it is interesting to consider the ability of the SIED to explain the results from other dichotic studies. For instance, some listeners have a higher threshold for dichotic detection of a 500-Hz pure tone in low-noise noise compared with Gaussian noise (Hall *et al.*, 1998; Eddins and Barber, 1998; Goupell, 2012). Low-noise noise (LNN) has less fluctuation in its envelope compared with Gaussian noise, because LNN is generated by manipulating the phases of each frequency component to reduce envelope fluctuations, whereas Gaussian noise has random phases for each frequency component. Goupell (2012) could predict a significant amount of several listeners' detection variance for just-noticeable-differences in interaural correlation for LNN stimuli using two models: a normalized cross-correlation model with envelope compression (Bernstein *et al.*, 1999) and the "independent-center" model (Goupell and Hartmann, 2007). Although fitting is involved in these models, his results show that envelope fluctuation is a possible cue to explain some listeners' performance. In addition, Hall *et al.* (1998) suggested that these listeners could benefit from listening in the "dips" for the Gaussian noises that have larger fluctuations.

Another possible explanation for the difference in thresholds for LNN and Gaussian noise is related to differences in the size of the SIED cue, as a result of the increased envelope fluctuations for Gaussian noise. Inspection of the SIED cues from a set of random Gaussian noises and LNNs showed that, although mean DVs were similar for these two types of noises, at signal-to-noise ratios (SNR) close to listeners' thresholds, the SIED cues from Gaussian noises were more variable across the maskers than for LNN (by approximately a factor of two), as expected.

Henning (1973) tested two listeners for frequency-modulation and amplitude-modulation discrimination under both diotic and dichotic conditions. His results show that at low SNR, listeners have significantly lower discrimination thresholds under dichotic conditions than under diotic conditions; at high SNR, listeners have similar discrimination thresholds for the two conditions. Henning further demonstrated that results from the amplitude-modulation discrimination task could be predicted using Durlach's EC model (1963) and the Webster-Jeffress models (1951). The SIED cue provides an alternative explanation for results from amplitude-modulation discrimination because envelope cues are available for the modulated stimuli. At low SNRs, the SIED cue was available for the dichotic condition, but not for the diotic condition; listeners' thresholds would be therefore lower for the dichotic than the diotic condition if they used the SIED cue. However, at high SNRs, the SIED cue would decrease for the dichotic condition because the tones would dominate the envelope; tone signals have flatter envelopes than noises, suggesting that the SIED would be less effective at high SNRs. Simulation results from amplitude- and frequency-modulated stimuli showed that the variance of the SIED cues decreased at low SNRs compared to SIED cues at high SNRs (by approximately one-half).

Because all three cues studied here (ILD, ITD, and SIED) depend on the interaural differences introduced by the addition of out-of-phase tones to in-phase noise, none of these cues exist for the noise-alone waveforms presented during the dichotic detection task. In order to predict false-alarm rates, potential sources of binaural differences in response to noise-alone waveforms must be considered. One way to achieve this goal is to apply physiological models with realistic statistical properties, such as responses from model auditory-nerve fibers and central neurons, or to introduce multiplicative noises (Bernstein and Trahiotis, 2008; Ewert and Dau, 2004). In addition, convergence of model auditory-nerve fibers with mismatched center frequencies could also provide binaural differences in response to noise-alone waveforms (Joris *et al.*, 2006). The analysis of narrowband detection results presented here suggests that an exploration of models that include combinations of different frequency channels deserves further study. Future studies will focus on physiological models, in which predictions of detection patterns for both hit and false-alarm rates can be computed for the narrowband and wideband detection conditions.

ACKNOWLEDGEMENTS

This work was supported by grant NIH-NIDCD R01-DC010813. We would like to thank Kristina Abrams, Kelly-Jo Koch, Dr. Tianhao Li, Douglas Schwarz, and the students in the lab for their helpful suggestions on preparing the manuscript. We would also like to thank Dr. Scott Isabelle and Dr. Steven Colburn for providing their stimuli and data.

[1] Listeners' detection thresholds in Evilsizer et al. (2002) study were described as $\frac{E_s}{N_o}$,

which was computed as

$$\frac{E_s}{N_o} = Overall\,Tone\,Level\,(dB\,SPL) - Noise\,Spectrum\,Level\,(dB\,SPL) + 10\log_{10}(Duration).$$

The noise level was computed as

$$Overall\ \ Noise\ \ Level = Noise\ \ Spectrum\ \ Level\,(dB\,SPL) + 10\log_{10}(Bandwidth).$$

As a result, signal-to-noise ratio (SNR) was calculated as

$$\begin{aligned} SNR &= Overall\ \ Tone\ \ Level - Overall\ \ Noise\ \ Level \\ &= \frac{E_s}{N_o} - 10\log_{10}(Duration) - 10\log_{10}(Bandwidth). \end{aligned}$$

Bibliography

Bernstein, L. R., and Trahiotis, C., (2008). "Binaural signal detection, overall masking level, and masker interaural correlation: revisiting the internal noise hypothesis," J. Acoust. Soc. Am. 124, 3850-3860.

Bernstein, L. R., van de Par, S., and Trahiotis, C., (1999). "The normalized interaural correlation: accounting for NoSpi thresholds obtained with Gaussian and "low-noise" masking noise," J. Acoust. Soc. Am. 106, 870-876.

Blodgett, H. C., Jeffress, L. A., and Taylor, R. W., (1958). "Relation of masked threshold to signal-duration for interaural phase combination," Am. J. Psychol. 71, 283-290.

Blodgett, H. C., Jeffress, L. A., and Whitworth, R. H., (1962). "Effect of noise at one ear on the masked threshold for tone at the other," J. Acoust. Soc. Am. 34, 979-981.

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

Colburn, H. S., Isabelle, S. K., and Tollin, D. J., (1997). "Modeling binaural detection performance for individual masker waveforms," in *Binaural and Spatial Hearing in real and virtual environments*, edited by R.H. Gilkey and T. Anderson (Erlbaum, Englewood Cliffs, NJ), Chap.25, pp. 533-556.

Culling, J. F., and Colburn, H. S., (2000). "Binaural sluggishness in the perception of tone sequences and speech in noise," J. Acoust. Soc. Am. 107, 517-527.

Davidson, S. A., Gilkey, R. H., Colburn, H. S., and Carney, L. H., (2006). “Binaural detection with narrowband and wideband reproducible noise maskers. III. Monaural and diotic detection and model results,” J. Acoust. Soc. Am. 119, 2258-2275.

Davidson, S. A., Gilkey, R. H., Colburn, H. S., and Carney, L. H., (2009a). “An evaluation of models for diotic and dichotic detection in reproducible noises,” J. Acoust. Soc. Am. 126, 1906-1925.

Davidson, S. A., Gilkey, R. H., Colburn, H. S., and Carney, L. H., (2009b). “Diotic and dichotic detection with reproducible chimeric stimuli,” J. Acoust. Soc. Am. 126, 1889-1905.

De Boer, E., and de Jongh, H. R., (1978). “On cochlear encoding: potentialities and limitations of the reverse correlation technique,” J. Acoust. Soc. Am. 63, 115-135.

Dolan, T. R., and Robinson, D. E., (1967). “Explanation of masking-level difference that result from interaural intensive disparities of noise,” J. Acoust. Soc. Am. 42, 977-981.

Durlach, N. I., (1963). “Equalization and cancellation theory of binaural masking-level differences,” J. Acoust. Soc. Am. 35, 1206-1218.

Eddins, D. A., and Barber, L. E., (1998). “The influence of stimulus envelope and fine structure on the binaural masking level difference,” J. Acoust. Soc. Am. 103, 2578-2589.

Evilsizer, M. E., Gilkey, R. H., Mason, C. R., Colburn, H. S., and Carney, L. H., (2002). “Binaural detection with narrowband and wideband reproducible maskers: I. Results for human,” J. Acoust. Soc. Am. 111, 333-345.

Ewert, S. D., and Dau, T., (2004). “External and internal limitations in amplitude-modulation processing,” J. Acoust. Soc. Am. 116, 478-490.

Green, D. M., (1964). “Consistency of auditory detection judgments,” Psychol. Rev. 71, 392-407.

Gilkey, R. H., Robinson, D. E., and Hanna, T. E., (1985). “Effects of masker waveform and signal-to-masker phase relation on diotic and dichotic masking by reproducible noise,” J. Acoust. Soc. Am. 78, 1207-1219.

Gilkey, R. H., and Robinson, D. E., (1986). “Models of auditory masking: A molecular psychophysical approach,” J. Acoust. Soc. Am. 79, 1499-1510.

Goupell, M. J., (2012). “The role of envelope statistics in detecting changes in interaural correlation,” J. Acoust. Soc. Am. 132, 1561-1572.

Goupell, M. J., and Hartmann, W. M., (2007). “Interaural fluctuations and detection of interaural incoherence. III. Narrowband experiments and binaural models,” J. Acoust. Soc. Am. 122, 1029-1045.

Grantham, D. W., and Wightman, F. L., (1979). “Detectability of stimuli pulsed tone in the presence of a masker with time-varying interaural correlation,” J. Acoust. Soc. Am. 65, 1509-1517.

Hafter, E. R., (1971). “Quantitative evaluation of a lateralization model of masking-level differences,” J. Acoust. Soc. Am. 50, 1116-1122.

Hall, J. W. 3rd, Grose, J. H., and Hartmann, W. M., (1998). “The masking-level difference in low-noise noise,” J. Acoust. Soc. Am. 103, 2573-2577.

Henning, G. B., (1973). “Effect of interaural phase on frequency and amplitude discrimination,” J. Acoust. Soc. Am. 54, 1160-1178.

Henry, K. S., and Heinz, M. G., (2012). “Diminished temporal coding with sensorineural hearing loss emerges in background noise,” Nat. Neurosci. 15, 1362-1364.

Isabelle, S. K., (1995). “Binaural detection performance using reproducible stimuli,” Ph.D. thesis, Boston University, Boston, MA.

Isabelle, S. K., and Colburn, H. S., (1987). “Effects of target phase in narrowband frozen noise detection data,” J. Acoust. Soc. Am. 82, S109-S109.

Isabelle, S. K., and Colburn, H. S., (1991). “Detection of tones in reproducible narrow-band noise,” J. Acoust. Soc. Am. 89, 352-359.

Isabelle, S. K., and Colburn, H. S., (2004). “Binaural detection of tones masked by reproducible noise: Experiment and models,” Report BU-HRC 04–01.

Johannesma, P. I. M., van Gisbergen, J. A. M., and Grashuis, J. L., (1971). “Forward and backward analysis of temporal relations between sensory stimulus and neural response,” Internal Report (Lab. of Medical Physics, University of Nijmegen, the Netherlands).

Joris, P. X., Van de Sande, B., Louage, D. H., and van der Heijden, M., (2006). “Binaural and cochlear disparities,” Proc. Natl. Acad. Sci. USA. 103, 12917-12922.

Kolarik, A. J., and Culling, J. F., (2009). “Measurement of the binaural temporal window using a lateralization task,” Hear Res. 248, 60-68.

Kollmeier, B., and Gilkey, R. H., (1990). “Binaural forward and backward masking: evidence for sluggishness in binaural detection,” J. Acoust. Soc. Am. 87, 1709-1719.

Levitt, H., (1971). “Transformed up-down methods in psychoacoustics,” J. Acoust. Soc. Am. 49, 467-477.

Mao, J., Vosoughi, A., and Carney, L. H., (2013). “Predictions of diotic tone-in-noise detection based on a nonlinear optimal combination of energy, envelope, and fine-structure cues,” J. Acoust. Soc. Am. 134, 396-406.

Oruç, İ., Maloney, L. T., and Landy, M. S., (2003). “Weighted linear cue combination with possibly correlated error,” Vision Res. 43, 2451-2468.

Richards, V. M., (1992). “The delectability of a tone added to narrow bans of equal energy noise,” J. Acoust. Soc. Am. 91, 3424-3435.

Schönfelder V. H., and Wichmann, F. A., (2013). “Identification of stimulus cues in narrow-band tone-in-noise detection using sparse observer models,” J. Acoust. Soc. Am. 134, 447-463.

van der Heijden, M., and Joris, P. X., (2010). “Interaural correlation fails to account for detection in a classic binaural task: dynamic ITDs dominate N0Spi detection,” J. Assoc. Res. Otolaryngol. 11, 113-131.

Webster, F. A., (1951). “The influence of interaural phase on masked thresholds. I. The role of time-deviation,” J. Acoust. Soc. Am. 23, 452-462.

Zhang, X., (2004). “Cross-frequency coincidence detection in the processing of complex sounds,” Ph.D. thesis, Boston University, Boston, MA.

Zheng, L., Early, S. J., Mason, C. R., Idrobo, F., Harrison, J. M., and Carney, L. H., (2002). “Binaural detection with narrowband and wideband reproducible noise maskers: II. Results for rabbits,” J. Acoust. Soc. Am. 111, 346-356.

Zurek, P. M., (1991). "Probability distributions of interaural phase and level differences in binaural detection stimuli," J. Acoust. Soc. Am. 90, 1927-1932.

Chapter 4

Tone-in-Noise Detection using Envelope Cues: Comparison of Signal-Processing-based and Physiological Models

4.1 Abstract

Tone-in-noise detection tasks with reproducible noise maskers have been used to identify cues that listeners use to detect signals in noisy environments. Previous studies have shown that energy, envelope, and fine-structure cues are significantly correlated to listeners' performance for detection of a 500-Hz tone in noise. In this study, envelope cues were examined for both diotic and dichotic tone-in-noise detection using stimulus-based signal processing and physiological models. For stimulus-based envelope cues, a modified envelope-slope (ES) model with a band-pass filter was used for the diotic condition and the binaural slope of the interaural envelope difference (SIED) model was used for the dichotic condition. For physiological envelope cues, noise-alone and tone-plus-noise stimuli were passed through model auditory-nerve (AN) fibers, cochlear nucleus (CN), and inferior colliculus (IC) cells. The model IC cell was simulated as a modulation filter. The average rate of synapse output and response fluctuations from the model IC cell were examined. Previous studies have shown that a significant amount of the variance across reproducible noise maskers in listeners' detection results can be explained by stimulus-based envelope cues. In this study, it is shown that basic neural mechanisms based on physiological envelope cues predict a similar amount of the variance in listeners' performance across noise maskers.

4.2 Introduction

Speech identification in the presence of background noise is more difficult for listeners with hearing loss, even when using hearing aids, than for listeners with normal hearing. Envelope cues are important for detecting tones in reproducible noises (Davidson *et al.*, 2006; Davidson *et al.*, 2009; Mao *et al.*, 2013) and are robust to roving-level conditions (Richards, 1992). However, these studies have been based on signal-processing style models that extract the envelope from the stimulus directly, using the Hilbert transform or related techniques. In this study, envelope cues were analyzed for diotic and dichotic tone-in-noise detection using both stimulus-based and physiological models.

Tone-in-noise detection with reproducible noises has been used to identify cues that listeners use for this task (Evilsizer *et al.*, 2002; Davidson *et al.*, 2006; Davidson *et al.*, 2009; Isabelle, 1995). For diotic detection, energy within a critical band (Fletcher, 1940) can predict a significant amount of the variance in listeners' performance across different noise maskers, but the energy cue fails in roving-level conditions, in which the overall sound level varies in each trial (Kidd *et al.*, 1989). An envelope-based cue, ES, has been shown to be robust for the roving-level condition, and can predict a significant amount of the variance in listeners' detection performance (Richards, 1992; Zhang, 2004; Davidson *et al.*, 2009; Mao *et al.*, 2013). Models based on temporal fine-structure using the phase-opponency model (Carney *et al.*, 2002) can also predict a small but significant portion of the variance in listeners' performance across waveforms. A recent study (Mao *et al.,*

2013) shows that model predictions based on an optimal combination of energy and temporal cues approach the predictable variance in detection patterns (the common variance among different listeners' performance) for the diotic condition.

For dichotic detection, the interaural level and time cues (ILD, ITD), and the combinations of these two cues, have been used to predict listeners' performance (Isabelle, 1995; Davidson *et al.*, 2009). Although some of these dichotic cues can explain a significant portion of the variance in some listeners' performance, these predictions were substantially lower than the predictable variance. A binaural envelope cue, the slope of the interaural envelope difference (SIED), yields significantly better predictions than ILD and/or ITD cues (Chapter 3). Thus, among stimulus-based models, those using envelope cues successfully predict listeners' performance for both diotic and dichotic conditions.

In addition to models based on properties of the stimulus waveforms, several physiological models have been used to predict listeners' interaural time, frequency, and level discrimination thresholds based on model auditory-nerve (AN) responses (Colburn, 1973, 1977; Heinz *et al.*, 2001). Dau *et al.* (1996) and Breebaart *et al.* (2001) simulated signal processing in the auditory system by including band-pass filters, rectification, and adaptation for masked detection. Models of auditory processing at different levels along the pathway have also been proposed. Model AN fibers (Zilany and Bruce, 2006, 2007; Zilany *et al.*, 2009) simulate responses to noises, tones and complex stimuli. Nelson and

Carney (2004) introduced a same-frequency inhibition and excitation (SFIE) model for cochlear nucleus (CN) and inferior colliculus (IC) cells to explain envelope processing.

Because envelope cues are important for detection in noise, it is interesting to investigate whether and how envelope cues are processed along the auditory pathway, and whether physiological models can perform as well as the signal-processing-based models in predicting listeners' performance. In this study, it was hypothesized that similar amounts of the variance in listeners' performance could be predicted using stimulus-based and physiological models for envelope processing along the auditory pathway. For stimulus-based models, a modified ES model with a band-pass filter was used for the diotic condition, and the SIED model was used for the dichotic condition. For physiological models, envelope information was computed as the average rate of synapse output and response fluctuations from monaural and binaural model IC cells. Physiological model responses were analyzed at the level of the IC, as the IC is sensitive to envelope information (Joris *et al.*, 2004; Nelson and Carney, 2007). Given the success of ES and SIED models for predicting listeners' performance, physiological models were used to investigate whether similar cues could be extracted using basic neural mechanisms

4.3 Datasets

Listeners' detection performance for diotic and dichotic tone-in-noise detection was obtained from previous experiments (Evilsizer *et al.*, 2002; Davidson *et al.*, 2006; Chapter 3). Detection in the presence of reproducible noise maskers, a set of pre-generated random noises, was tested on each listener in these previous studies. In each trial, either a noise-alone or tone-plus-noise waveform was randomly chosen from the set of reproducible waveforms. Listeners responded “tone present” or “tone not present”, and their performance was described in terms of hit rate (proportion of correct response of “tone present” for tone-plus-noise waveforms), and false-alarm (FA) rate (proportion of responding “tone present” for noise-alone waveforms) for each reproducible masker waveform. The set of hit and FA rates across the ensemble of maskers are referred to as a detection pattern (Davidson *et al*., 2006). Figure 4.1 shows the detection pattern of the average listener (*i.e.*, averaged performance across individual listeners) for diotic narrowband waveforms.

Data from two different listening conditions were used in this study: diotic, in which identical in-phase noise-alone and tone-plus-noise waveforms were presented at the two ears, and dichotic, in which out-of-phase tones were added to the noise waveforms at the two ears. Both narrowband (452-552 Hz) and wideband (100-3000 Hz) waveforms were used for diotic and dichotic conditions. The narrowband maskers were created by extracting their 100-Hz bandwidth spectrum from the wideband maskers. The spectrum level of the noise waveform was 40 dB SPL (overall noise level was 60 dB SPL for narrowband waveforms and approximately 75 dB SPL for wideband waveforms). For the predictions here, the 500-Hz tone level was set to each listener’s threshold. Listeners’ detection thresholds for the dichotic condition were approximately 10 dB lower than those in the diotic condition. This decrease in threshold, the well-known binaural masking level difference (Moore, 2003), was due to the binaural differences introduced in the dichotic condition. In this study, data from a total of eight listeners for the diotic condition (S1-S4 from Evilsizer *et al.*, 2002 and S5-S8 from Davidson *et al.*, 2006), and six listeners for the dichotic condition (S1-S4 from Evilsizer *et al.*, 2002 , and S9-S10 from Chapter 3) were analyzed.

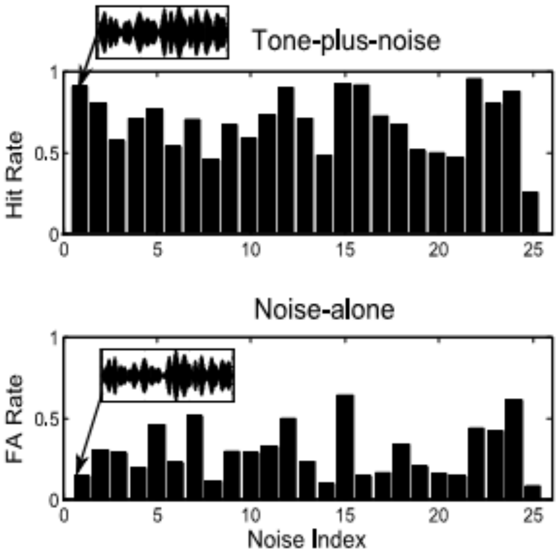

Figure 4.1: Detection pattern (hit and FA rates) of the average listener for diotic narrowband waveforms. The horizontal axis shows noise index; the insets show examples of tone-plus-noise (top) and noise-alone (bottom) waveforms. Note that listeners' responses vary across reproducible waveforms; responses were highly consistent within and across individual listeners for this stimulus condition.

4.4 Methods

In this study, given that envelope cues for tone-in-noise detection are processed in the auditory pathway, it was hypothesized that cues obtained with basic neural mechanisms of the responses from the model IC cells yield similar predictions as the stimulus-based envelope cues for predicting listeners' detection performance.

4.4.1 Diotic models for tone-in-noise detection

Two types of envelope-based cues were used in this study: the modified stimulus-based ES cue and the physiologically-based envelope cue from model IC responses.

4.4.1.1 Stimulus-based Model

The original ES model (Richards, 1992; Zhang, 2004; Davidson *et al.*, 2006) quantifies changes in envelope fluctuations. Because the addition of a tone to a narrowband noise waveform results in a decrease of the envelope fluctuations, a low value for the decision variable (DV, representing a certain feature of the waveform) indicates that the testing waveform is more likely to be a tone-plus-noise stimulus. By inspecting the frequency components of envelopes from tone-plus-noise and noise-alone stimuli, it was determined that the largest differences in envelope energy were within 50-150 Hz (Mao *et al.*, 2013). Thus, a sixth-order bandpass filter with center frequency of 120 Hz (Q=1) was used to extract the envelope frequency range of interest. Figure 4.2a shows the schematic diagram of the modified ES model. The Hilbert transform is used to compute the analytic signal from the output of a fourth-order gammatone filter (center frequency of 500 Hz). The envelope is obtained from the analytic signal and the DV of the model was calculated as the integral of the half-wave rectified slope of the envelope. The difference between this modified model and the original ES model (Richards, 1992; Zhang, 2004; Davidson *et al.*, 2006) is that a tenth-order lowpass filter (cut-off frequency at 250 Hz, aiming to exclude the high frequency fine-structure components) is replaced by the bandpass filter to extract envelope cues from the most informative frequency range. A previous study showed that predictions based on the modified ES model (Mao *et al.*, 2013) were more consistent with listeners' performance than those using the original ES model.

4.4.1.2 Physiological Model

In the physiological model, the stimulus is passed through a series of phenomenological models along the ascending auditory pathway (Fig. 4.2b). First, a human-version of the AN model (Zilany *et al.*, 2009; Ibrahim and Bruce, 2010; Zilany *et al.*, 2013) is used to obtain the AN synapse output. The input to the AN model is first processed by a middle ear filter, followed by a set of bandpass filter paths that provided inputs to the inner hair cell (IHC). The IHC response provides the input to the synapse model, which provides the final model AN response. This AN model has been shown to simulate responses to a range of different stimuli accurately, including pure tones, forward masked stimuli, and amplitude modulated (AM) stimuli (Zilany *et al.*, 2009). Next, model AN responses are used as inputs to a CN model (Nelson and Carney, 2004). Inhibitory and excitatory AN responses tuned to the same frequency are processed through lowpass filters, representing convolution with post-synaptic potential waveforms, and are then combined to provide the response of the CN model. Krishna and Semple (2000), and Nelson and Carney (2007) showed that approximately half of the IC cells have bandpass tuning to AM. The IC cell is simulated with a modulation filter to represent this tuning in the model. Specifically, the IC responses are modeled by a bandpass modulation filter, with a peak, or best modulation frequency (BMF), that receives its input from the synapse output of the CN model.

The SFIE-type IC model used by Nelson and Carney (2004) acts as a modulation filter, and a number of center frequencies can be achieved by carefully choosing time constants for the excitatory and inhibitory inputs. The *Q*-value of the SFIE-type filter is approximately 1.5, and preliminary results showed that better predictions were obtained with broader filters. In this study, a sixth-order bandpass filter with more flexible center frequencies and a *Q*-value of 1 was used (details see Appendix). The order of the filter used here was determined by the phase range obtained from physiological recordings of IC cells in awake rabbits (unpublished observations). Two basic neural mechanisms were used to obtain envelope cues from the physiological model: rate, which was computed as the averaged synapse output, and fluctuations, which was obtained from the integral of the half-wave rectified derivative of the model response.

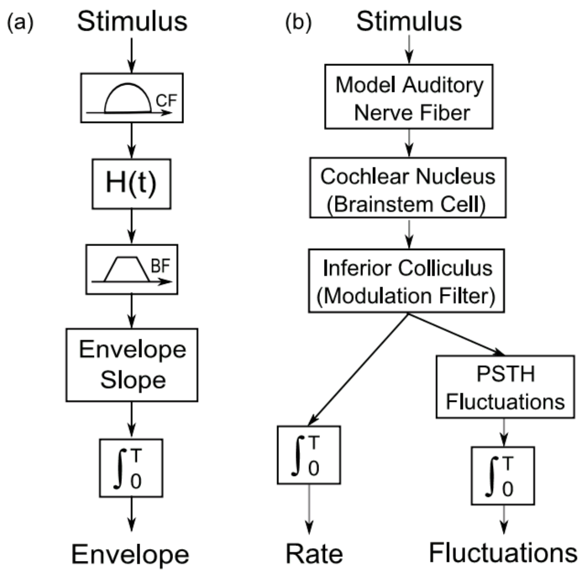

Figure 4.2: A schematic diagrams of the monaural envelope models. (a) Stimulus-based modified ES model: The envelope was obtained from the analytic waveform computed from the Hilbert transform of a fourth-order gammatone filtered waveform; a sixth-order bandpass filter with center frequency of 100 Hz was used to extract the envelope frequency range of interest. Cue values were computed as the integral of the half-wave rectified slope of the envelope fluctuations. (b) Physiological envelope model: the stimulus was passed through the AN and CN phenomenological models, and the IC modulation filter to obtain the synapse output. The neural envelope cue was computed based on the rate and fluctuation of the model synapse output.

4.4.2 Dichotic Models for Tone-in-noise Detection

Similar to the diotic condition, both stimulus-based and physiological envelope models were used in the dichotic condition. In this condition, listeners were tested with identical noise stimuli to the two ears in noise-alone trials, and thus there were no binaural difference cues on these trials. Therefore, only hit rates were predicted for the stimulus-based models in the dichotic condition. For the physiological cues, binaural differences were obtained by using cell inputs with mis-matched center frequencies, and FA rates could be predicted.

4.4.2.1 Stimulus-based Model

For the dichotic condition, binaural differences occur because of the addition of out-of-phase tones to in-phase noises at the two ears. The SIED model (Mao *et al.*, 2011) focuses on the binaural envelope difference cues. Figure 4.3a shows a schematic diagram of the SIED model, in which envelopes from the contralateral and ipsilateral sides are extracted from the analytic signal computed from a fourth-order gammatone filtered stimulus. The binaural envelope difference was calculated based on the difference between the computed monaural envelopes. Specifically, the SIED cue value was calculated as the time integral of the half-wave rectified slope of the envelope differences. It has been shown that the SIED cue represents a nonlinear combination of interaural time and level differences (Mao *et al.*, 2011).

4.4.2.2 Physiological Model

Computation of the dichotic physiological envelope cue is shown in Fig. 4.3b. Similar to the diotic physiological model (Fig. 4.2b), binaural stimuli are first passed through a series of phenomenological cell models along each monaural pathway. Model synapse outputs are obtained from the AN and the CN models for both contralateral and ipsilateral sides. The excitatory response from the contralateral CN model is combined with a delayed (2 ms) inhibitory response from the ipsilateral CN model via an inhibitory interneuron. The combination of the CN outputs represents the binaural difference, which includes both interaural time and level differences. The combined excitatory and inhibitory inputs are sent to the IC modulation filter. For the IC model, a bandpass modulation filter was used to extract the envelope frequency around 50 Hz that contained the largest envelope difference related to tone presence. Envelope cues in terms of rate and response fluctuations were obtained from the model IC synapse output.

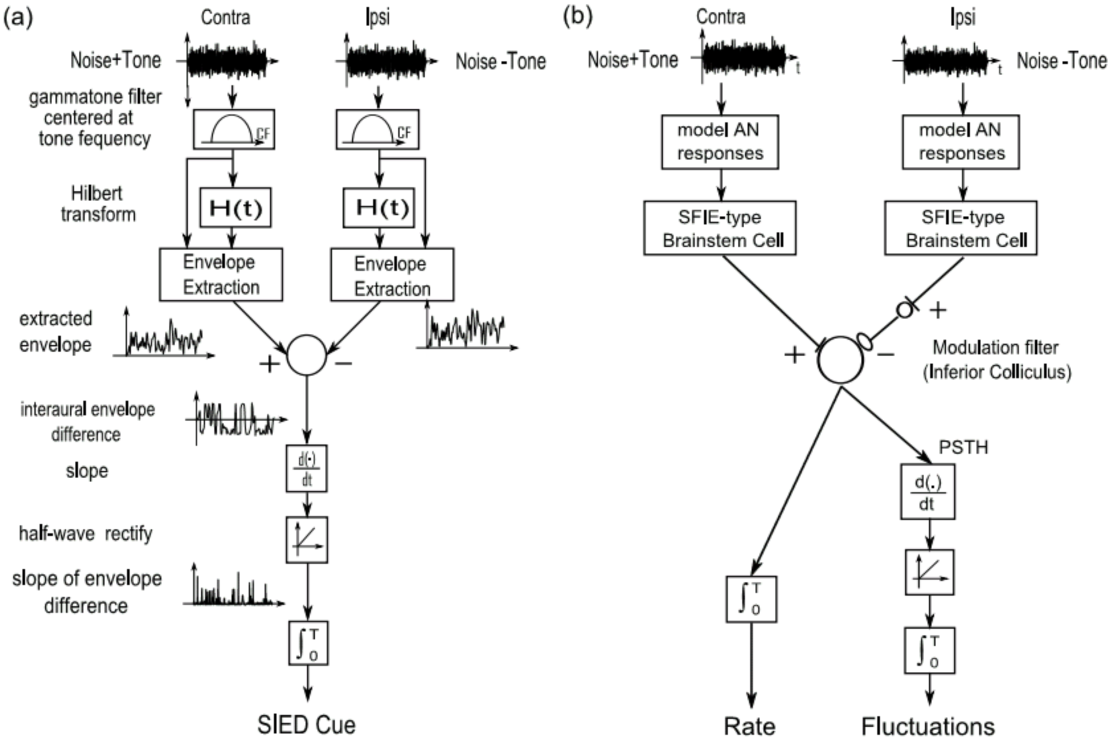

Figure 4.3: A schematic diagram of the dichotic envelope models. (a) Dichotic stimulus-based SIED cue: envelope was extracted using analytic signal computed from the Hilbert transform of the fourth-order gammatone filtered waveforms. The SIED cue was computed as the time integral of the half-wave rectified slope of the envelope difference at the two ears. (b) Dichotic physiological envelope cue: binaural stimuli were passed through the AN and the CN models; excitatory contralateral and delayed inhibitory ipsilateral CN outputs were combined to compute the binaural envelope difference. Responses from the CN outputs were sent to the IC cell. The IC cell was simulated with a bandpass modulation filter, and the envelope cue from the IC cell was computed based on the average rate and fluctuation of the model's response.

4.4.3 Evaluation of Model Predictions

Model predictions based on envelope cues were evaluated by comparing them to listeners' detection patterns. For each model, a DV was computed for each waveform. The proportion of the variance in the detection pattern explained by the model was computed as the squared Pearson product-moment correlation coefficient between the DVs and the z-score of listeners' detection patterns (Davidson *et al.*, 2009; Mao *et al.*, 2013). The variance predicted by each model was compared with the significance level (p<0.05). In addition, the variance explained by the stimulus-based and physiological models were compared to test the hypothesis that stimulus-based and physiological envelope cues could predict similar amounts of the variance in listeners' responses.

4.5 Results

In this section, model predictions using stimulus-based and physiological envelope cues are shown. Model AN fibers with different center frequencies and IC bandpass modulation filters with different best modulation frequencies were used in the physiological models. Basic neural mechanisms were used to compute cues from the IC model responses: rate and synchrony from the synapse output, and fluctuations of the model responses. Predictions computed using a synchrony cue are not shown here because synchrony to the 500-Hz tone was not significantly correlated to listeners' detection patterns.

Detection patterns were highly correlated across different pairs of listeners in the diotic narrowband and wideband, and dichotic wideband conditions (Mao *et al.*, 2013;

Chapter 3), indicating that listeners used a similar strategy to detect tones in noise in each of these conditions. In this study, model predictions are only shown for the average listeners in these three conditions. For the dichotic narrowband condition, in which listeners' patterns were not significantly correlated (Chapter 3), model predictions are shown for individual listeners. Model predictions using stimulus-based envelope cues have been reported in previous studies (Mao *et al.*, 2013 (Chapter 2); Chapter 3). Quantitative comparisons of stimulus-based and physiological envelope cues are shown in Tables 4.1 and 4.2.

4.5.1 Diotic Physiological Cues

Figure 4.4a-c shows model predictions of average listeners' narrowband detection patterns using stimulus-based envelope cues, average rates and fluctuations computed from the model IC cell responses. Predictions based on the same cues for the average listeners' wideband detection patterns are shown in Fig. 4.4d-f. In each panel, the x-axis shows the model center frequencies and the y-axis shows the proportion of variance in the detection pattern that is explained by the model.

In both narrowband and wideband conditions, the trends of model predictions across different frequency channels were similar, with the highest correlation to listeners' detection patterns obtained at or near 500-Hz tone frequency. In addition, maximal predictions from the stimulus-based envelope cue and the physiological rate and fluctuations cues were similar in these two conditions (Fig. 4.4a-c for narrowband; Fig. 4.4d-f for wideband). In the narrowband condition, the frequency range that yielded the highest correlation to listeners' hit rates was approximately 530 Hz (the target tone frequency was 500 Hz) for both stimulus-based and physiological cues. The reason that the maximal correlation to listeners' performance occurred for the 530-Hz channel was likely due to the phase properties of the gammatone filter, as no significant difference in envelope energy was observed across these channels. In the wideband condition, maximal correlations to listeners' detection patterns were obtained from model cells tuned near the tone frequency.

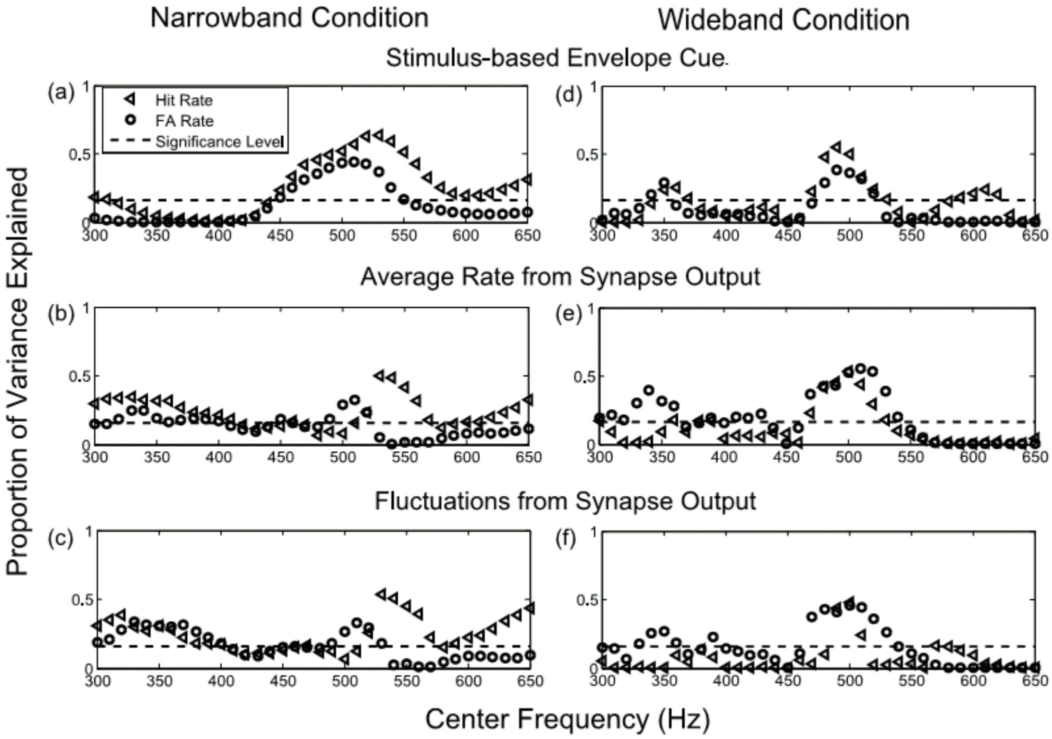

Figure 4.4: Stimulus-based and physiological model predictions of the average listener's hit (triangles) and FA (circles) rates based on stimulus-based envelope cues (a: narrowband, d: wideband), average rate (b: narrowband, e: wideband) and fluctuations (c: narrowband, f: wideband) computed from the model IC cell responses. The x-axis shows the center frequencies of the model cells and the y-axis shows the proportion of variance explained by the model. The black dotted line indicates the level required for significant predictions (p<0.05).

4.5.2 Dichotic Physiological Cues

For the stimulus-based SIED cue, FA rates cannot be predicted because there are no interaural differences in the identical noise-alone stimuli that were presented to the listener. In the physiological models, assuming that model IC cells receive inputs from

AN synapse outputs with mis-matched center frequencies (Joris *et al.*, 2006), predictions of both hit and FA rates can be computed. In this section, each figure shows dichotic model predictions with different combinations of mis-matched center frequencies. As described previously, the average listener was used in the wideband condition because listeners' detection pattern were highly correlated with each other, and individual listeners were used for the narrowband condition because listeners seemed to use different cues for narrowband stimuli.

In Fig. 4.5, dichotic model predictions of hit and FA rates for the average listener in the wideband condition (a-b) and three individual listeners in the narrowband condition (S1: c-d, S3: e-f, S4: g-h) are shown. These individual listeners were chosen because their results were representative of the other individual listeners. In each panel, the axes show the center frequencies of the model cells that received stimuli presented to the left and right ears. Predictions from the matched 500-Hz frequency channels are shown at the lower left corner.

For the wideband condition, the trends of model predictions using mis-matched AN inputs are similar: the highest correlation of hit rate was observed for models cells that received left and right inputs with AN tuning near the tone frequency of 500 Hz; for FA rates, predictions with ipsilateral inputs around 500 Hz were high (Fig. 4.5a-b). For the narrowband condition, the trends in the predictions varied across listeners: some were best predicted by frequency channels around 500 Hz, others were best predicted by frequency channels away from the tone frequency for both ears (Fig. 4.5c-g). For instance, S3’s detection patterns were best predicted by using rate information from frequency channels near 500 Hz, and S4’s patterns were best predicted by frequency channels that were approximately one-critical bandwidth apart. The diversity in these results implies that different listeners use different strategies for narrowband tone-in-noise detection, which also may explain the in low correlations between listeners’ detection patterns for this condition. In both narrowband and wideband conditions, model predictions were not symmetric around the matched center frequencies of 500 Hz. This is partly due to the fact that the physiological cues were obtained by combining positive contralateral and negative ipsilateral CN inputs, and exchanging the contralateral and ipsilateral inputs does not yield the same results.

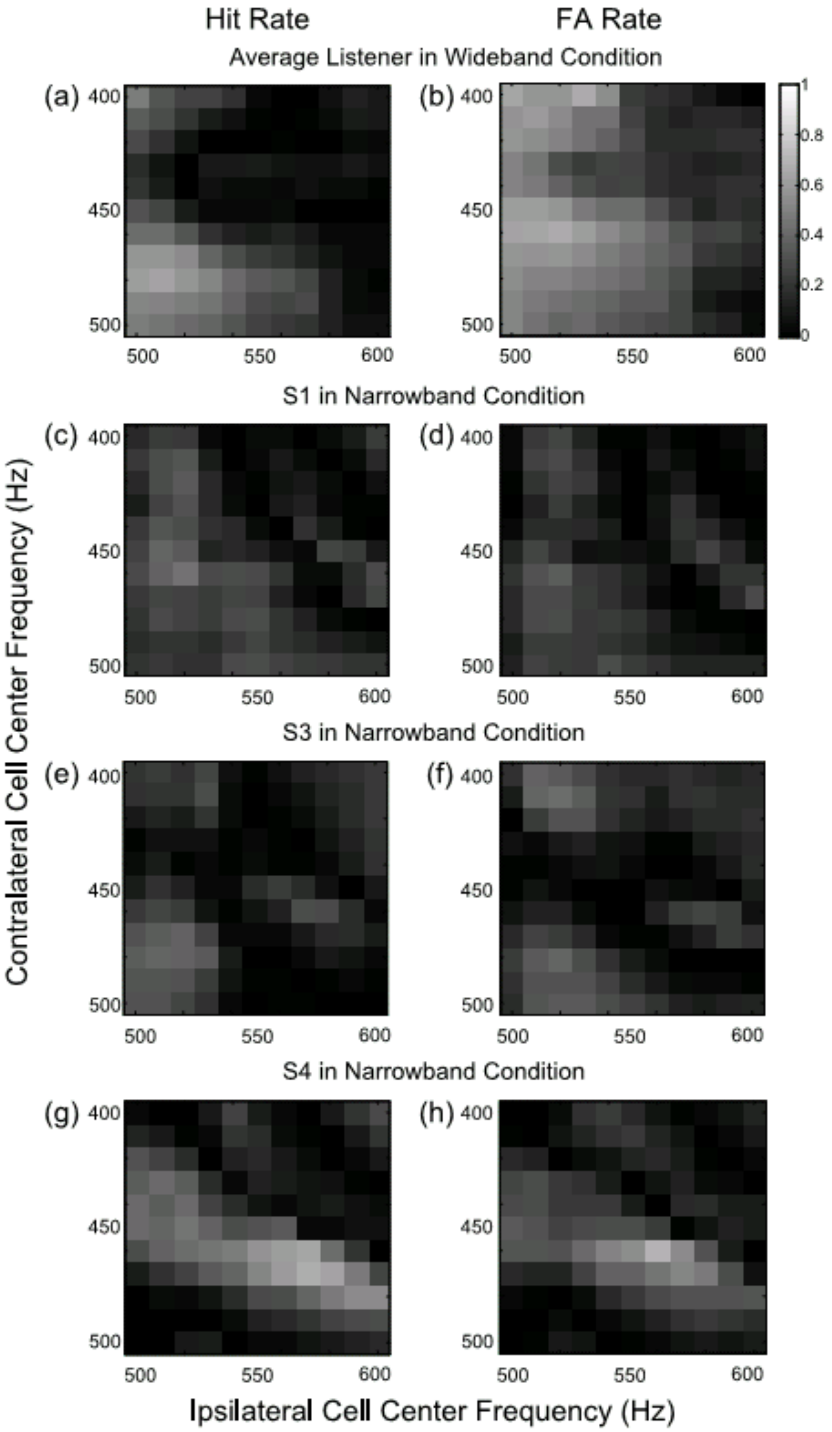

Figure 4.5: Physiological model predictions of the average listener's wideband hit and FA rates (a-b), and individual listeners' narrowband hit and FA rates (S1: c-d, S3: e-f, S4:

g-h) based on the average rate of the model IC synapse output. The x- and y-axis show the center frequencies of model cells receiving the stimuli presented to the left and right ears. The grayscale value shows the proportion of variance explained by the model.

Predictions based on the model response fluctuations for the average listener in the wideband (Fig. 4.6a-b) and individual listeners in the narrowband condition (S1: Fig. 4.6c-d, S3: Fig. 4.6e-f, and S4: Fig. 4.6g-h) are shown in Fig. 4.6. The overall trends in Fig. 4.6 are similar to results from Fig. 4.5, listeners' detection patterns were best predicted by a similar combination of frequency channels tuned near 500 Hz in the wideband condition, whereas different combinations of frequency channels yielded better predictions of listeners' patterns in the narrowband condition. However, there are some detailed differences between the trends in Figs. 4.5 and 4.6. In the wideband condition, for fluctuation cues the best frequency channels are located closer to 500 Hz compared with rate cues, though predictions from 600-Hz contralateral and 400-Hz ipsilateral inputs also yielded good predictions. For the narrowband condition, model predictions for the dichotic conditions based on model response fluctuations yielded a substantially higher correlation for some listeners' patterns than predictions using the rate cue.

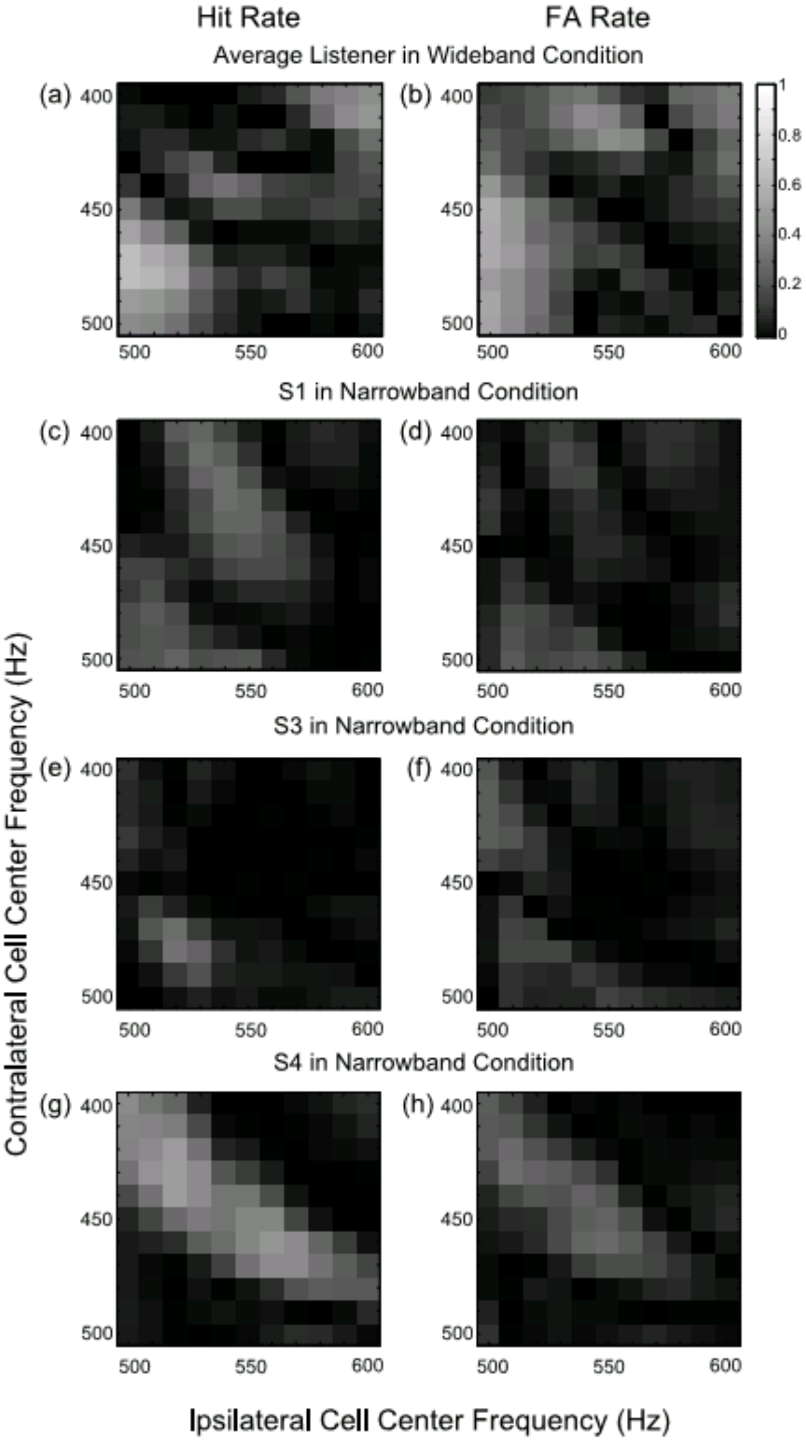

Figure 4.6: Physiological model predictions of the average listener's wideband hit and FA rates (a-b), and individual listeners' narrowband hit and FA rates (S1: c-d, S3: e-f,

and S4: g-h) based on fluctuations of the model IC synapse output. The x-and y-axis show the center frequencies of model cells receiving the stimuli presented to the left and right ears. The grayscale value shows the proportion of variance explained by the model.

Comparisons of model predictions for the average listeners using stimulus-based and physiological cues are shown in Table 4.1. Table 4.2 shows predictions for individuals in the dichotic narrowband conditions, because the average listener was not used in this condition. Both hit and FA rates were predicted using the physiological envelope cues for both diotic and dichotic conditions. For the stimulus-based cues, hit and FA rates were predicted for the diotic condition, whereas only the hit rate for the dichotic condition was predicted. As shown from the table, both types of physiological envelope models predicted similar amounts of variance of the average listeners' detection patterns in the wideband condition, and these predictions were similar to the stimulus-based cues. For narrowband dichotic condition, predictions from fluctuation cues were slightly, but not significantly better than rate cues for some listeners, and predictions from these physiological cues were substantially better than the SIED cue for S3 and S4.

Table 4.1: Stimulus-based and physiological model predictions of diotic and dichotic tone-in-noise detection patterns for the average listener. For each model, the center frequency for the model cell is 500 Hz and the best modulation frequency for the IC bandpass modulation filter is 100 Hz in the diotic condition and 50 Hz for the dichotic condition.

		Narrowband condition			Wideband condition		
		Stim ES	Physio Rate	Physio Fluctuations	Stim ES	Physio Rate	Physio Fluctuations
Diotic condition	Hit	0.53	0.50	0.54	0.50	0.53	0.48
	FA	0.46	0.32	0.33	0.38	0.52	0.46
Dichotic condition (Hit)		NA	NA	NA	0.53	0.46	0.52

Table 4.2: Stimulus-based and physiological model predictions for individual listeners' hit rates in the dichotic narrowband condition.

Model Predictions	Stimulus-based SIED Cue	Physio Rate		Physio Fluctuations	
	Hit	Hit	FA	Hit	FA
S1	0.18	0.31	0.23	0.30	0.22
S2	0.31	0.15	0.36	0.22	0.36
S3	0.00	0.26	0.28	0.31	0.24
S4	0.16	0.52	0.56	0.50	0.29
S5	0.10	0.28	0.41	0.38	0.30
S6	0.13	0.14	0.22	0.12	0.15

4.6 Discussion

Model predictions based on stimulus-based envelope cues and basic neural mechanisms from physiological IC model responses were tested with listeners' detection patterns for diotic and dichotic reproducible noises. For both listening conditions, similar or larger amounts of the variance in the average listener's detection patterns was explained by stimulus-based ES (diotic condition) and SIED (dichotic condition) cues and by the rate and fluctuation cues from the physiological model for IC responses.

In previous studies, model predictions for diotic and dichotic reproducible noises are computed from different types of cues. Davidson *et al.* (2009) evaluated several models in both binaural conditions. For the diotic condition, commonly used models are based on energy (Fletcher, 1940; Gilkey *et al.*, 1986), envelope (Richards, 1992; Zhang, 2004; Davidson *et al.*, 2006), fine-structure (Carney *et al.*, 2002), and template-matching-based temporal cues (Dau *et al.*, 1996; Breebaart *et al.*, 2001). For the dichotic condition, models are based on interaural cues, such as energy-based equalization and cancellation (Durlach, 1963), the normalized cross-correlation (Isabelle, 1995), interaural level and time difference cues (Isabelle, 1995), and different types of linear combination of interaural level and time differences (Isabelle and Colburn, 1987, 2004; Goupell and Hartmann, 2007; Davidson *et al.*, 2009).

In the current study, physiological monaural and binaural envelope cues were analyzed, motivated by the robustness of stimulus-based envelope cues for predicting listeners' performance in both diotic and dichotic conditions (Richards, 1992; Davidson *et al.*, 2006; Mao *et al.*, 2013; Mao *et al.*, 2011). Stimulus-based envelope cues (ES and SIED) have been used in previous studies (Mao *et al.*, 2013; Chapter Three). The physiological envelope cues studied here were based on the average rate and response fluctuations computed from a physiological model for IC neuron. The rate computation from the model IC responses can be interpreted as the response energy from the envelope-sensitive cell, and the fluctuations of the IC responses are similar to the envelope-slope cues used in the stimulus-based models. In addition, the synchrony of IC responses to the 500-Hz tone and synchronized-rate (the product of synchrony and rate) values were also computed, but significant correlations between DVs based on these metrics and listeners' detection performance were not observed.

For predictions in the diotic condition, average rates and response fluctuations of the IC model yielded similar maximal correlations to listeners' detection patterns as stimulus-based envelope cues. For the dichotic wideband condition, "envelope-slope" type fluctuation cues gave slightly, but not significantly, better predictions than predictions from "envelope energy" type rate cues. In addition, these physiological fluctuations cues better predicted listeners' narrowband dichotic condition than rate cues. Results from basic neural mechanisms involved in envelope processing suggest that physiological envelope cues are as reliable as the stimulus-based envelope cues in predicting listeners' tone-in-noise detection results.

Results from Figs. 4.5 and 4.6 indicate that for the dichotic wideband condition, for which listeners' patterns were highly correlated, similar frequency channels were used across listeners to detect tones in noise. In the dichotic narrowband condition, different combinations of frequency channels were most correlated to different listeners' patterns. Similar results were observed in Chapter Three (Fig. 3.7) when predicting dichotic detection patterns using SIED cue with gammatone filters that had mis-matched center frequencies. Both results suggest that different listeners use different cues (i.e., frequency channels) for the dichotic narrowband condition.

In addition to the success of envelope cues in exploring tone-in-noise detection, envelope cues have also been investigated for speech perception. Envelope cues are available for all complex stimuli that have amplitude modulations. Envelope cues have been investigated by purposely removing the amplitude modulations in these stimuli, leaving only the frequency modulations (e.g., temporal fine-structure speech, Lorenzi *et al.*, 2006). However, envelope cues can still be recovered from these flat-envelope stimuli by applying a narrowband filter (Drullman, 1995). This envelope recovery may explain the ability of listeners with normal hearing to understand temporal fine-structure speech, whereas listeners with hearing loss have difficulty. Lorenzi *et al.* (2006) showed that these listeners with hearing loss could use envelope cues recovered from temporal fine-structure cues, although their ability to use reconstructed envelope cues is worse than that of listeners with normal hearing. Swaminathan and Heinz (2012) further investigated the relative contribution of neural envelope and fine-structure cues. They showed that neural envelope was the primary cue, and neural fine-structure cue contributed in the noisy environments. Henry and Heinz (2012) found that chinchillas with sensorineural hearing loss have degraded peripheral temporal coding in noisy backgrounds, which explained the difficulty in noisy environments for listeners with hearing loss.

Although predictions of listeners' detection patterns based on envelope cues explained a significant amount of listeners' performance in the current study, these predictions were lower than the predictable variance for the diotic condition (Mao *et al.*, 2013). The predictable variance describes the proportion of the variation in detection patterns that is common among all listeners, and is used as a benchmark for model predictions. As shown in previous studies (Mao *et al.*, 2013), predictions based on an optimal nonlinear combination of stimulus-based energy and temporal cues approaches the predictable variance. In the current model, physiological envelope cues were analyzed from one model IC neuron with inputs from one model AN fiber. Assuming that physiological cues are Gaussian-distributed, it would be worth investigating model predictions based on an optimal combination of these cues across different frequency channels, using the likelihood-ratio-based method as in Mao *et al.* (2013).

In this study, the hypothesis that physiological envelope cues were as reliable as stimulus-based envelope cues in predicting listeners' tone-in-noise detection patterns was tested. In conclusion, predictions from physiological cues were similar to stimulus-based cues in diotic wideband and narrowband and dichotic wideband conditions. For the dichotic narrowband condition, in which listeners seemed to use different strategies, predictions from physiological cues explained substantially more of the variance of listeners' detection patterns than the stimulus-based binaural envelope cue for some listeners.

Appendix

The six-order bandpass filter (H) is computed by cascading three second-order bandpass filter (H_1, H_2, and H_3). The formula for the second-order bandpass filter (H_i) is

$H_i(z) = \frac{1-\alpha_i}{2}\frac{1-z^{-2}}{1-\beta_i(1+\alpha_i)z^{-1}+\alpha_i z^{-2}}$, where β is related to the center frequency, f_i, of H_i by $\beta_i = \cos(2\pi f_i)$, and α is related to the 3-dB bandwidth, W_i by $\alpha_i = \frac{1-\sin(2\pi W_i)}{\cos(2\pi W_i)}$.

Bibliography

Breebaart, J., van der Par, S., and Kohlrausch, A. (2001a). “Binaural processing model based on contralateral inhibition I. Model structure,” J. Acoust. Soc. Am. 110, 1074–1088.

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. Acust. Acust. 88, 334–347.

Carney, L. H., Mao, J., Koch, K.-J., and Doherty, K. A., (2013), “Modeling detection of 500-Hz tones in reproducible noise for listeners with sensorineural hearing loss,” *Proceedings of Meetings on Acoustics*, vol. 19, Montreal, Canada, June 2013.

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

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

Dau, T., Püschel, D., and Kohlrausch, A. (1996). “A quantitative model of the “effective” signal processing in the auditory system. I. Model structure,” J. Acoust. Soc. Am. 99, 3615–3622.

Davidson, S. A., Gilkey, R. H., Colburn, H. S., and Carney, L. H., (2006). “Binaural detection with narrowband and wideband reproducible noise maskers. III. Monaural and diotic detection and model results,” J. Acoust. Soc. Am. 119, 2258-2275.

Davidson, S. A., Gilkey, R. H., Colburn, H. S., and Carney, L. H., (2009). “An evaluation of models for diotic and dichotic detection in reproducible noises,” J. Acoust. Soc. Am. 126, 1906-1925.

Durlach, N. I., (1963). “Equalization and cancellation theory of binaural masking-level differences,” J. Acoust. Soc. Am. 35, 1206-1218.

Drullman, R., (1995). “Temporal envelope and fine structure cues for speech intelligibility,” J. Acoust. Soc. Am. 97, 585-592.

Evilsizer, M. E., Gilkey, R. H., Mason, C. R., Colburn, H. S., and Carney, L. H., (2002). “Binaural detection with narrowband and wideband reproducible maskers: I. Results for human,” J. Acoust. Soc. Am. 111, 333-345.

Fletcher, H. (1940). “Auditory patterns,” Rev. Mod. Phys. 12, 47–65.

Gilkey, R. H., and Robinson, D. E., (1986). “Models of auditory masking: a molecular psychophysical approach,” J. Acoust. Soc. Am. 79, 1499-1510.

Goupell, M. J., and Hartmann, W. M., (2007). “Interaural fluctuations and detection of interaural incoherence. III. Narrowband experiments and binaural models,” J. Acoust. Soc. Am. 122, 1029-1045.

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.

Henry, K. S., and Heinz, M. G., (2012). “Diminished temporal coding with sensorineural hearing loss emerges in background noise,” Nat. Neurosci. 15, 1362-1364.

Ibrahim, R. A., and Bruce, I. C., (2010). “Effects of peripheral tuning on the auditory nerve’s representation of speech envelope and temporal fine structure cues,” in *The Neurophysiological Bases of Auditory Perception*, eds. E. A. Lopez-Poveda, A. R., Palmer, and R. Meddis, Springer, NY, chapter 40, pp. 429-438.

Isabelle, S. K., (1995). "Binaural detection performance using reproducible stimuli," Ph.D. thesis, Boston University, Boston, MA.

Isabelle, S. K., and Colburn, H. S., (1987). "Effects of target phase in narrowband frozen noise detection data," J. Acoust. Soc. Am. 82, S109-S109.

Isabelle, S. K., and Colburn, H. S., (2004). "Binaural detection of tones masked by reproducible noise: Experiment and models," Report BU-HRC 04–01.

Joris, P. X., Schreiner, C. E., and Rees, A., (2004). "Neural processing of amplitude-modulated sounds," Physiol Rev. 84, 541-577.

Joris, P. X., Van de Sande, B., Louage, D. H., and van der Heijden, M., (2006). "Binaural and cochlear disparities," Proc. Natl. Acad. Sci. USA. 103, 12917-12922.

Kidd, G. Jr., Mason, C. R., Brantley, M. A., and Owen, G. A. (1989). "Roving-level tone-in-noise detection," J. Acoust. Soc. Am. 86, 1310-1317.

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.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C., (2006). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proc. Natl. Acad. Sci. USA. 103, 18866-18869.

Mao, J., Vosoughi, A., and Carney, L. H., (2011). "Stimulus-based diotic and dichotic models that combine cues for detection of tones in reproducible noise," *Conference of* *the Acoustical Society of America*, Abstract: *Journal of the Acoustical Society of America*, vol. 129, pp. 2489, Seattle, WA, May 2011.

Mao, J., Vosoughi, A., and Carney, L. H., (2013). "Predictions of diotic tone-in-noise detection based on a nonlinear optimal combination of energy, envelope, and fine-structure cues," J. Acoust. Soc. Am. 134, 396-406.

Moore, B. C. J., (2003). *An introduction to the psychology of hearing* (Elsevier Science & Technology Books).

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., (2007). "Neural rate and timing cues for detection and discrimination of amplitude-modulated tones in the awake rabbit inferior colliculus," J. Neurophyiol. 91, 522-539.

Richards, V. M., (1992). "The delectability of a tone added to narrow bans of equal energy noise," J. Acoust. Soc. Am. 91, 3424-3435.

Swaminathan, J., and Heinz, M. G., (2012). "Psychophysiological analyses demonstrate the importance of neural envelope coding for speech perception in noise," J. Neurosci. 32, 1747-1756.

Zhang, X., (2004). "Cross-frequency coincidence detection in the processing of complex sounds," Ph.D. thesis, Boston University, Boston, MA.

Zilany, M. S. A., and Bruce, I. C., (2006). "Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery," J. Acoust. Soc. Am. 120, 1446-1466.

Zilany, M. S. A., and Bruce, I. C., (2007). "Representation of the vowel /ε/ in normal and impaired auditory nerve fibers: model predictions of responses in cats," J. Acoust. Soc. Am. 122, 402-417.

Zilany, M. S., Bruce, I. C., Nelson, P. C., and Carney, L. H. (2009). "A phenomenological model of the synapse between the inner hair cell and auditory nerve: long-term adaptation with power-law dynamics," J. Acoust. Soc. Am. 126, 2390-2412.

Zilany, M. S. A., Bruce, I. C., and Carney, L. H., (2013). "Improved parameters and expanded simulation options for a model of the auditory periphery," ARO poster.

Chapter 5

Summary and Discussion

The goal of this thesis was to identify cues that listeners used in diotic and dichotic tone-in-noise detection experiments. Different types of models were proposed to predict listeners' performance in these two conditions. It was found that listeners likely used a nonlinear combination of energy and temporal cues for the diotic condition and used a binaural envelope cue for the wideband dichotic condition. Envelope processing along the auditory pathway based on physiological models predicted a similar or larger amount of the variance in listeners' detection patterns as the stimulus-based envelope cues for both diotic and dichotic conditions. In addition, our analysis revealed that different listeners apparently used different strategies for the narrowband dichotic condition.

5.1 Summary of Novel Results

In Chapter 2, a nonlinear cue-combination model based on a logarithmic likelihood-ratio test of energy and temporal cues was used to predict listeners' diotic detection patterns. A significant improvement of model predictions was observed using the likelihood-ratio test model compared to those using single energy or temporal cues or linear combinations of these cues. In Chapter 3, a binaural SIED model was proposed and explained a larger amount of listeners' wideband hit rates than any available binaural models. Furthermore, the SIED cue was shown to be a nonlinear combination of ILD and ITD cues. In addition to the SIED model, a linear cue-combination model that accounted for the correlation between ILD and ITD cues was also tested for the dichotic condition. In Chapter 4, a physiological model using responses from the model IC cell was used to predict listeners’ diotic and dichotic detection patterns. Basic neural mechanisms, such as averaged rate from synapse output and fluctuations from the model cell response, were used to predict a significant amount of the variance in listeners’ detection patterns. Physiological model predictions were similar to stimulus-based model predictions using envelope cues.

In summary, effective cues yielding predictions that were highly correlated to listeners’ detection patterns have been identified in this thesis. Successful identification of these cues for tone-in-noise detection is the first step to understand how listeners detect more complex signals. The stimulus-based and physiological envelope cues presented here and the strategy for optimal combination of available cues contributes to the understanding of the cocktail party effect.

5.2 Application of the Proposed Models to Other Studies

The following section discusses applications of the models presented proposed here to data from other studies in the lab. These data were from fixed- and roving-level tone-in-noise detection experiments from listeners with hearing loss and from budgerigars. A detailed presentation of these studies is beyond the scope of this thesis, but they are presented here as examples of the potential utility of these models.

5.2.1 LRT Model Predictions and the Envelope Cue for Listeners with Hearing Loss

For normal-hearing listeners, predictions from a likelihood-ratio-based nonlinear cue-combination model were highly correlated to their diotic detection patterns, and these predictions approached the predictable variance (Chapter 2). The same model was used to predict diotic tone-in-noise data for hearing-loss listeners (Carney *et al.*, 2013). However, model predictions based on the likelihood-ratio model were no better than those from the best single-cue model, suggesting that listeners with hearing loss do not use an optimal cue-combination strategy for detecting tones in noise. The failure of the optimal nonlinear cue-combination model might be related to the degraded temporal fine-structure cues for listeners with hearing loss, because energy and temporal envelope cues still predicted a significant amount of the variance in their detection patterns.

The diotic ES and dichotic SIED cues predicted a significant amount of normal-hearing listeners' detection patterns (Chapters 2 and 3). Given the robustness of envelope cues, it was interesting to know whether listeners with hearing loss could use envelope cues as well. In a recent study (Mao *et al.*, 2013), roving-level reproducible-noise stimuli (roving range 20 dB) were used to test thresholds and detection patterns in the diotic condition for listeners with hearing loss. It was shown that listeners with hearing loss had thresholds in the narrowband and wideband conditions that differed by less than 2 dB from their corresponding fixed-level thresholds; their detection patterns were highly correlated in these two conditions. Thus, it is likely that they used a similar strategy (or cue) in both experiments. Because the energy cue is unreliable in the roving-level condition and the envelope cues are the same in both conditions, these results suggested that listeners with hearing loss used the envelope cue in both conditions. The reason for the high correlations observed between listeners' detection patterns and the energy cues in the fixed-level condition is presumably related to the fact that energy and envelope cues are highly correlated.

5.2.2 LRT Model and Envelope Cue Predictions for Budgerigars

Budgerigars are good subjects for psychoacoustical studies because they can mimic sounds and are easy to train. Three English budgerigars were tested with fixed- and roving-level tone-in-noise detection with reproducible noises. In the fixed-level experiment, Carney *et al.* (2012) showed that budgerigars' detection thresholds were higher than thresholds from normal-hearing listeners, but comparable to those from listeners with hearing loss. Similar to predictions for listeners with hearing loss, predictions of budgerigars' diotic detection patterns based on the likelihood-ratio test model were not significantly better than those from energy or envelope models. This result suggested that budgerigars used both energy and envelope cues but not temporal fine-structure cues.

These budgerigars were further tested with roving-level experiments using narrowband and wideband noises. Similar to results from the listeners with hearing loss, no significant change of budgerigars' detection thresholds and patterns was observed for the wideband condition. However, for the narrowband condition, budgerigars' detection thresholds were significantly higher than those in fixed-level conditions, and their detection patterns were different. These results suggest that budgerigars use the energy cue for the narrowband conditions, and use the envelope cue for the wideband condition. The multiple-detector model (Chapter 2) uses energies in different bands and is also robust for the roving-level condition, thus it is possible that budgerigars use energy cues from the multiple-detector model for the wideband condition.

5.3 Future Study to Investigate the Dichotic Narrowband Conditions

Another problem worth investigating is the pair-wise correlation of listeners' detection patterns for the narrowband dichotic condition. Tone-in-noise detection studies from Evilsizer *et al.* (2002), Isabelle and Colburn (1991), and additional listeners tested recently (Chapter 3) showed that performance among different listeners was not significantly correlated in the dichotic narrowband condition. Interestingly, these listeners were also tested with diotic and dichotic, narrowband and wideband stimuli, and their detection patterns were highly correlated with each other in the other conditions. It is possible that the decreased correlations might be due to the different perceptions of tone presence in these conditions. To be more specific, narrowband stimuli are qualitatively different from wideband stimuli: even the noise-alone stimuli are more tone-like in the narrowband condition. Thus, it would be interesting to test listeners using stimuli of different bandwidths and to investigate whether there is a systematic change in consistency as a function of bandwidth.

Besides the different perceptions of tone presence in these conditions, it is also likely that different listeners used different strategies to detect tones in dichotic narrowband stimuli. Model predictions based on the SIED and physiological envelope cues (Chapters 3 and 4) showed that the best frequency channels (or combinations of frequency channels), which gave highest correlation to listeners' patterns, differed among individual listeners.

5.4 Implication for the Envelope Cues

The physiological model used in Chapter 4 was a single-channel model, computed from responses of one model auditory-nerve fiber and one model inferior colliculus cell. Results showed that frequency channels away from the tone frequency can also provide cues for tone presence and different listeners seem to used different frequency channels, thus it is possible that the inclusion of multiple channels would yield improved predictions.

As described above, the envelope cue is robust for normal-hearing listeners and those with hearing loss. Because technology for current hearing aids focuses on recovering temporal fine-structure, it is possible that focusing on envelope cues would be a promising alternative. In addition, the current study was based on tests of listeners with headphones, therefore reverberation was not included in the stimuli. In the real world (free field), reverberation is unavoidable, and it would be interesting and important to test how envelope cues would be affected in this condition.

Bibliography

Carney, L. H., Abrams, K. S., Mao. J., Schwarz, D. M., and Idrobo, F., (2012). "The Budgerigar as a Model for Human Detection of Tones in Noise," *Conference of the Society for Neurosicence*, New Orleans, LA, October 2012.

Carney, L. H., Mao. J., Kock, K-J., and Doherty. K. A., (2013). "Modeling Detection of 500-Hz Tones in Reproducible Noise for Listeners with Sensorineural Hearing Loss," *Proceedings of Meetings on Acoustics*, vol. 19, Montreal, Canada, June 2013.

Evilsizer, M. E., Gilkey, R. H., Mason, C. R., Colburn, H. S., and Carney, L. H. (2002). "Binaural detection with narrowband and wideband reproducible maskers: I. Results for human," J. Acoust. Soc. Am. 111, 333-345.

Isabelle, S. K., and Colburn, H. S., (1991). "Detection of tones in reproducible narrow-band noise," J. Acoust. Soc. Am. 89, 352-359.

Mao, J., Doherty, K. A., Kock, K-J., and Carney, L. H., (2013). "Effects of sensorineural hearing loss on roving-level tone-in-noise detection," *Conference of the American Auditory Society*, Scottsdale, AZ, March 2013.