

**Detection of Tones in Reproducible Noise:
Psychophysical and Computational Studies of Stimulus
Features and Processing Mechanisms**

Sean A. Davidson

ABSTRACT:

This dissertation combines psychophysical experiments with computational modeling efforts in order to help explain how humans detect sounds of interest in the presence of competing background noise. Specifically, tone-in-noise detection was examined for low-frequency tones (500 Hz) in the presence of reproducible narrowband masking noises. Both the N_0S_0 (noise and signal presented at the same phase to the two ears) and N_0S_π (noise presented at the same phase to the two ears and signal presented 180° out of phase between the two ears) interaural configurations were tested. Two psychophysical detection experiments are described that used prerecorded, or *reproducible*, masking waveforms in conjunction with multiple-regression data analyses. These experiments were designed to determine the dominant stimulus features used by individual listeners to compute detection cues. Candidate stimulus features included stimulus energy, temporal fine structure (e.g., zero crossings, or the stimulus carrier), temporal envelope, or a linear combination of carrier and envelope. Results indicated that listeners used energy cues when they were available for detection under N_0S_0 conditions, but that they also used temporal processing. Results also indicated that listeners did not separately process envelope and carrier under N_0S_0 or N_0S_π conditions. Computational modeling efforts were designed to approximate physiologically plausible stimulus processing along with several different decision devices. Several recent psychophysical detection models failed at predicting detection statistics for individual waveforms, demanding that new explanatory models for masked detection be examined. The characterization of detection cues used by listeners with normal hearing will lead to improved hearing aids. This improvement could occur either through the preservation of stimulus features found to be critical for detection in noise, or by mimicking the types of processing occurring in the healthy auditory system.

**Detection of Tones in Reproducible Noise:
Psychophysical and Computational Studies of Stimulus Features
and Processing Mechanisms**

By

Sean Adam Davidson

B.S. Syracuse University, 2002
M.S. Syracuse University, 2003

DISSERTATION

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

May 2007

Approved _____

Date _____

© Copyright 2007
Sean Adam Davidson
All Rights Reserved

Table of Contents

Abstract.....	i
List of Figures.....	xii
List of Tables.....	xiv
Acknowledgments.....	xvi
1 General Introduction.....	1
1.1 Psychophysical experiments.....	1
1.2 Modeling efforts.....	4
1.3 Relevant Physiology.....	6
1.4 Significance.....	7
2 Diotic and dichotic detection with chimaeric stimuli.....	8
Abstract.....	8
2.1 Introduction.....	8
2.2 Methods.....	10
2.2.1 General design.....	10
2.2.2 Stimuli.....	10
2.2.3 Training.....	13
2.2.4 Testing.....	14
2.2.5 Analysis.....	14
2.3 Results and discussion.....	15
2.3.1 Reliability of the data and detection performance.....	15
2.3.2 Within-subject comparisons of detection patterns estimated with baseline and chimaeric stimuli.....	20
2.3.2.1 N_0S_0 Stimuli.....	27
2.3.2.2 N_0S_π Stimuli.....	28
2.3.3 Comparisons between subjects	29
2.3.4 Comparisons between interaural configurations.....	32
2.3.5 Implications for computational modeling.....	33
2.3.5.1 Comparisons between detection patterns estimated with baseline and chimaeric stimuli.....	33
2.3.5.2 Comparisons between interaural configurations.....	33
2.4 Summary and future directions.....	34
3 An evaluation of models for diotic and dichotic detection in reproducible noises.....	35
Abstract.....	35
3.1 Introduction.....	35
3.1 Methods.....	39
3.2.1 General modeling strategy.....	39
3.2.2 Individual model implementations.....	39
3.2.2.1 Diotic models.....	39
3.2.2.1.1 Critical-band model.....	40
3.2.2.1.2 Multiple-detector model.....	40
3.2.2.1.3 Envelope-slope model.....	41
3.2.2.1.4 Dau model.....	41
3.2.2.1.5 Breebaart model.....	42

3.2.2.1.6	Phase-opponency model.....	43
3.2.2.2	Dichotic models.....	43
3.2.2.2.1	Isabelle (1995) and Goupell (2005) decision variables.....	43
3.2.2.2.2	Four-channel model.....	44
3.2.2.2.3	Breebaart model.....	46
3.2.2.2.4	Mismatch model.....	46
3.3	Results and discussion.....	47
3.3.1	Diotic models.....	47
3.3.1.1	Individual model results.....	47
3.3.1.1.1	Critical-band model.....	47
3.3.1.1.2	Multiple-detector model.....	47
3.3.1.1.3	Envelope-slope model.....	53
3.3.1.1.4	Dau model.....	53
3.3.1.1.5	Breebaart model.....	54
3.3.1.1.6	Phase-opponency model.....	54
3.3.1.2	Comparisons between models.....	55
3.3.1.3	Effect of stimulus energy.....	56
3.3.1.4	Between-condition comparisons of model decision variables for study 4.....	56
3.3.2	Dichotic models.....	62
3.3.2.1	Individual model results.....	62
3.3.2.1.1	Isabelle (1995) and Goupell (2005) decision variables.....	62
3.3.2.1.2	Four-channel model.....	63
3.3.2.1.3	Breebaart model.....	63
3.3.2.1.4	Mismatch model.....	63
3.3.2.2	Comparisons between Models.....	72
3.3.2.3	Between-condition comparisons of model decision variables for study 4.....	72
3.4	Summary and conclusions.....	75
4	General discussion and summary.....	78
4.1	Summary.....	78
4.2	An analysis of model decision variables using the methods of Ch. 2.....	79
4.3	Implications and suggestions for future experiments.....	81
4.3.1	Psychophysical experiments.....	81
4.3.2	Modeling exercises.....	82
4.3.2.1	Diotic models.....	82
4.3.2.2	Template-based models.....	82
Appendix: Diotic and dichotic detection under restricted-cue conditions.....		84
A.1	Introduction.....	84
A.1.1	Background.....	84
A.1.2	Experimental Design.....	85
A.2	Methods.....	87
A.2.1	Stimuli.....	87
A.2.2	Training.....	88
A.2.3	Testing.....	89
A.3	Results and discussion.....	90

A.3.1 Reliability of the data and detection performance.....	90
A.3.2 Comparisons between cue conditions.....	95
A.3.2.1 N_0S_0	96
A.3.2.2 N_0S_π	102
A.3.3 Comparisons between subjects and between interaural configurations.....	103
A.4 Conclusions and future directions.....	106
Bibliography	108

List of Figures

1-1	Schematic illustration of a detection pattern.....	2
1-2	Schematic illustration of experimental methodology.....	5
2-1	Chimaeric stimulus construction and multiple-regression analysis procedure.....	11
2-2	General liner model predictions for the $N_0S_0 E_1C_1$ detection pattern.....	21
2-3	General liner model predictions for the $N_0S_0 E_2C_2$ detection pattern.....	22
2-4	General liner model predictions (N_0S_0) grouped by predictor.....	23
2-5	General liner model predictions for the $N_0S_\pi E_1C_1$ detection pattern.....	24
2-6	General liner model predictions for the $N_0S_\pi E_2C_2$ detection pattern.....	25
2-7	General liner model predictions (N_0S_π) grouped by predictor.....	26
3-1	Block diagrams of N_0S_0 models.....	40
3-2	Block diagrams of N_0S_π models.....	45
3-3	Proportion of variance of $P(Y T+N)$ explained by N_0S_0 models.....	48
3-4	Proportion of variance of $P(Y N)$ explained by N_0S_0 models.....	49
3-5	Proportion of predictable variance of $P(Y T+N)$ explained by N_0S_0 models.....	50
3-6	Proportion of predictable variance of $P(Y N)$ explained by N_0S_0 models.....	51
3-7	Multiple-detector model weights.....	52
3-8	Dau-model templates.....	54
3-9	Breebaart-model templates and frequency weights (N_0S_0).....	55
3-10	Proportion of variance of $P(Y T+N)$ explained by the N_0S_π models.....	60
3-11	Proportion of predictable variance of $P(Y T+N)$ explained by the N_0S_π models.....	61
3-12	Model weights for decision variables based on both ITDs and ILDs.....	64
3-13	Proportion of variance of $P(Y T+N)$ explained by the N_0S_π models (cont.).....	65
3-14	Proportion of variance of $P(Y N)$ explained by the N_0S_π models (cont.).....	66
3-15	Proportion of predictable variance of $P(Y T+N)$ explained by the N_0S_π models (cont.).....	67
3-16	Proportion of predictable variance of $P(Y N)$ explained by the N_0S_π models (cont.).....	68
3-17	Four-channel model weights.....	69
3-18	Breebaart-model templates and frequency weights (N_0S_π).....	70
A-1	Comparisons between detection patterns estimated under the different cue conditions for the N_0S_0 interaural configuration.....	97
A-2	Comparisons between detection patterns estimated under different cue conditions for the N_0S_π interaural configuration.....	98
A-3	Comparison of variances of detection patterns under different cue conditions.....	99

List of Tables

2-1	N_0S_0 reliability and detection performance statistics.....	16
2-2	N_0S_0 reliability statistics: $P(Y T+N)$ and $P(Y N)$	17
2-3	N_0S_π reliability and detection performance statistics.....	18
2-4	N_0S_π reliability statistics: $P(Y T+N)$ and $P(Y N)$	19
2-5	Comparisons between subjects: $P(Y W)$	30
2-6	Comparisons between subjects: $P(Y T+N)$ and $P(Y N)$	31
2-7	Comparisons between interaural configurations: $P(Y W)$	32
2-8	Comparisons between interaural configurations: $P(Y T+N)$ and $P(Y N)$	32
3-1	Correlations between N_0S_0 model decision variables: $P(Y T+N)$	57
3-2	Correlations between N_0S_0 model decision variables: $P(Y N)$	58
3-3	Correlations between the Ch.2 cue conditions of N_0S_0 models.....	59
3-4	Mismatch model parameters.....	71
3-5	Correlations between N_0S_π model decision variables (low SNR).....	73
3-6	Correlations between N_0S_π model decision variables (high SNR).....	74
3-7	Correlations between the Ch.2 cue conditions of N_0S_π models (low SNR).....	76
3-8	Correlations between the Ch.2 cue conditions of N_0S_π models (high SNR).....	77
4-1	Contribution of envelope and carrier to model decision variables.....	80
A-1	Expected correlations between cue conditions for naive hypotheses.....	86
A-2	N_0S_0 reliability and detection performance statistics.....	91
A-3	N_0S_0 reliability statistics: $P(Y T+N)$ and $P(Y N)$	92
A-4	N_0S_π reliability and detection performance statistics.....	93
A-5	N_0S_π reliability statistics: $P(Y T+N)$ and $P(Y N)$	94
A-6	Proportion of variance predicted by an energy model in RNRE and LNRE conditions.....	101
A-7	Comparisons between subjects: $P(Y W)$	103
A-8	Comparisons between subjects: $P(Y T+N)$ and $P(Y N)$	104
A-9	Comparisons between interaural configurations: $P(Y W)$	106
A-10	Comparisons between interaural configurations: $P(Y T+N)$ and $P(Y N)$	106

Acknowledgments

This dissertation would not exist without the support of a number of individuals. My advisor, Dr. Laurel Carney, is the noblest of scientists and engineers. Her work ethic is the envy of others, and her modesty and generosity command respect and admiration. My “distant” advisor, Dr. Robert Gilkey, is an expert psychoacoustician with an unmatched eye for detail. I hold Dr. Carney and Dr. Gilkey in the highest regard, and I look up to them both as mentors, and as friends. I appreciate the time and effort they have devoted to this project and will always be grateful for their help and support. Dr. Steven Colburn was also very involved with this project. His insightful comments were always appreciated (as was his organization and hosting of the yearly “Binaural Bash” at Boston University). I also thank my committee members, Drs. Marc Howard, Karen Doherty, Bob Smith, and Beth Prieve, for providing helpful comments as readers. I thank Dr. Marty Sliwinski for providing valuable statistical advice.

At the inception this project, Drs. Mike Anzalone, Paul Nelson, Felicia Gai, and Lauren Calandruccio were my fellow graduate students and friends. I have learned very much from each and every one of them and express my gratitude for their support and encouragement. I am also grateful for the companionate support of my mother, father, and fiancé.

This work would not have been possible without the ongoing financial support of the National Institute on Deafness and other Communication Disorders [DC 01641 (LHC) and DC 07798 (SAD)].

CHAPTER 1

General Introduction

The human auditory system's ability to resolve signals of interest in background noise is remarkable. A notable and often-described example of this ability is the cocktail-party effect, a situation in which a person is able to understand a particular conversation ("signal") amongst a number of competing conversations ("noise"). The psychophysical experiments and computational modeling efforts described in this thesis were designed to help understand how the auditory system detects signals in noisy environments in the context of masked detection. Pure tones were used as target stimuli and narrow-band noise waveforms were used as masking stimuli. The term masking simply refers to the use of the noise waveform to make the tone difficult to detect (i.e., the noise is masking the perception of the tone). Here the focus was on both diotic (in which the noise and signal were presented in-phase to the two ears, denoted as N_0S_0) and dichotic (in which the noise was presented in-phase to the two ears, and the signal was presented 180° out of phase to one ear with respect to the other, denoted as N_0S_π) detection experiments that used a small set of *reproducible*, or prerecorded, stimulus waveforms.

1.1 Psychophysical Experiments

Traditional studies of auditory masking phenomena used randomly fluctuating noises and were concerned with detection thresholds for tones in various noise bandwidths, levels, and stimulus configurations (e.g., Diercks and Jeffress, 1962; Egan, 1965; Shaw *et al.*, 1947). These studies describe what are now commonly referred to as masking level differences (MLDs), or differences in detection thresholds between stimulus configurations. The largest MLDs are observed between the N_0S_0 and N_0S_π interaural configurations. (As much as a 20-dB reduction in threshold for N_0S_π conditions has been observed with respect to N_0S_0 conditions; Durlach and Colburn, 1978; Moore, 2003). Durlach and Colburn (1978) provide an excellent review of the pioneering work of researchers such as Jeffress, Licklider, and Hirsh. These studies focused on obtaining thresholds while varying specific physical parameters of the stimuli. Summarizing their results; the MLD is greatest for signals 180° out of phase between the ears and noise maskers presented in phase to the 2 ears ($N_0S_0 - N_0S_\pi$, Jeffress *et al.*, 1952), for noise spectrum levels above about 40 dB SPL (Diercks and Jeffress, 1962), for narrow noise bandwidths (Metz *et al.*, 1968), and for signal frequencies below about 500 Hz (Hirsh, 1948).

More recent masked-detection studies use what are known as *reproducible* maskers. Studies employing reproducible maskers are capable of precisely characterizing human detection performance for individual masking waveforms (Pfafflin and Matthews, 1965; Ahumada *et al.*, 1975; Gilkey *et al.*, 1985; Siegel and Colburn, 1989; Isabelle and Colburn, 1991; Evilsizer *et al.*, 2002; and Davidson *et al.*, 2006). The experiments described in these studies use repeated presentations of several different masking waveforms presented in a random order within each block. On each trial, either a noise alone (N) or a tone-plus-noise (T+N) waveform is presented, and the listener indicates whether the tone was heard. Over many trials, reliable detection statistics are established for each individual noise waveform (e.g., Siegel and Colburn, 1989). These statistics include hit rate, $[P(Y|T+N)]$, or the probability of responding "yes, tone present" to a particular tone-plus-noise waveform] and false-alarm rate, $[P(Y|N)]$ or the probability of responding "yes, tone present" to a particular noise-alone waveform]. $P(Y|T+N)$ is calculated for each individual T+N waveform as the number of trials with "tone present" responses divided by the total number of trials that particular T+N waveform was presented. $P(Y|N)$ is calculated for

each individual N waveform as the number of trials with “tone present” responses divided by the total number of trials that particular N waveform was presented.

The group of $P(Y|T+N)$ and $P(Y|N)$ values for all noises (computed separately for each waveform) is referred to as a detection pattern [$P(Y|W)$], or the probability of responding “tone present” for each Waveform, [i.e., if the $P(Y|T+N)$ values and $P(Y|N)$ values for each waveform are considered as a single group of probabilities, $P(Y|W)$ results]. An illustration of a detection pattern is shown in Fig. 1-1. Hit rates are shown in the upper panel and were computed from T+N trials. False-alarm rates are shown in the lower panel and were computed from N trials. Corresponding hit and false-alarm rates for 3 masking waveforms are described below to give the reader an intuitive feel for the data presented in the following chapters. The $P(Y|T+N)$ and $P(Y|N)$ values for masker 3 are both large, indicating that the listener almost always perceived a tone, regardless of its actual presence. The $P(Y|T+N)$ and $P(Y|N)$ values for masker 8 are both small, indicating that the listener almost never perceived a tone, regardless of its presence. The $P(Y|T+N)$ value for masker 25 is large, indicating that the listener perceived the tone on T+N trials, but the $P(Y|N)$ value for masker 25 is low, indicating that the listener did not perceive the tone on N trials. The main findings of experiments using reproducible maskers (and thus

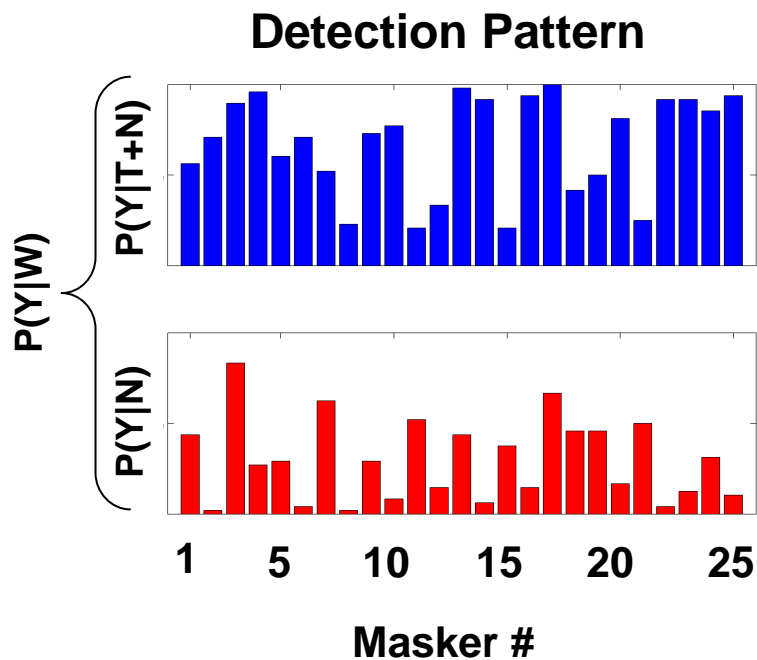


Figure 1-1. Schematic illustration of a detection pattern. Hit rates, or probabilities of “yes, tone present” responses for each tone-plus-noise waveform are shown in the upper panel [$P(Y|T+N)$]. False-alarm rates, or probabilities of “yes, tone-present” responses for each noise-alone waveform are shown in the lower panel [$P(Y|N)$]. When both $P(Y|T+N)$ and $P(Y|N)$ are considered together, the term $P(Y|W)$ is used, or the probability of a “yes, tone present” response for any waveform.

estimating detection patterns) are summarized below. Modeling efforts are described in Sec. 1.2.

Three recent studies describing detection with reproducible noises are summarized below. Gilkey *et al.* (1985) estimated detection patterns with wideband (100-3000 Hz) noises under both the N_0S_0 and N_0S_π interaural configurations, and used several different tone-to-

masker phase values. They found that $P(Y|W)$ values were correlated between the N_0S_0 and N_0S_π conditions and that $P(Y|T+N)$ values varied significantly with the tone-to-masker phase value.

Isabelle and Colburn (1991) performed a similar experiment with 1/3-octave noise bandwidths centered at 500 Hz. Using their narrowband noises, they found small correlations between $P(Y|W)$ values for N_0S_0 and N_0S_π conditions, which differed from the wideband results of Gilkey *et al.* (1985). Isabelle and Colburn did find significant differences in $P(Y|T+N)$ values with tone-to-masker phase in the N_0S_0 condition, which is in agreement with the result of Gilkey *et al.*, however, they found negligible differences in $P(Y|T+N)$ values with tone-to-masker phase in the N_0S_π condition.

In Evilsizer *et al.* (2002), both narrowband (100 Hz) and wideband (2900 Hz) reproducible noise waveforms were used to mask a 500-Hz tone in both the N_0S_0 and N_0S_π interaural configurations using the same set of subjects. Also, the frequency content of the noise in the 100-Hz region surrounding the tone frequency (450-550Hz) was identical for the 100-Hz and 2900-Hz waveforms. Results were consistent with the previously described studies (Gilkey *et al.*, 1985; Isabelle and Colburn, 1991). The results of the Evilsizer *et al.* (2002) study provide a detailed template for testing models because they provide detection patterns *and* information about the relationships between detection patterns for different bandwidths and interaural configurations. This was not possible in previous work, because different masking waveforms and subjects were used across studies and bandwidths.

Detection patterns vary depending on noise bandwidth and stimulus configuration. In order to quantify these relationships, correlations between detection patterns for different stimulus configurations and bandwidths have been calculated and are summarized below (Gilkey *et al.*, 1985; Isabelle and Colburn, 1991; and Evilsizer *et al.*, 2002). Four relationships are present in detection patterns: (1) Detection patterns for narrowband N_0S_0 and narrowband N_0S_π configurations are not correlated. (2) Detection patterns for narrowband N_0S_π and wideband N_0S_π configurations are not correlated. (3) Detection patterns for wideband N_0S_0 and wideband N_0S_π configurations are are correlated. (4) Detection patterns for narrowband N_0S_0 and wideband N_0S_0 configurations are are weakly correlated.

These relationships indicate several features of tone-in-noise detection. The processing strategies used for wideband N_0S_0 and N_0S_π detection tasks, although yielding different overall thresholds, produce relatively similar $P(Y|T+N)$ and $P(Y|N)$ for each waveform, indicating that some components of diotic and binaural processing are similar. The relatively weak correlation between detection patterns for narrowband and wideband stimuli (Evilsizer *et al.*, 2002; Davidson *et al.*, 2006) indicates that either additional energy within a critical band (in the filter skirts) or in adjacent critical bands affects the detection patterns.

Several factors motivate the further study of temporal contributions to tone-in-noise detection: Green (1983) has shown that in a 2-interval, 2-alternative tone-in-noise detection task, when levels¹ of stimuli are roved within a trial and across intervals, critical-band model thresholds increase by about 1/4 of the rove range. Kidd *et al.* (1989) provides psychophysical data that differs from the threshold increase predicted by the critical band model. Additionally, the work of Eddins and Barber (1998) has shown that thresholds differ for stimuli of equal

¹ Note that throughout this document, “level” will refer to the overall energy present in a stimulus waveform, while “energy” refers to the energy present at the output of a 75-Hz BW auditory filter, corresponding to 1 ERB at 500 Hz (Glasberg and Moore, 1990). If the bandwidth of the stimulus is narrow relative to the bandwidth of the auditory filter, level and overall energy are virtually identical.

energy but differing envelope fluctuations, phenomena that the critical band model cannot explain.

One of the goals of this dissertation was to begin a systematic exploration of cues that may be used for tone-in-noise detection. Rather than comparing detection patterns estimated at different noise bandwidths (such as in the work described above), stimulus features were altered in specific ways to observe the effect of those alterations on the detection patterns. This concept is illustrated in Fig. 1-2 A. If the stimulus features altered were substrates for detection cues, the detection pattern was expected to change. If the stimulus features altered were not substrates for detection cues, no change in the detection pattern was expected.

Figure 1-2 B illustrates the method for comparing the patterns before (ordinate) and after (abscissa) the stimulus features were altered. First, z scores were computed for each probability value in the detection patterns in order to normalize the data. Then, the z scores from each pattern were compared using standard linear regression techniques. The comparison was quantified by the r^2 value, or the square of the correlation coefficient, for the experiment described in the Appendix, or by the R^2 statistic for the multiple regression procedure described in Ch. 2. Both of these statistics quantify the proportion of variance explained in the “baseline” detection pattern by one or more “altered” detection patterns.

The preliminary experiment described in the Appendix explored the roles of energy and temporal structure (using low-noise noise, Pumplin, 1985) in shaping the detection patterns. Detection patterns that were estimated in conditions with overall energy equalized (within T+N and N stimuli) were compared to patterns that had differences in overall energy from stimulus to stimulus (but had otherwise identical temporal structures). This comparison illustrated the effect of preserving temporal structure, and was performed using both Gaussian and low-noise noise. Detection patterns with energy differences from stimulus to stimulus were compared to sets having the same energy differences, but different temporal structures, testing the effect of altering the temporal structure of the stimulus waveforms while preserving corresponding waveform energies between conditions.

The experiment described in the Appendix was a precursor of the experiment described in Ch. 2, which also explored detection under both N_0S_0 and N_0S_π conditions. However, overall energies were equalized for all stimuli in the N_0S_0 conditions to prevent the use of energy as a detection cue. Baseline detection patterns were predicted using detection patterns estimated from sets of stimuli sharing the same temporal envelopes (but different temporal fine structures), or from sets of stimulus sharing the same temporal fine structures (but different temporal envelopes), or from a combination of the two.

1.2 Modeling Efforts

Several models for masked detection have been applied to psychophysical data collected with reproducible maskers (Ahumada and Lovell, 1971; Gilkey and Robinson, 1986; Isabelle, 1995; Colburn *et al.*, 1997). A black-box representation of the modeling procedure is shown in Fig. 1-2 C. The challenge for modeling data collected with reproducible maskers is prediction of psychophysical detection patterns, which is accomplished by producing sets of model decision variables for the reproducible stimuli. If the decision variables resulting from the model and data correlate perfectly, a particular noise waveform that effectively masks the tone for the listener will also effectively mask the tone for the model. Conversely, if a listener responds to a particular noise waveform by consistently indicating that the tone is present (regardless of its actual presence), then model should produce a decision variable indicating tone presence.

Predicting the detection pattern is a more critical test than predicting an average threshold (which the model must do as well).

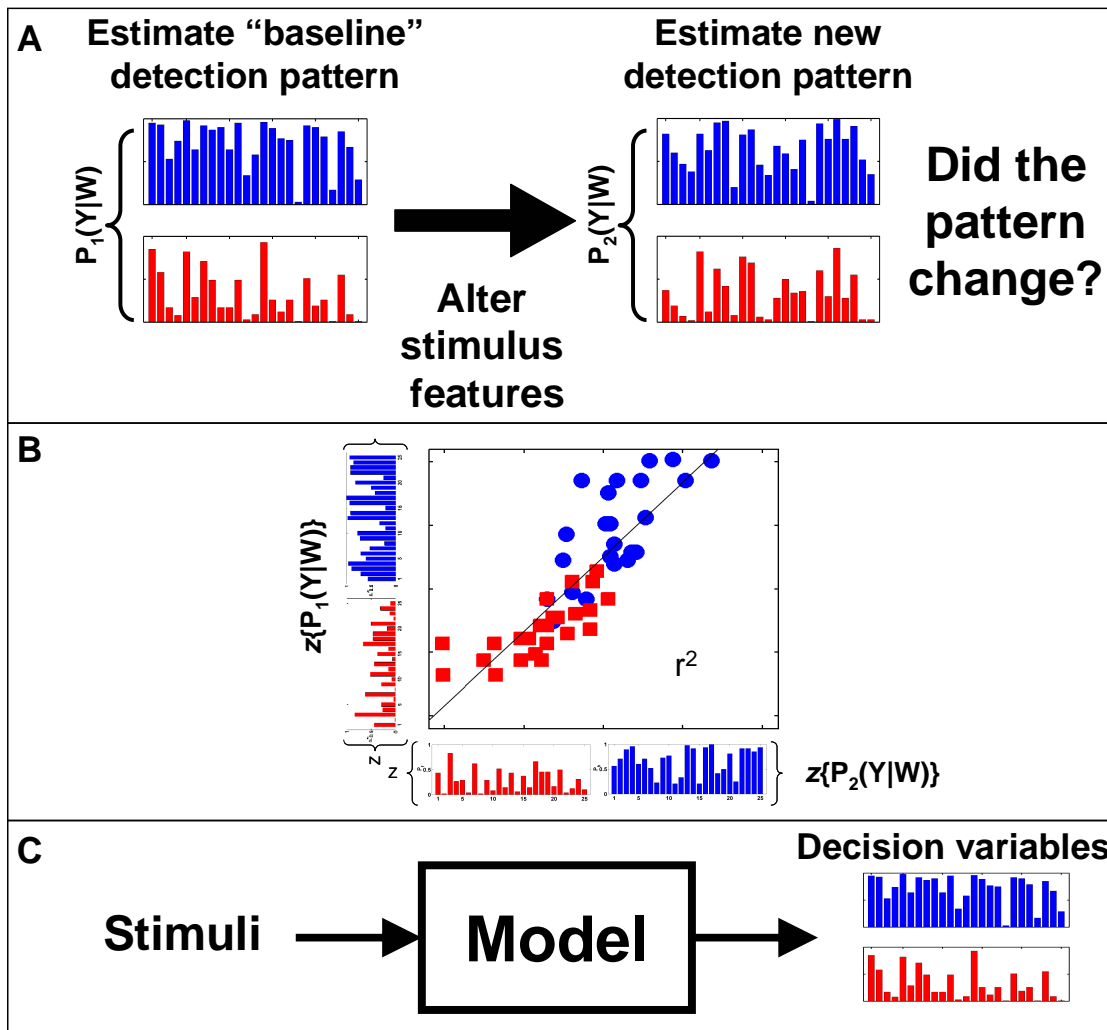


Figure 1-2. A. Simplified schematic illustrating the general strategy of the experiments in the Appendix and in Ch. 2. If the stimulus features altered were used in the detection process, the pattern is expected to change. B. Example illustration of a quantitative comparison between two detection patterns. The z scores of the individual probabilities in each detection pattern are used to normalize the data. The z scores of the “baseline” condition (1) are on the ordinate and the z scores of the “new” condition (2) are on the abscissa. The r^2 value is calculated as the square of the correlation coefficient. C. Schematic illustrating the modeling procedures in Ch.3. In the N_0S_0 case, the model produces a single decision variable for each reproducible noise waveform. In the N_0S_π case, the model produces a single decision variable for each pair of left and right stimulus waveforms.

Initial attempts to predict N_0S_0 detection patterns were based on a linear combination of energy at the output of a series of band-pass filters (Ahumada and Lovell, 1971 and Ahumada *et al.*, 1975). These studies used a weighting scheme determined by fitting model detection patterns to human detection patterns collected using wideband reproducible noises. Gilkey and Robinson (1986) extended this work using a multiple-detector model that could explain about 71 percent of the variance in N_0S_0 detection patterns. Multiple-detector models are essentially multi-channel

extensions of the critical-band model (Fletcher, 1940). Chapter 3 will examine the abilities of a multiple-detector style model to predict the N_0S_0 results from the experiments in the Appendix and Ch. 2, as well as those from Evilsizer *et al.* (2002) and will present comparisons of these predictions to those from a basic energy model (Fletcher, 1940) and predictions from more recent psychophysical models (Dau *et al.*, 1996a, b and Breebaart *et al.*, 2001a, b, c).

Few have attempted to predict N_0S_π detection patterns. Isabelle (1995) and Colburn *et al.* (1997) tested several stimulus-based binaural models with narrowband reproducible noises. These included models based on the normalized cross correlation of the stimulus waveforms, interaural time and level differences (ITDs and ILDs respectively), combinations of ITD and ILD, lateral position, time-deviation and the equalization-cancellation (EC) model (Durlach, 1972; Colburn *et al.*, 1997). The overall conclusion from these studies is that none of the models examined predicted more than about 50 percent of the variance in N_0S_π detection patterns. Goupell and colleagues (e.g., Goupell, 2005; Goupell and Hartman, 2005, 2006) modeled data from a related group of experiments using modified versions of decision variables described by Isabelle (1995). The task in these experiments was to detect interaural coherence using a set of reproducible noises. (It should be noted that this task is very different from tone-in-noise detection in that there is no specific signal present to be detected or modeled.) Results from those studies were more promising [i.e., a larger proportion of variance in subjects' responses was explained in Goupell (2005) than in Isabelle (1995)].

Chapter 3 examines several signal-processing-style detection models. Models were tested under both the N_0S_0 and N_0S_π interaural configurations with both narrowband and wideband stimuli. Energy-related diotic models included the critical-band (Fletcher, 1940) and multiple-detector models (e.g., Gilkey *et al.*, 1986). The envelope-based models included the Dau *et al.*, (1996a,b), Breebaart *et al.* (2001a, b, c) and envelope-slope (Richards, 1992) models. A temporal phase-opponency model (Carney *et al.*, 2002) was also examined. Binaural models included two of the Isabelle (1995) decision variables and also two Goupell (2005) decision variables based on temporal interactions of interaural time and level differences. A relatively modern excitatory-inhibitory processing strategy was also examined in the form of the Breebaart *et al.* (2001a) model. The utility of decision devices based on templates was also considered. [Here templates are defined as averages of stored peripherally- or centrally- transformed representations of the stimulus waveforms as described in Dau *et al.* (1996a,b) and Breebaart *et al.* (2001a, b, c).] Symmetrical processing strategies incorporating interaural mismatches of frequency and delay channels were examined, as well as a model inspired by the results of McAlpine *et al.* (2001) and Marquardt and McAlpine (2001), using the output of only 4 interaural delay channels, with delay dependent on the best frequency of the channel.

1.3 Relevant Physiology

No physiological detection studies have used reproducible-noise stimulus paradigms similar to the psychophysical studies mentioned above. This is not surprising, given the number of complications such a study would create, the most critical of which is the question of how to construct ensembles of reproducible stimuli for cells of differing center frequencies. One study has attempted to examine an ensemble of reproducible waveforms in order to compare neurological responses within and across waveforms (Shackelton and Palmer, 2006). The primary conclusion from that study was that only 27 percent of the variance in responses could be explained by across-stimulus variability. This finding is at odds with the results of the psychophysical detection studies using reproducible noises as described above (e.g., Siegel and Colburn, 1989; Isabelle and Colburn, 1991), which generally find very reliable (and different)

detection statistics for individual waveforms. There are some notable differences between the studies that may explain this discrepancy. First, Shackelton and Palmer's stimuli were noises of varying interaural correlation or interaural time differences only, neither of which have been shown to predict $P(Y|W)$ measured in psychophysical tone-in-noise detection experiments. Second, they used the average rates of individual neurons for their computations, which neglect the use of temporally-varying cues. Third, they recorded from anesthetized Guinea pig inferior colliculus (IC). If animals had been attending to these stimuli, the results may have differed. Finally, the psychophysical studies involving reproducible noise used stimuli 100 ms in duration or longer, generally with 10-ms onset and offset ramps, whereas Shackelton and Palmer used 50-ms stimuli with 2-ms onset and offset ramps. Such stimuli would be considered too distracting to be used in a psychophysical task due to the spectral spread of energy at the onsets and offsets causing perceptual clicks.

1.4 Significance

Overall, this combination of psychophysical experimentation and modeling work will help us understand how humans extract signals of interest from the environment. Such an understanding will be achieved by identifying and manipulating cues present in carefully constructed stimuli and using these cues to predict detection patterns. Understanding how the detection process is performed on a waveform-by-waveform basis is fundamentally different from predicting average thresholds. Such a waveform-specific model allows for an understanding of cues that may also be employed in processing more complicated stimuli such as amplitude-modulated tones and eventually speech. Once these cues are more completely characterized, better hearing aids, cochlear implant processors, and noise-reduction systems may be developed by either better preserving or enhancing the cues used by the brain to understand signals in noise.

Diotic and dichotic detection with chimaeric stimuli

ABSTRACT

Hit rates and false-alarm rates were estimated from a closed set of reproducible tone-plus-noise and noise-alone waveforms under several different conditions in order to identify the components of sound waveforms from which detection cues were derived. Two sets of corresponding waveform envelopes and fine structures were combined to form four sets of stimuli. Two of the sets shared the same envelopes but had different fine structures, and two sets shared the same fine structures but had different envelopes. Detection patterns estimated for each of the four sets were compared to reveal the listeners' reliance on either fine structure or envelope for stimuli presented under both the N_0S_0 and N_0S_π interaural configurations. Results varied across listeners, but in general suggested that detection cues were based on a combination of both waveform envelope and fine structure. Implications for computational models are discussed.

2.1 INTRODUCTION

Over the past few years, many researchers, using a variety of different approaches, have investigated the question of whether the auditory system processes envelope and fine structure separately (e.g. van de Par and Kohlrausch 1998; Kohlrausch *et al.*, 1997; Eddins and Barber, 1998; Breebaart *et al.*, 1999; Smith *et al.* 2002, Joris, 2003; and Zeng *et al.*, 2004). It has long been known that auditory-nerve (AN) responses phase lock to individual cycles and to the envelope of stimuli with low-frequency carriers (Kiang *et al.*, 1965) and to the envelopes of stimuli with high-frequency carriers (Joris and Yin, 1992; Kay, 1982). Thus the auditory system codes both cycle-by-cycle information and envelope information at low frequencies, and information based only on the stimulus envelope at high frequencies, raising the question: Which type of processing is used for detecting a low-frequency tone in a narrowband noise waveform: envelope or fine timing? This dichotomous thinking has produced a variety of psychophysical models that rely on envelope (e.g., Dau *et al.*, 1996a, b, Eddins and Barber, 1998) or on fine structure (e.g., Moore, 1975). Models that rely on the entire stimulus waveform have also been described (e.g., Durlach, 1963; Colburn, 1977; Breebaart *et al.*, 2001a).

Several relatively recent studies have examined the roles of stimulus envelope and fine structure in perception using chimaeric stimuli. For example, Smith *et al.* (2002) define the concept of chimaeric stimuli as a stimulus created by combining the envelopes of certain sounds (music, speech, noise etc.) with the fine structures of other sounds. They tested speech recognition and sound localization using various chimaeras and suggested that speech identification appeared to be based on envelope, whereas sound localization appeared to be based on fine structure. Zeng *et al.* (2004) later showed, by creating chimaeras with directionally conflicting interaural-time differences (ITDs; which were embedded in the fine structure) and interaural-level differences (ILDs; which were embedded in the envelope), that sound localization was not based entirely on signal fine structure. In the present study, chimaeric stimuli will be used in a tone-in-noise detection experiment to examine the use of cues based on stimulus fine-structure and stimulus envelope.

If the discussion is restricted to low-frequency *diotic* stimuli, the concept of basing detection cues on stimulus fine structure or envelope is simply a question of whether the cue is computed from fast or slow fluctuations in the stimulus waveform. However, this concept

requires some clarification for *dichotic* stimuli that are processed binaurally. Previous researchers (e.g., Davidson *et al.*, 2006; Isabelle, 1995; and Richards, 1992) have implemented signal-processing-style detection models that operate on signal envelopes extracted from the narrowband analytic signal using the Hilbert transform. For binaural processing, such models compute ITDs from the stimulus fine structure and ILDs from the stimulus envelope, while neglecting the possible use of ITDs based on the stimulus envelope.

Results from several physiological studies help illustrate more realistic binaural processing of envelope and fine structure. These studies have focused on the medial-superior olive (MSO; e.g., Goldberg and Brown, 1969; Yin and Chan, 1990), where ipsilateral and contralateral excitatory inputs converge (EE); the lateral-superior olive (LSO; e.g., Boudreau and Tsuchitani, 1968; Joris and Yin, 1995; Joris, 1996) where ipsilateral excitatory inputs and contralateral inhibitory inputs converge (IE); and the inferior colliculus (IC; e.g., Hind *et al.*, 1963, Kuwada *et al.*, 1987, Fitzpatrick *et al.*, 2000, and Joris, 2003) where cell responses show many characteristics of both the MSO and LSO cell types.

In general, the MSO is biased toward low frequencies and is commonly described as a center where cells are sensitive to ITDs present in the stimulus fine structure (e.g., Yin and Chan, 1990). Joris (1996) found using amplitude-modulated (AM) binaural-beats, that the relatively few high-frequency MSO cells also have weak envelope ITD sensitivity. Joris and Yin (1995) found that high-frequency cells in the LSO encode ILDs and ITDs in stimulus envelopes. Further, they showed that LSO cells sensitive to both ITDs and ILDS had smaller dynamic ranges (i.e., change in rate) for coding ITD than for coding ILDs (if only the physiologically-relevant range of possible ITDs was considered). They also found that low-frequency IE-type LSO cells encode ITDs in the stimulus fine structure. Tollin and Yin (2005) describe the responses of several low-frequency (<1.5 kHz) neurons in the LSO that were sensitive to both to ITDs and ILDs.

Joris (2003) shows, for cells in the IC, that phase locking to the stimulus envelope occurs with higher gain and at wider bandwidths for cells tuned to high frequencies than to low frequencies. The lowest frequency for which the binaural processing of stimulus envelope applies is not entirely clear. For the purposes of this study, we will consider ITDs based on either stimulus fine structure or stimulus envelope and ILDs based on stimulus envelope to be plausible cues for detecting a 500-Hz tone in noise.

The present study is related to a psychophysical experiment completed by van de Par and Kohlrausch (1998) using multiplied-noise maskers, which are created by modulating a sinusoidal carrier with a low-pass noise. The resulting noise waveform is centered at the carrier frequency and has twice the bandwidth of the low-pass noise. By carefully manipulating the phases of the signal waveforms, van de Par and Kohlrausch (1998) created binaural stimuli having *only* ITDs or *only* ILDs (that is, having the same fine structure or same envelopes, respectively; envelope ITDs were ignored). They found similar N_0S_π thresholds for ITD-only and ILD-only stimuli at low frequencies (<1000 Hz) using 25-Hz masker bandwidths, supporting the notion that either fine-structure based or envelope-based cues could have been used by subjects in the detection task. However, they also found that some listeners had masking-level differences (MLDs) for ITD-only stimuli at high frequencies (4000 Hz) for which physiological coding of fine-timing information is weak. These MLDs were attributed to peripheral transformations that may have converted frequency modulations to amplitude modulations (Blauert, 1981).

At low frequencies, the question of whether envelope or fine-structure based decision variables can be separated, and if so, which dominates the detection process remains. The study

described here was designed to “passively” investigate the use of cues for low-frequency tone-in-noise detection based directly on stimulus envelope or stimulus fine structure, or a linear combination of both, using chimaeric stimuli. The word “passively” is used to remind the reader that no stimulus modifications² were performed in order to prevent the use of one interaural cue over another (i.e., as in van de Par and Kohlrausch, 1998; and). Both N_0S_0 and N_0S_π stimuli will be considered.

2.2 METHODS

2.2.1 General Design

This experiment was designed to investigate the contribution of cues based on stimulus fine-structure and cues based on stimulus envelope as putative decision variables for a tone-in-noise detection task. Four sets of reproducible stimuli were created, two pairs of which shared stimulus fine structures (or carriers) and two pairs of which shared stimulus envelopes. The two primary conditions, E_1C_1 and E_2C_2 (E denoting envelope; and C denoting carrier), were created first. The third and fourth sets of stimuli, E_1C_2 and E_2C_1 , were created by recombining the envelopes and carriers of the first two sets. (Details regarding the construction of stimuli will be discussed below.) Detection patterns were estimated for each of the four sets of stimuli. The detection patterns were compared within and across subjects using standard regression techniques.

Experimental procedures were adapted from those of Davidson, *et al.* (2006), Evilsizer *et al.* (2002) and Gilkey *et al.* (1985). As in the previous experiments, listeners performed tone-in-noise detection under diotic and dichotic conditions using reproducible noises. Training and testing procedures were performed in a double-walled sound attenuating booth (Acoustic Systems, Austin, TX). Six subjects completed the experiment, all of whom had previous listening experience. S3 was the author of the present paper. S2 and S5 had extensive training with psychophysical tasks.

2.2.2 Stimuli

Stimuli were created and controlled by MATLAB software (Mathworks, Natick, MA) and presented with a TDT System III (Tucker Davis Technologies, Gainesville, FL) RP2 D/A converter.

Initially, two sets of 25 noise waveforms were generated. Each waveform was created in the frequency domain by randomly selecting 5 magnitudes from a Rayleigh distribution and 5 phase values from a uniform distribution on the interval $[-\pi, \pi]$. The inverse Fourier transform was used to generate the time-domain noise waveforms. All waveforms were 100 msec in duration and had 50-Hz bandwidths centered at 500 Hz. Each of the noise waveforms was normalized to the overall level of 57 dB SPL, which corresponds to a 40-dB SPL spectrum level. Tones were added at the levels determined during training (see below) for both the N_0S_0 and N_0S_π interaural configurations. For the N_0S_0 condition, the resulting tone-plus-noise (T+N) stimuli were once again normalized to the overall level of 57 dB SPL to eliminate cues based on overall level. The N_0S_π T+N stimuli were not normalized in order to avoid introducing static interaural level differences. The resulting sets of stimuli are denoted E_1C_1 and E_2C_2 . A schematic illustrating these conditions is shown in Fig. 2-1 A. The waveform shown for each condition is a representative example. The detection pattern icons are included for each condition to remind the reader that the same type of processing occurred for each T+N and N waveform in both N_0S_0 and N_0S_π conditions.

² with the exception of stimulus energies in the N_0S_0 conditions, which were all normalized to the same overall level.

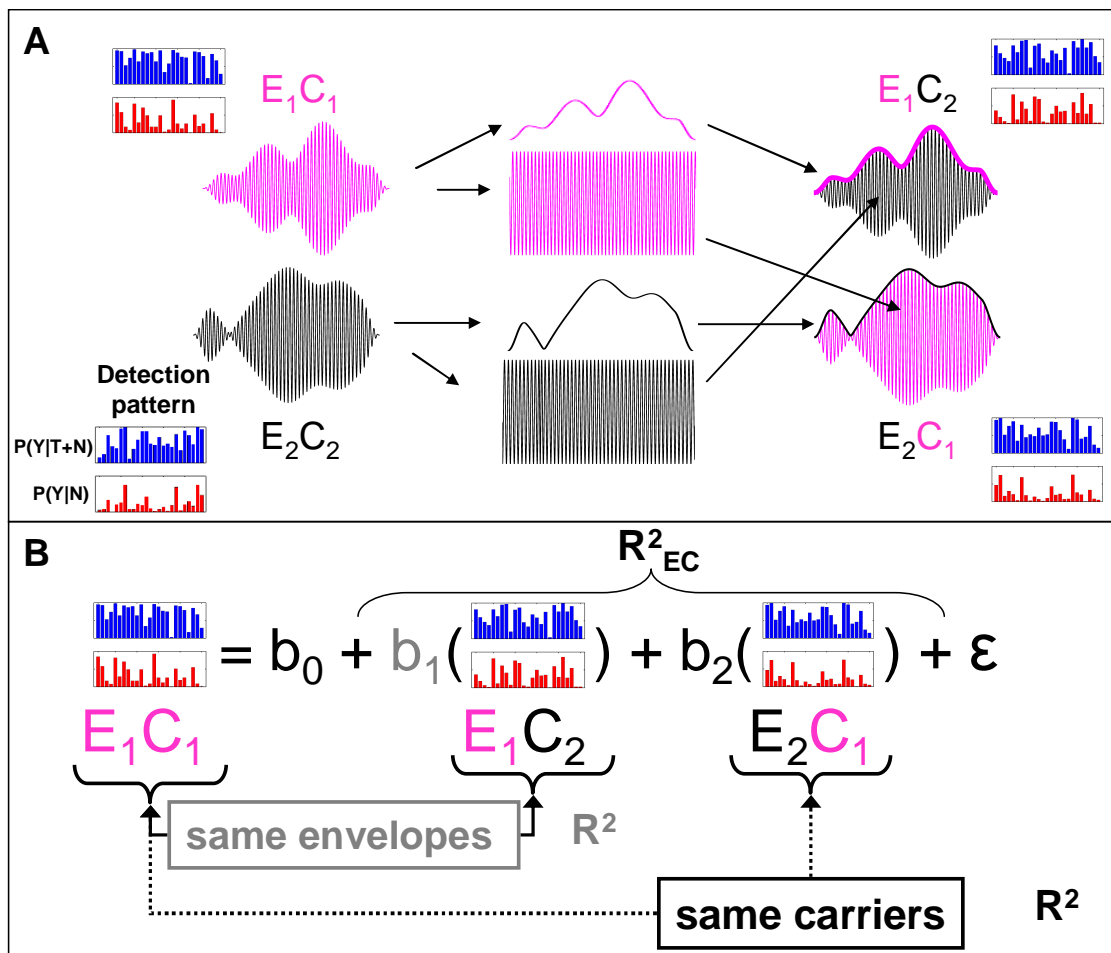


Figure 2-1. A. Schematic illustration of the stimulus-construction procedure Envelopes (E) and carriers (C) were separated from the E_1C_1 and E_2C_2 stimuli using the Hilbert transform. The envelopes and carriers were exchanged and recombined to create chimaeric waveforms E_1C_2 and E_2C_1 . Detection patterns are present to remind the reader that each stimulus waveform is representative of an entire set of stimuli. A more detailed description of the stimuli (including distortion control procedures) is given in the text. B. Illustration of the multiple-regression procedure for the E_1C_1 condition Chimaeric detection patterns sharing envelopes (E_1 in the example above) and sharing carriers (C_1 above) were used to predict the detection pattern from the baseline condition (E_1C_1 above). The b coefficients represent the slopes of the regression lines used in the linear statistical model. The b_0 coefficient is always equal to zero because variability linearly associated with the baseline condition not in the model (E_2C_2 above) was removed (see text for details). The ϵ term represents error variance. R^2 values were computed for envelope (gray) carrier (black) and a linear combination of envelope and carrier (R^2_{EC}). If envelope dominated the detection process, it was expected that the E_1C_1 and E_1C_2 detection patterns would be the same and the R^2 envelope = 1. If carrier dominated the detection process, it was expected that the E_1C_1 and E_2C_2 detection patterns be the same and the R^2 carrier = 1.

Each T+N and noise-alone (N) stimulus waveform was then decomposed into its complex analytic function using the Hilbert transform to extract signal envelope and fine structure (Oppenheim *et al.*, 1999). The envelopes from the first set of stimuli (E_1) were combined with the carriers from the second set of stimuli (C_2) to produce a third set of stimuli (E_1C_2). The envelopes from the second set of stimuli (E_2) were combined with the carriers from the first set of stimuli (C_1) to produce a fourth set of stimuli (E_2C_1). Thus, four sets of stimuli were created, with the latter two sets created from combinations of the envelopes and carriers from the first and second sets of stimuli. Stimuli from the latter two sets will be referred to as chimaeric stimuli. Stimuli from the first two sets will be referred to as baseline stimuli. Signal presence and corresponding waveforms were preserved across the sets. For example, the right-ear, N_0S_π , T+N waveform number 1 from the third set of stimuli was created with the envelope from right-ear, N_0S_π T+N waveform number 1 from the first set of stimuli and the carrier from the right N_0S_π T+N waveform number 1 from the second set of stimuli.

The process of assembling chimaeric stimuli in some cases resulted in waveforms with spectral splatter and temporal distortions. Because these distortions had the potential to interfere with the task and cause unintended interaural differences (in the N_0S_π condition), the waveforms were tested for excessive spectral splatter and eliminated based on specific criteria. If absolutely no spectral splatter was allowed, the stimulus-creation algorithm would eventually create four sets of stimuli with identical corresponding waveforms (see below). Chimaeric stimuli were therefore eliminated if their bandwidth exceeded 60 Hz at a magnitude 15 dB below the waveform's spectral peak, and a bandwidth of 100 Hz at a magnitude 25 dB below the waveform's spectral peak. (Recall that the baseline stimuli had a bandwidth of 50 Hz.) In the event a waveform was eliminated, the corresponding waveforms across all 4 sets of stimuli were also eliminated. Stimuli that were eliminated tended to have large frequency modulations in the carrier that were temporally positioned at relatively high envelope values. Such a combination naturally increased the bandwidth of the waveform.

Two new baseline waveforms were created using random noise, and corresponding chimaeric stimuli were created. The stimuli were scaled, tones added, and the resulting waveforms tested. The process continued for all waveforms in all 4 conditions. The algorithm ran for approximately 12 hours on a Pentium M computer (1.86 GHz), and eliminated thousands of candidate stimulus waveforms before converging on the set used in the present study (the exact number of waveforms eliminated was not recorded). This process resulted in stimuli that had minimal distortions and spectral splatter, but also produced significant differences in masking across waveforms.

Blauert (1981) and Zeng *et al.* (2004) have pointed out that when relatively broadband stimuli are filtered with a filter narrower than the stimulus bandwidth, an envelope may be recovered. This was not likely to occur given the approximate 75-Hz critical bandwidth at 500 Hz, and the fact that a 50-Hz noise bandwidth was used. Nevertheless, stimuli were diagnostically tested for possible envelope recovery by filtering all stimuli with a 50-Hz bandwidth, 4th-order gammatone filter at center frequencies from 400-600 Hz in 1 Hz steps. Envelopes were then recovered from the stimuli by half-wave rectification and filtering with a first-order low-pass filter with an 8 Hz cut-off frequency. First, envelopes from the filtered chimaeric stimulus sets (E_1C_2 and E_2C_1) were compared to the envelopes from the filtered original stimulus sets (E_1C_1 and E_2C_2 respectively). Under no cases (i.e., at any filter center frequency or for any waveform) did the correlation value fall below 0.977. Then, the correlations between the envelopes extracted in the E_1C_1 and E_2C_2 conditions were subtracted

from the correlations between the envelopes extracted in the E_1C_2 and E_2C_2 conditions, and also from the correlations between the envelopes extracted from the E_2C_1 and E_1C_1 conditions. This comparison examined whether the envelope of the baseline conditions (e.g., E_1) was recovered from the carriers of the chimaeric conditions (e.g., E_2C_1). Under no cases did the correlations differ by more than 0.05, indicating that at these stimulus and filter bandwidths, the recovery of envelope information from stimulus-fine structure by peripheral filtering was unlikely.

2.2.3 Training

Training procedures were similar to those described in Davidson *et al.* (2006) and will be briefly summarized here. An extensive training paradigm was used encourage subjects to form detection strategies that remained constant over the duration of the experiment in order to establish stable performance for the final testing procedure, which was a single-interval task using large numbers of trials at threshold. Threshold is defined here for each subject as the E_s/N_0 value in dB where $d' \approx 1$. Three separate training tasks were completed, and each task was progressively more similar to the final testing procedure. The training procedures used 50-Hz bandwidth, 100-ms duration noise waveforms that were generated randomly on each trial (i.e., not the reproducible stimuli used in the testing procedure). Randomly-generated noise was used to prevent any possible learning of reproducible stimuli.

The following training and testing procedures were conducted under both the N_0S_0 and N_0S_π interaural configurations. In general, subjects received stimuli from only one interaural configuration per session (2-3 hours). The use of N_0S_0 or N_0S_π stimuli alternated by session for all subjects but S2 and S4. S2 and S4 completed all training (and testing) for the N_0S_π condition first, and then completed all training and testing for the N_0S_0 condition. This change was made to reduce possible confusion of the diotic and dichotic cues created by switching interaural configurations between sessions. (Note that S3 completed the experiment both alternating interaural configurations by session and also by completing the N_0S_π interaural configuration first. Detection patterns from the two testing methods were highly correlated.)

In rare cases, stimuli from both interaural configurations were presented in the same session (such as to finish a particular training or testing paradigm). During those sessions, presentation of the blocks of stimuli never alternated between the two configurations. The initial listening configuration was randomized across subjects.

During the first training procedure, each subject completed 10-15 repetitions of a two-interval two-alternative forced-choice tracking procedure with trial-by-trial feedback to estimate a level where $d'_{2AFC} = 0.77$. Each track had a fixed length of 100 trials. The step size was maintained at 4 dB for the first 2 reversals and dropped to 2 dB thereafter. Thresholds were estimated by averaging tone levels at all but the first 4 or 5 reversals in the track such that the number of reversals averaged was even. Subjects were instructed to “select the interval containing the tone” and learned the task using the trial-by-trial feedback.

During the second training procedure, a single-interval, fixed-level task with feedback was used to encourage stable performance at each subject’s approximate threshold. The instructions for the single-interval tasks were to “determine whether the tone was present” on each trial and to click on a button labeled either “tone” or “no tone.” Approximately 10 blocks containing 100 trials each were completed at +3, +1 and -1 dB relative to the threshold established in the two-interval task. Throughout the single-interval training procedures (and the testing procedure described in Sec. 2.2.4), d' and bias (β , MacMillan and Creelman, 1991) were monitored. The d' values calculated from these blocks were used as an accurate estimate of tone level where d' was approximately equal to unity, rounded to 0.5- or 1-dB resolution.

Approximately 10 blocks were then run at that tone level. If a subject's threshold changed, the tone level was once again adjusted with 0.5 or 1-dB resolution until d' returned to unity.

After a stable tone level was established, the feedback was removed, and subjects completed approximately 10 100-trial blocks without feedback in order to determine whether d' values would remain near unity after feedback was removed. In rare cases, tone-levels were adjusted with 1-dB resolution such that $d' \approx 1$. The block length was then increased to 400 trials, and subjects completed 5 more blocks.

If a listener was noticeably biased (i.e., β departed more than 15 percent from unity, with unity indicating an equal probability of guessing "tone" or "no tone") the subject was given verbal feedback to "try and make an equal number of tone and no tone responses." Subjects were also notified that $\beta < 1$ indicates too many "tone" responses and $\beta > 1$ indicates too many "no tone" responses. The values of d' and β were computed using $P(Y|T+N)$ (the probability of a "yes" response conditional on individual T+N waveform, or hit rate) and $P(Y|N)$ (the probability of a "yes" response conditional on individual N waveform, or false-alarm rate) across all stimulus waveforms from the four cue conditions, and were not monitored within each of the conditions. No attempt was made to control for variations in values of d' and β computed for the individual envelope and carrier sets (e.g., E_1C_1) during the course of the experiment.

2.2.4 Testing

The testing procedure was identical to the final training procedure except that the reproducible noises described in Sec. 2.2.2 were used as stimuli. Before each 400-trial block, 20 practice trials (that did not use reproducible stimuli) were presented with feedback. Each T+N and N stimulus from each of the 4 stimulus sets was presented twice in a randomly interleaved order in each 400-trial block. A total of 50 blocks were presented to each listener such that 100 presentations of each T+N and each N waveform were presented at the final tone level in both the N_0S_0 and N_0S_π conditions.

The narrowband-noise waveforms used in training were random and did not include chimaeric stimuli. As a result, the tone level determined from the training procedure did not necessarily represent the level where $d' \approx 1$ for each subject when using the sets of reproducible noise waveforms. In these cases, the tone level was adjusted in 0.5 or 1-dB steps until $d' \approx 1$ for each subject. The tone level was adjusted at least once for each listener, which was most likely a consequence of the specific stimuli selected with the distortion control algorithm. Learning during this process was unlikely, as the long training procedure with feedback was designed to encourage subjects to establish a fixed decision strategy. Feedback was never presented while testing with the reproducible noise waveforms.

2.2.5 Analysis

The reliability of the data was first verified by calculating the d' , β , r^2 , and χ^2 statistics. Following these computations, detection patterns [$P(Y|W)$, or the probability of a "yes" response conditional on a particular stimulus waveform] were analyzed across the various stimulus conditions and interaural configurations.

Detection patterns were first compared within subjects and across the chimaeric stimulus sets, which required several steps: Initially, all $P(Y|W)$ values were converted to z -scores. The general strategy was then to predict each of the baseline detection patterns (E_1C_1 or E_2C_2) with the chimaeric detection patterns (E_2C_1 and E_1C_2) using multiple regression. This regression procedure is illustrated in Fig. 2-1 B. To the extent that detection cues were based on the carrier, the detection patterns for conditions E_1C_1 and E_2C_1 should be the same, and detection patterns for conditions E_2C_2 and E_1C_2 should be the same. To the extent that detection cues were based on

envelope, the detections patterns for conditions E_1C_1 and E_1C_2 should be the same and detection patterns for conditions E_2C_2 and E_2C_1 should be the same.

The method used for predicting the detection pattern in the E_1C_1 condition is described in this section (as shown in Fig. 2-1 B). The method for predicting the E_2C_2 detection pattern can be obtained by interchanging the subscripts 1 and 2 in the following description. An additional analysis was also performed in which the data for the 2 conditions described above were pooled by either envelope predictor or carrier predictor.

The goal was to determine the influence of envelope and carrier specifically. A multiple-regression approach was used to incorporate the E_2C_1 and E_1C_2 detection patterns as predictors for the E_1C_1 detection pattern. It was first of interest to reduce the variability associated with the E_2C_2 pattern in the remaining detection patterns. Accordingly, the E_2C_2 detection pattern was regressed on each of three remaining patterns, and the residuals from this regression were used for subsequent analysis. The net effect was to “block” or “partial out” the variability associated with E_2C_2 . Next, several simple regressions were performed. (Note that each regression used *only* residuals from the above operation.) First two simple linear regressions were performed to predict E_1C_1 using both E_2C_1 and E_1C_2 as predictors individually; these regressions indicated the proportion of variance explained (in terms of R^2) respectively by the carrier (because C_1 was held constant) and by the envelope (because E_1 was held constant). Recall that the variability associated with E_2C_2 was “blocked.” Then, E_1C_1 was simultaneously regressed on E_2C_1 and E_1C_2 to compute the proportion of variance explained by a linear combination of both envelope and carrier. Incremental F-tests (Edwards, 1979) were performed to determine if the proportion of predicted variance in the E_1C_1 detection pattern was significantly increased by incorporating carrier in addition to envelope alone, or envelope in addition to carrier alone.

Comparisons between stimulus configurations and subjects were quantified using the square of Pearson’s correlation coefficient, which is directly comparable to the regression predictions.

2.3 RESULTS AND DISCUSSION

Several comparisons are made in the following sections: First the reliability of the data is addressed. Then, detection patterns estimated with the baseline and chimaeric stimuli are compared within subjects. Detection patterns are then compared between subjects and between interaural configurations. Finally, implications for explanatory models are discussed. An average subject (S_{avg}) was considered for all but the between-subject comparisons and the within-subject comparisons made in each of the four stimulus conditions. The average subject was created by computing $P(Y|W)$ across all subject responses.

2.3.1 Reliability of the data and detection performance

Tables 2-1 through 2-4 show reliability and detection performance statistics for the N_0S_0 (Tables 2-1 and 2-2) and N_0S_π (Tables 2-3 and 2-4) interaural configurations. The threshold-tone level where $d' \approx 1$ is given terms of E_S/N_0 . The resulting d' and β values calculated across and within the four stimulus sets are also shown. The training procedure was relatively successful in finding overall d' values near 1 with the possible exception of S1 in the N_0S_π condition (Table 2-3). No procedure was implemented to control the d' and β values within the individual stimulus sets. For N_0S_0 stimuli, d' values ranged from 0.51-1.14 and β values ranges from 0.70 to 1.32 (Table 2-1). For N_0S_π stimuli, d' values ranged from 0.54-1.11 and β values ranges from 0.57 to 1.35 (Table 2-3).

Table 2-1. Reliability and performance statistics for the N_0S_0 interaural configuration. One tone level (E_S/N_0) was used for each subject. Overall d' and β were computed using responses to waveforms in all conditions. Individual d' and β values are given for each of the 4 stimulus conditions. The coefficient of determination between responses from the first and the last half of the trials (r^2) and the proportion of predictable variance (V) are given for each condition. All r^2 values were significant ($p < 0.05$).

S	E_S/N_0	Overall		Condition	d'	β	P(Y W)	
		d'	β				r^2	$V_{P(Y W)}$
S1	10	0.87	0.93	E_1C_1	0.96	0.70	0.93	0.98
				E_2C_2	0.95	0.80	0.97	0.99
				E_1C_2	0.76	1.04	0.95	0.99
				E_2C_1	0.94	1.23	0.96	0.99
S2	10	0.88	0.99	E_1C_1	1.01	0.87	0.93	0.98
				E_2C_2	0.90	0.84	0.96	0.99
				E_1C_2	0.63	1.03	0.95	0.99
				E_2C_1	1.04	1.29	0.92	0.98
S3	10	1.02	1.07	E_1C_1	0.86	1.02	0.88	0.97
				E_2C_2	1.07	0.99	0.90	0.98
				E_1C_2	1.01	1.10	0.89	0.97
				E_2C_1	1.14	1.21	0.92	0.98
S4	11	0.96	0.95	E_1C_1	0.85	0.85	0.93	0.98
				E_2C_2	0.93	0.93	0.93	0.98
				E_1C_2	0.95	0.93	0.93	0.98
				E_2C_1	1.13	1.17	0.92	0.98
S5	11	0.86	0.99	E_1C_1	0.51	0.88	0.95	0.99
				E_2C_2	1.21	1.05	0.95	0.99
				E_1C_2	0.68	0.97	0.95	0.99
				E_2C_1	1.13	1.32	0.97	0.99
S6	11.5	0.94	0.97	E_1C_1	0.79	0.81	0.89	0.97
				E_2C_2	1.00	0.94	0.90	0.97
				E_1C_2	1.05	1.04	0.89	0.97
				E_2C_1	0.99	1.19	0.88	0.97
S_{avg}	10.58	0.92	0.98	E_1C_1	0.82	0.85	0.98	0.99
				E_2C_2	1.00	0.91	0.98	0.99
				E_1C_2	0.84	1.02	0.98	0.99
				E_2C_1	1.06	1.24	0.98	0.99

Table 2-2. Reliability statistics for $P(Y|T+N)$ and $P(Y|N)$ for the N_0S_0 stimuli. The χ^2 statistic, coefficient of determination between responses from the first and the last half of the trials (r^2), and the proportion of predictable variance (V) are shown. All χ^2 values were significant ($p < 0.001$) and all r^2 values were significant ($p < 0.05$).

S	Condition	P(Y T+N)			P(Y N)		
		χ^2	r^2	$V_{P(Y T+N)}$	χ^2	r^2	$V_{P(Y N)}$
S1	E ₁ C ₁	1371	0.91	0.98	1829	0.91	0.98
	E ₂ C ₂	1947	0.96	0.99	1720	0.96	0.99
	E ₁ C ₂	2198	0.94	0.99	2078	0.95	0.99
	E ₂ C ₁	1850	0.94	0.98	1694	0.96	0.99
S2	E ₁ C ₁	1543	0.89	0.97	1856	0.92	0.98
	E ₂ C ₂	2055	0.96	0.99	1783	0.93	0.98
	E ₁ C ₂	1737	0.94	0.98	1779	0.94	0.98
	E ₂ C ₁	1791	0.90	0.97	1557	0.89	0.97
S3	E ₁ C ₁	669	0.73	0.92	1011	0.85	0.96
	E ₂ C ₂	820	0.75	0.93	1148	0.87	0.97
	E ₁ C ₂	488	0.61	0.88	1431	0.89	0.97
	E ₂ C ₁	664	0.80	0.94	1226	0.85	0.96
S4	E ₁ C ₁	1350	0.91	0.98	1340	0.89	0.97
	E ₂ C ₂	1486	0.87	0.96	1547	0.92	0.98
	E ₁ C ₂	940	0.86	0.96	1628	0.92	0.98
	E ₂ C ₁	1176	0.83	0.95	1130	0.90	0.97
S5	E ₁ C ₁	2352	0.95	0.99	3017	0.95	0.99
	E ₂ C ₂	2402	0.95	0.99	1966	0.89	0.97
	E ₁ C ₂	1645	0.90	0.97	2310	0.96	0.99
	E ₂ C ₁	2104	0.97	0.99	2341	0.93	0.98
S6	E ₁ C ₁	1258	0.75	0.93	1645	0.93	0.98
	E ₂ C ₂	1760	0.90	0.97	1460	0.79	0.94
	E ₁ C ₂	1113	0.77	0.94	1620	0.87	0.97
	E ₂ C ₁	1561	0.79	0.94	1778	0.87	0.97
S _{avg}	E ₁ C ₁	4873	0.94	0.98	7659	0.98	0.99
	E ₂ C ₂	7798	0.97	0.99	6601	0.97	0.99
	E ₁ C ₂	3912	0.93	0.98	8530	0.98	0.99
	E ₂ C ₁	5409	0.95	0.99	6473	0.98	0.99

Table 2-3. Same as Table 2-1 but for the N_0S_π interaural configuration.

S	E_S/N_0	Overall		Condition	d'	β	P(Y W)	
		d'	β				r^2	$V_{P(Y W)}$
S1	0	0.78	0.91	E_1C_1	1.10	0.57	0.93	0.98
				E_2C_2	1.03	0.70	0.91	0.98
				E_1C_2	0.66	0.97	0.93	0.98
				E_2C_1	0.54	1.19	0.93	0.98
S2	-10	0.97	1.10	E_1C_1	0.85	1.35	0.90	0.97
				E_2C_2	0.87	1.11	0.91	0.98
				E_1C_2	1.09	0.96	0.94	0.98
				E_2C_1	1.10	0.96	0.91	0.98
S3	-17	1.01	0.99	E_1C_1	0.94	1.11	0.89	0.97
				E_2C_2	0.96	1.01	0.85	0.96
				E_1C_2	1.06	0.92	0.86	0.96
				E_2C_1	1.11	0.92	0.88	0.97
S4	-1	0.93	1.00	E_1C_1	0.92	0.85	0.90	0.98
				E_2C_2	1.03	0.94	0.95	0.99
				E_1C_2	0.79	1.03	0.93	0.98
				E_2C_1	1.02	1.25	0.93	0.98
S5	-16.5	0.91	1.02	E_1C_1	0.90	1.24	0.92	0.98
				E_2C_2	0.81	0.94	0.87	0.97
				E_1C_2	1.09	1.11	0.92	0.98
				E_2C_1	0.88	0.87	0.88	0.97
S6	-10	0.96	1.06	E_1C_1	0.87	0.95	0.89	0.97
				E_2C_2	0.81	1.22	0.84	0.96
				E_1C_2	1.08	0.96	0.87	0.97
				E_2C_1	1.11	1.12	0.88	0.97
S_{avg}	-9.08	0.92	1.01	E_1C_1	0.89	1.00	0.98	0.99
				E_2C_2	0.90	0.99	0.97	0.99
				E_1C_2	0.96	0.99	0.97	0.99
				E_2C_1	0.94	1.06	0.98	0.99

Table 2-4. Same as Table 2-2 but for the N_0S_π interaural configuration.

S	Condition	P(Y T+N)			P(Y N)		
		χ^2	r^2	$V_{P(Y T+N)}$	χ^2	r^2	$V_{P(Y N)}$
S1	E ₁ C ₁	885	0.94	0.98	1621	0.89	0.97
	E ₂ C ₂	1116	0.93	0.98	1516	0.85	0.96
	E ₁ C ₂	1859	0.94	0.98	1938	0.90	0.97
	E ₂ C ₁	1477	0.91	0.98	1445	0.93	0.98
S2	E ₁ C ₁	1283	0.89	0.97	639	0.75	0.93
	E ₂ C ₂	1383	0.88	0.97	872	0.82	0.95
	E ₁ C ₂	1188	0.91	0.98	970	0.85	0.96
	E ₂ C ₁	968	0.83	0.95	617	0.83	0.95
S3	E ₁ C ₁	844	0.81	0.95	530	0.75	0.93
	E ₂ C ₂	801	0.72	0.92	487	0.66	0.90
	E ₁ C ₂	909	0.72	0.92	366	0.63	0.88
	E ₂ C ₁	585	0.67	0.90	426	0.69	0.91
S4	E ₁ C ₁	921	0.90	0.97	1100	0.82	0.95
	E ₂ C ₂	1648	0.91	0.98	1759	0.96	0.99
	E ₁ C ₂	1390	0.91	0.98	1625	0.91	0.98
	E ₂ C ₁	1429	0.92	0.98	1343	0.84	0.96
S5	E ₁ C ₁	1388	0.89	0.97	623	0.87	0.96
	E ₂ C ₂	1117	0.82	0.95	658	0.78	0.94
	E ₁ C ₂	1220	0.88	0.97	614	0.74	0.93
	E ₂ C ₁	955	0.86	0.96	705	0.77	0.93
S6	E ₁ C ₁	1438	0.87	0.97	490	0.73	0.92
	E ₂ C ₂	1081	0.80	0.94	472	0.62	0.88
	E ₁ C ₂	1159	0.77	0.94	654	0.72	0.92
	E ₂ C ₁	1112	0.83	0.95	615	0.65	0.89
S _{avg}	E ₁ C ₁	3089	0.98	0.99	1287	0.87	0.97
	E ₂ C ₂	1470	0.92	0.98	744	0.76	0.93
	E ₁ C ₂	1953	0.91	0.98	978	0.83	0.95
	E ₂ C ₁	2595	0.96	0.99	1072	0.83	0.95

Tables 2-1 through 2-4 also include first-half, last-half correlations and proportions of predictable variance (Ahumada, 1971), both of which indicate the reliability of the data. These statistics are shown for cases when T+N and N were grouped together [P(Y|W), Tables 2-1 and 2-3] or considered separately [P(Y|T+N) and P(Y|N), Tables 2-2 and 2-4]. Note that Tables 2-2 and 2-4 also include χ^2 statistics, with larger values indicating more reliable detection patterns (for a more complete description see the Appendix, Sec. A.3.1.) In general, more variance was predictable for P(Y|W) because a larger number of waveforms was considered. Note that the vast majority of the variance was predictable in the N_0S_0 condition (>95 percent in most cases) for all sets of stimuli. Less variance was predictable in the N_0S_π condition than the N_0S_0 condition, but this amount always exceeded about 90 percent. In general, these results indicate that all of the estimated detection patterns were reliable.

Note the relatively low V and χ^2 values observed for S_{avg} in Table 2-4. Despite the fact that more trials were included when calculating these statistics as compared to those for individual subjects, the χ^2 values remained lower than some of those computed for individual subjects for P(Y|N). This indicated that when pooling responses across subjects, the detection patterns actually became less variable (i.e., flattened), which was consistent with the individual listeners using different strategies for this task.

2.3.2 Within-subject comparisons of detection patterns estimated with baseline and chimaeric stimuli

Before applying the analysis procedure described in the methods section, the detection patterns were checked for normality using the Lilliefors hypothesis test of composite normality (Sheskin, 2000), keeping the individual-test alpha level at 0.05. No family-wise error-rate correction was implemented in order to maintain a conservative test criterion. Only 2 of the 144 detection patterns [considering P(Y|W), P(Y|T+N), and P(Y|N) separately for the 4 stimulus sets in the 2 interaural configurations with 6 subjects] proved to be non-normal. Similarly, for all regression analyses, residuals were examined using the same test. Of the 324 regressions performed [3 predictors (E_1C_1 , E_2C_2 , and combined) x 3 detection-pattern components x 6 subjects x 3 predictor models (envelope carrier or both) x 2 interaural configurations], only 10 showed significantly ($p < 0.05$) non-normal residuals. Examination of residual plots failed to find any serious issues of heteroscedasticity (unequal error variances). Correlations between predictor variables in the same analysis were computed to check for multicollinearity (high correlation of predictor variables). Typical values for the r^2 between predictor variables ranged from 0 to 0.1 (and were insignificant) and in no case did the value exceed 0.31, indicating that the data did not exhibit a large degree of multicollinearity.

Figures 2-2 through 2-7 show scatter plots of z -scores computed for the various cue conditions. Probabilities of 0 and 1 (the z -scores of which are unbounded) were replaced with 1/100 and 99/100 respectively, which occurred for only 54 of the 2400

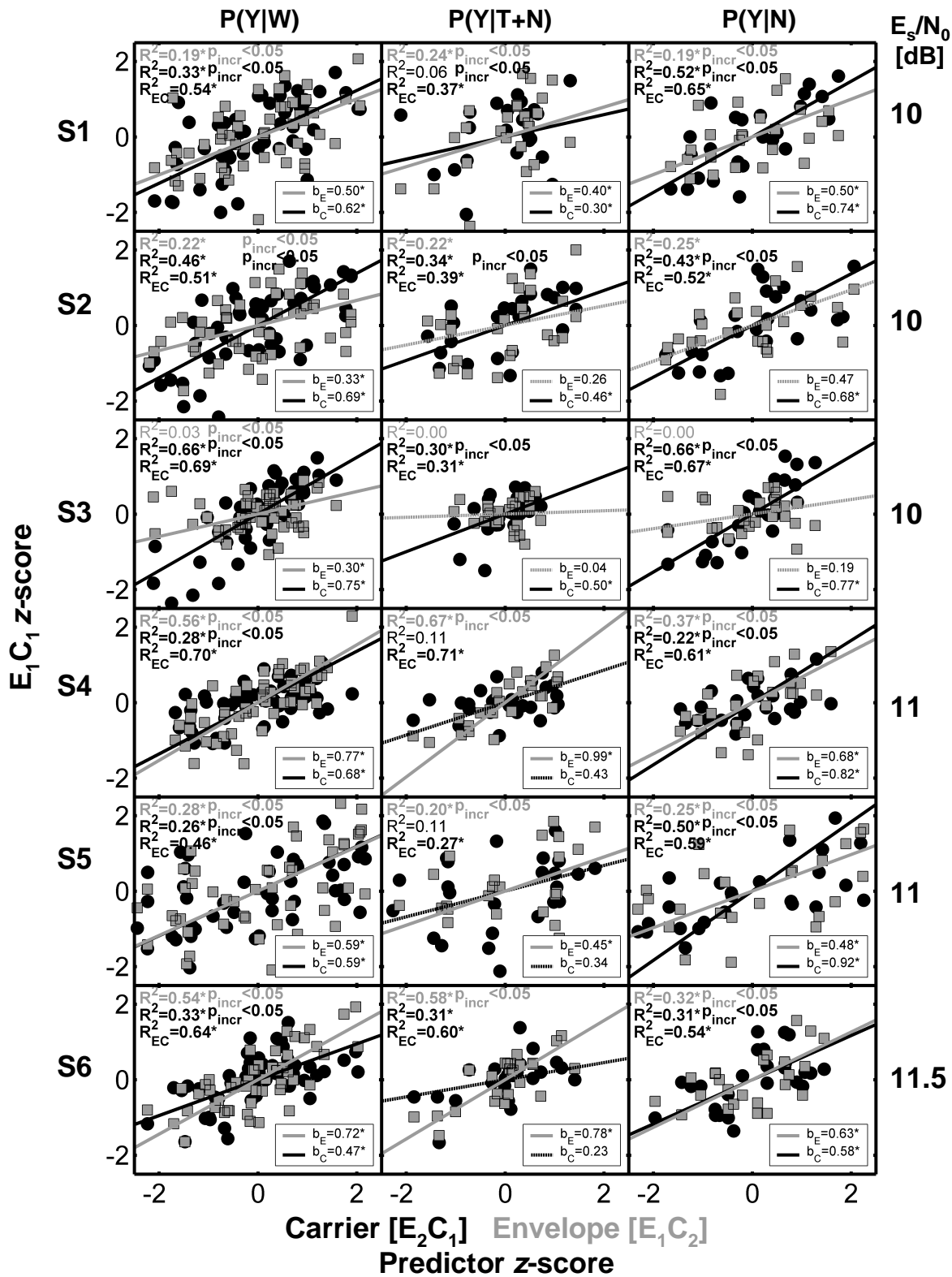


Figure 2-2. Comparisons between cue conditions for $P(Y|W)$, $P(Y|T+N)$ and $P(Y|N)$ (columns) for 6 subjects (rows). Envelope-based predictions are shown in grey (squares), while carrier-based predictions are shown in black (circles). Relative weights are shown by the b values. R^2_{EC} corresponds to the proportion of predictable variance using a linear combination of both envelope and carrier. Significant ($p < 0.05$) incremental F-test results are shown for envelope and/or carrier. Signal-to-noise ratio in dB (E_s/N_0) is shown to the right of each plot.

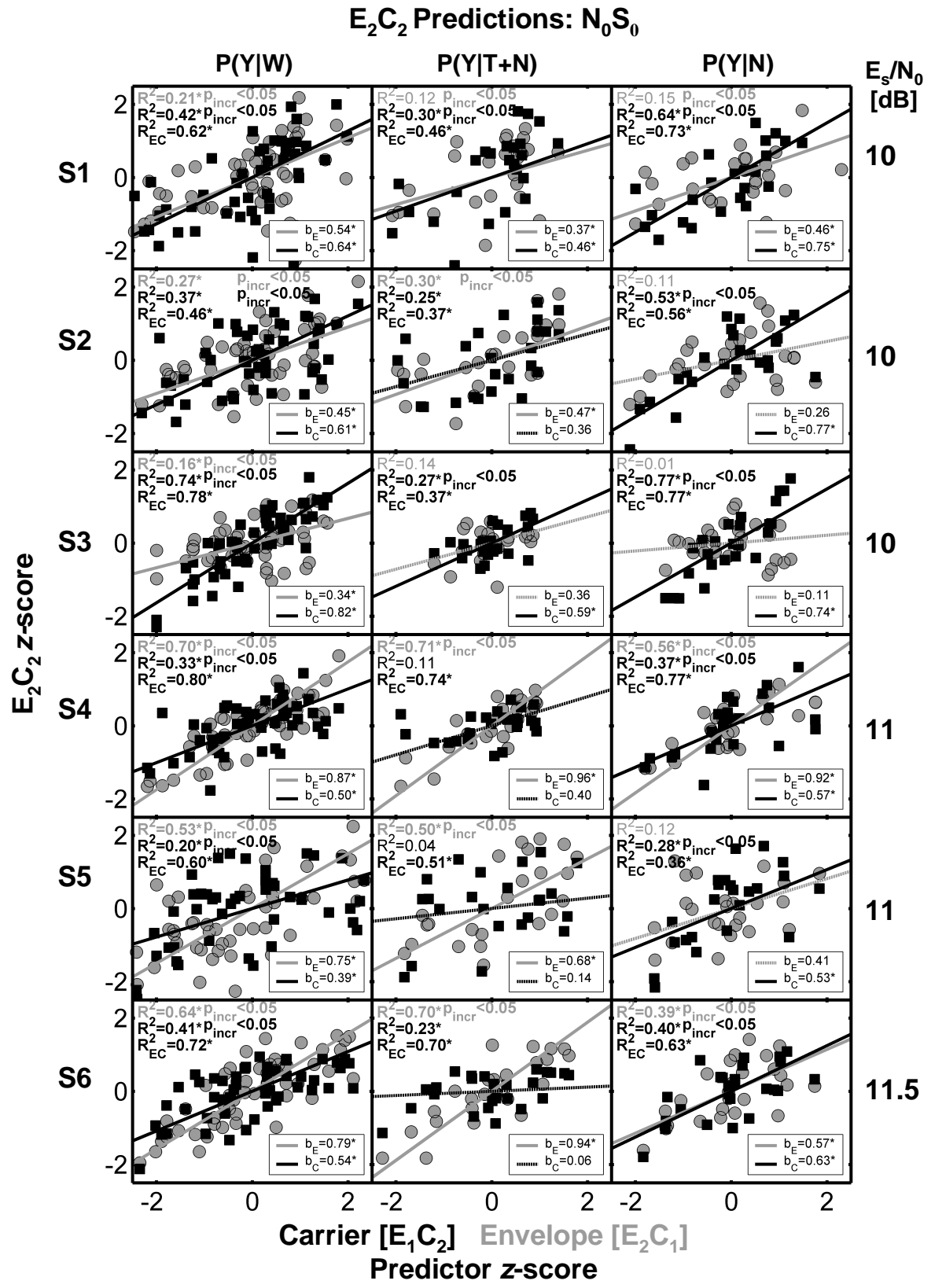


Figure 2-3. Same as Fig. 2-2 except predictions were made for the E_2C_2 stimulus condition.

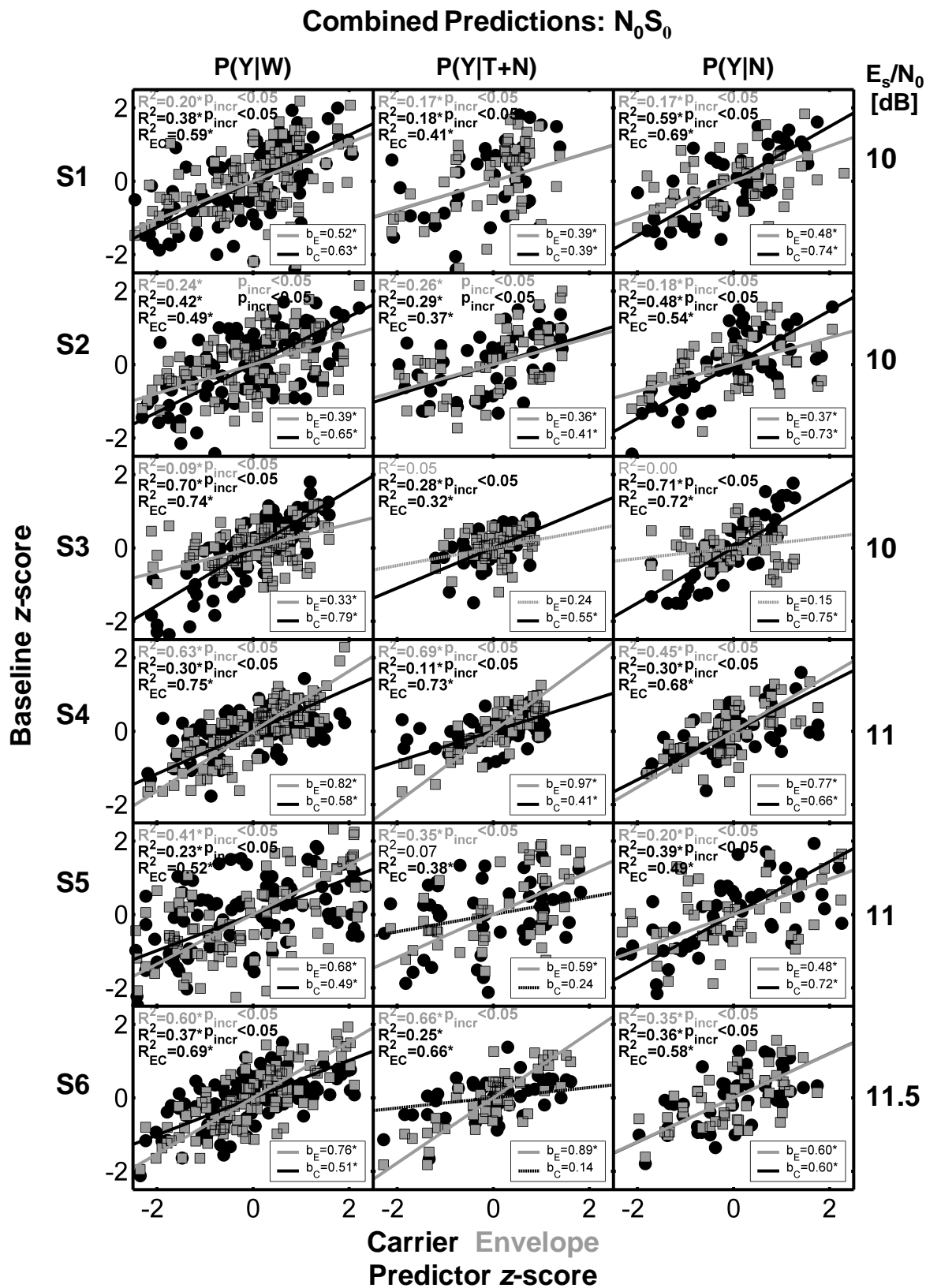


Figure 2-4. Combined predictions for E_1C_1 and E_2C_2 by predictor type: envelope or carrier. Predictions are for the N_0S_0 condition. See text for details.

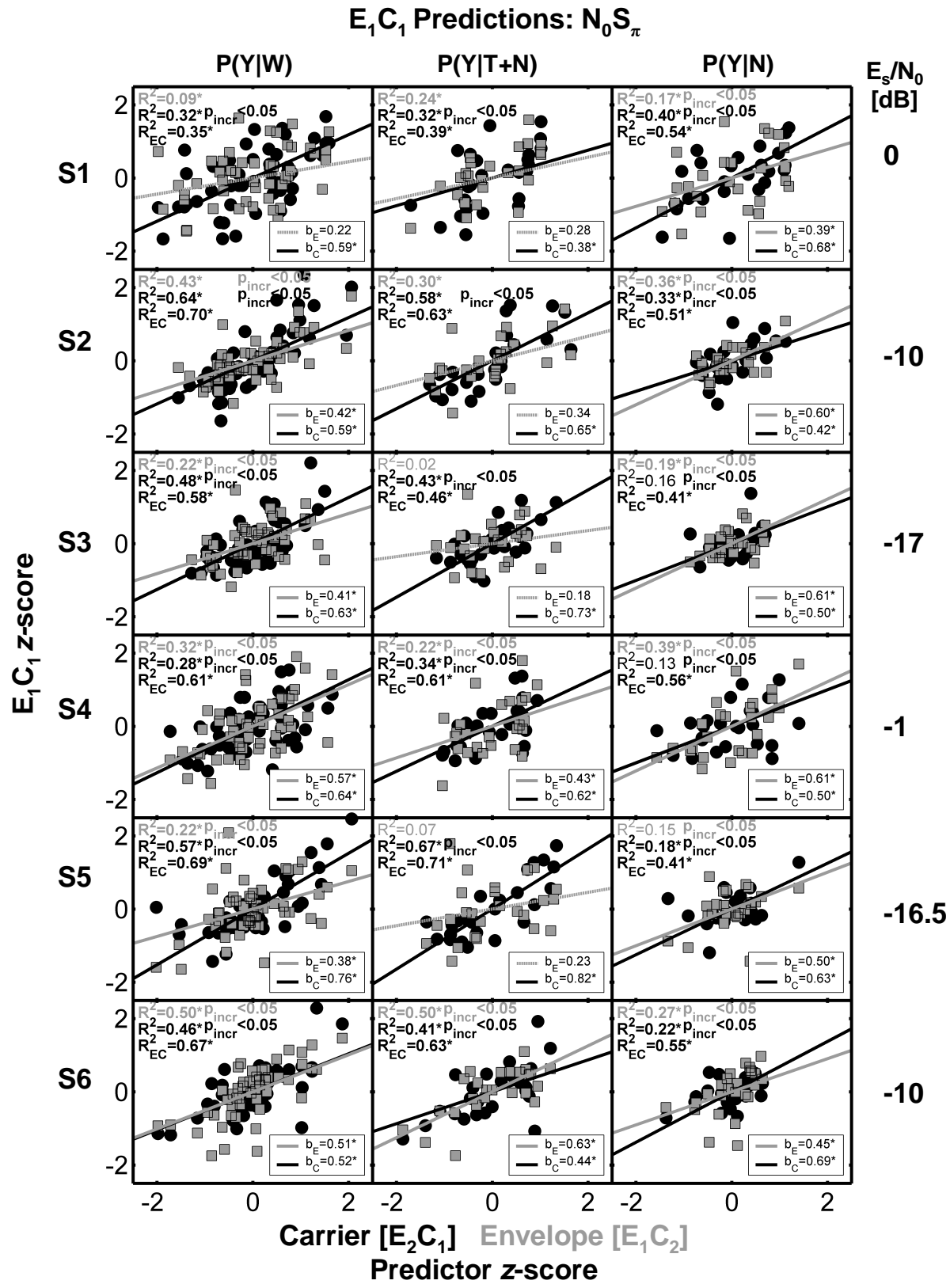


Figure 2-5. Same as Fig. 2-2 except N_0S_π stimuli were used.

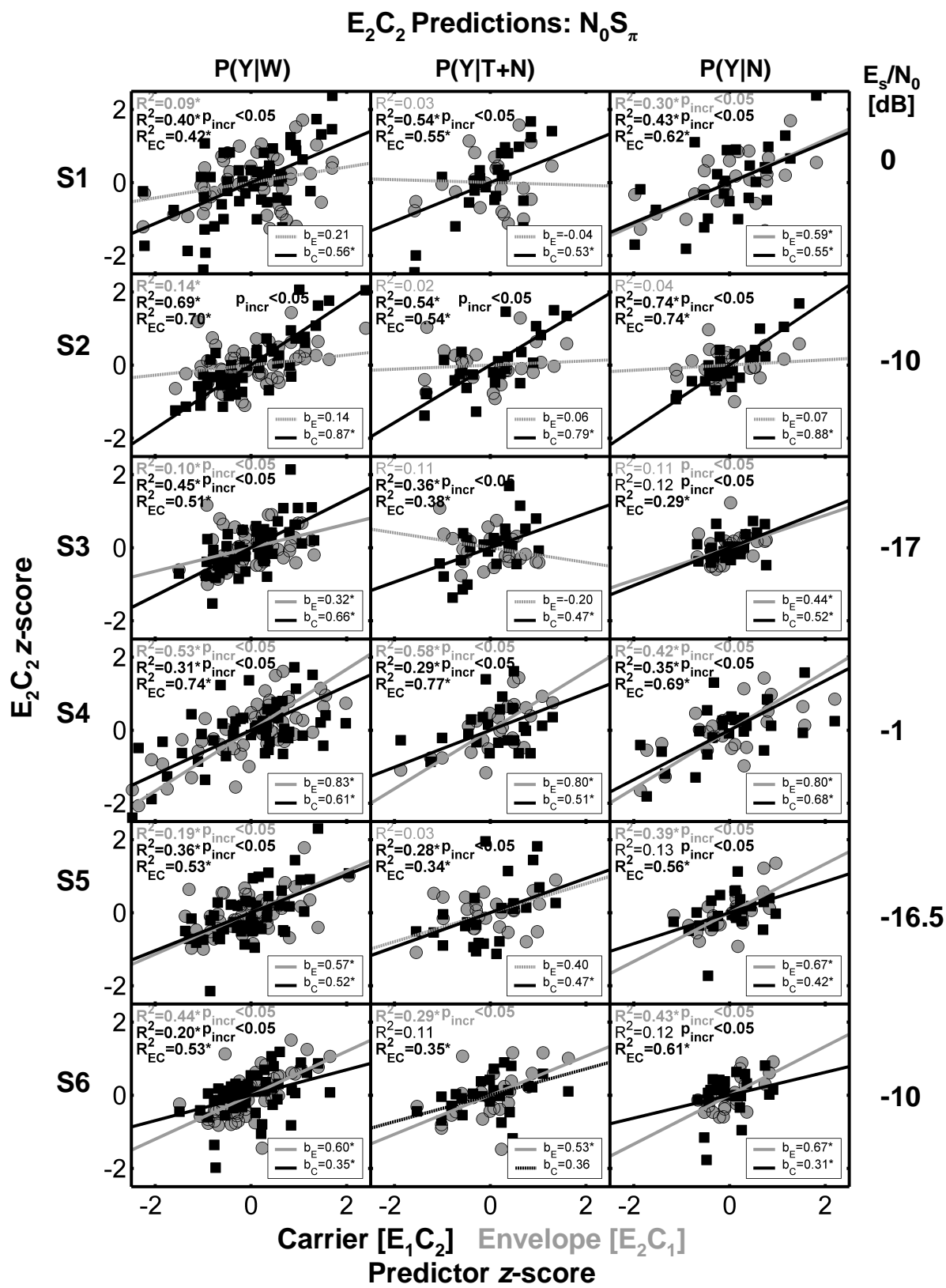


Figure 2-6. Same as Fig. 2-3 except predictions were made for the E_2C_2 stimulus condition.

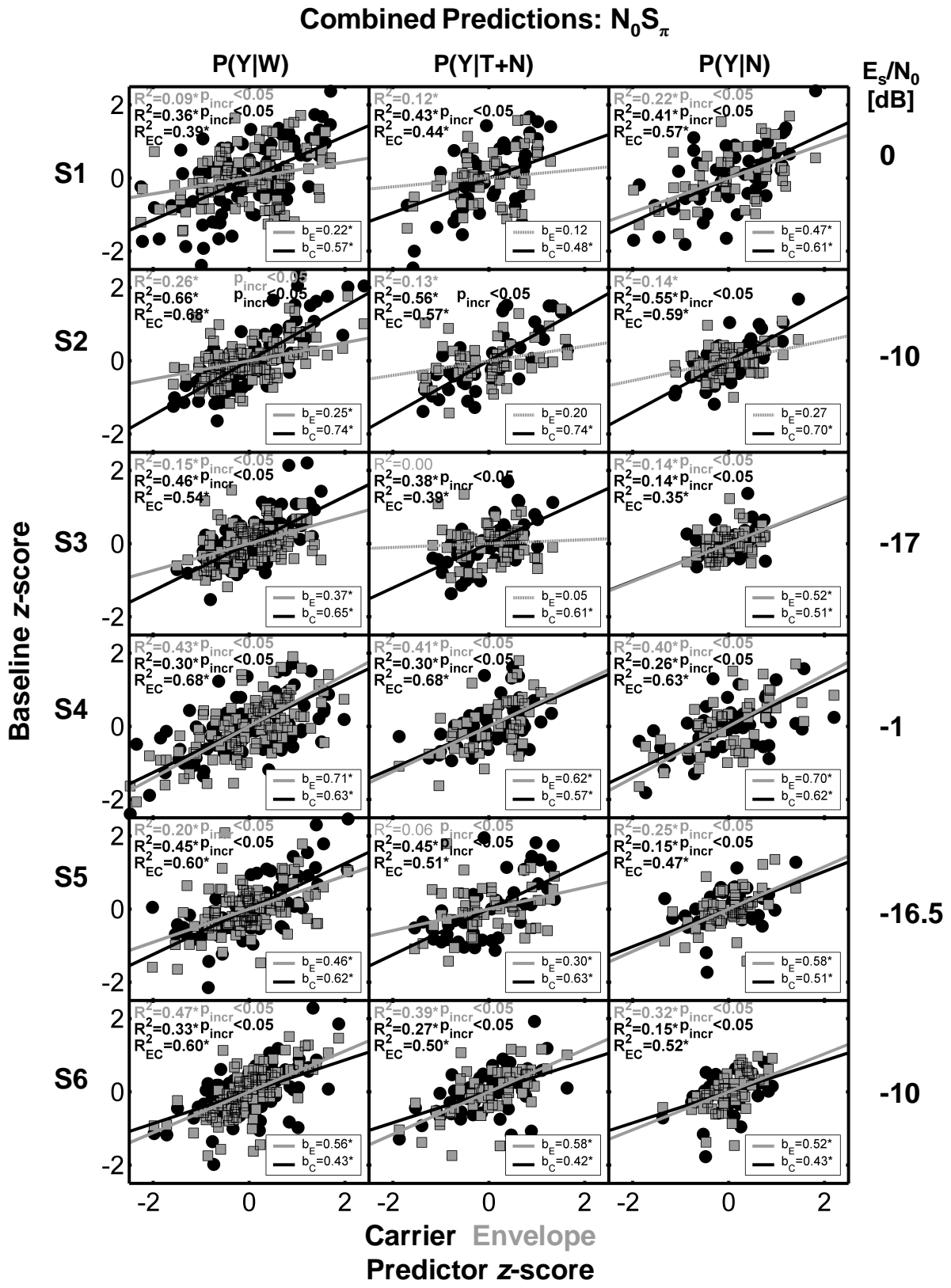


Figure 2-7. Same as Fig. 2-4 except predictions were made for the N_0S_π condition.

probabilities in the stimulus set [50 P(Y|W) x 2 interaural configurations x 4 stimulus conditions x 6 subjects]. In each panel, the detection patterns estimated from responses to chimaeric stimuli were used to predict the detection patterns estimated in the E₁C₁ or E₂C₂ conditions. The predictors are plotted on the abscissa of each panel. Envelope predictions are always shown in grey squares and carrier predictions are always shown in black circles. Regression slopes (b_E and b_C) are shown in each panel. The slope values were computed using the multiple regression procedure (i.e., both envelope and carriers were predictors) and thus are slightly different from the slopes that would be obtained using either envelope or carrier individually (as discussed below). If the carrier were a perfect predictor of the variance in the detection patterns, the black circles would fall exactly along the diagonal and b_C would equal one. Conversely, if the envelope were a perfect predictor of variance in the detection patterns, the gray squares would fall exactly along the diagonal and b_E would equal 1.

Three R^2 values are shown in each panel (Figs. 2-2 through 2-7). The upper grey R^2 corresponds to a prediction using only envelope; the middle black R^2 corresponds to a prediction using only carrier. The lower-black R^2_{EC} corresponds to a linear combination of envelope and carrier using the weights given by b_E and b_C respectively. Slopes for the individual envelope and carrier predictions (corresponding to the individual envelope and carrier R^2 values) are not shown. Significant R^2 values are in bold and denoted with an asterisk. Results from incremental-F tests (p_{incr}) indicate if the addition of envelope to a prediction based on carrier, or the addition of carrier to a prediction based on envelope, significantly increased the amount of predictable variance in the detection pattern plotted along the ordinate. (Note that the incremental-F test is equivalent to testing the significance of b_E or b_C). Recall that the variance linearly associated with the non-predictor condition was blocked in these predictions, as described above.

2.3.2.1 N₀S₀ stimuli

Figures 2-2 through 2-4 show results for the regression analyses described above for N₀S₀ stimuli. Chimaeric detection patterns were used to predict baseline detection patterns for each individual subject. Figures 2-2 and 2-3 show separate predictions for detection patterns estimated with E₁C₁ and E₂C₂ stimuli, respectively. It was of interest to determine if the predictions for the two baseline conditions were in agreement. To do so, the test of significant differences between correlated but non-overlapping correlations (Raghuathan *et al.*, 1996) was conducted for each combination of envelope and carrier (2), each subject (6), and for P(Y|W), P(Y|T+N), and P(Y|N), for a total of 36 tests. As with the tests of normality, we were interested in NOT rejecting the null hypothesis. To reduce the chance of type-II error, a family-wise error alpha level was not computed, and the individual alpha level for each test was maintained at 0.05. None of the 36 tests yielded significant differences ($p > 0.05$) between predictions for E₁C₁ (Fig. 2-2) and E₂C₂ (Fig. 2-3) for either envelope or carrier. Thus the data from Figs. 2 and 3 were combined for Fig. 2-4. The data from Figs. 2-2 and 2-3 were concatenated by predictor (either envelope or carrier), and data from the two baseline patterns were also concatenated. Thus, individual panels in Fig. 2-4 contain twice the number of data points as in Figs. 2-2 or 3. The weights and R^2 values in Fig. 2-4 are essentially the average of those in Figs. 2-2 and 2-3 for corresponding panels. One notable difference when combining predictions was the additional significant p_{incr} value for S4, P(Y|T+N) in Fig. 2-4, which was most likely an effect of doubling the number of data points.

Inspection of Fig. 2-4 reveals that both envelope and carrier were, in general, positively correlated to the each of the individual listener's decision variables (for all but four cases). The large number of significant p_{incr} values indicates that for most subjects, both envelope and carrier

contributed unique information that was correlated to the listeners' decision variables. However, there was some intersubject variability in the R^2 values observed in Fig. 2-4. In previous N_0S_0 detection experiments in which energy was not equalized, subjects' detection patterns were highly correlated to one another (e.g., Evilsizer *et al.*, 2002 and Davidson *et al.*, 2006). These high correlations indicate that the same or very similar decision variables were used by each subject in those studies. Recall that in the N_0S_0 condition of this experiment, overall stimulus levels were equalized to remove the availability of energy as decision variable. As a result, high intersubject correlations were not necessarily expected (for a complete discussion of intersubject correlations, see below), nor was the use of identical decision variables across subjects. In fact, the results shown in Figs. 2-2 through 2-4 (in addition to the relatively low intersubject correlations described in Sec. 2.3.3) suggest the use of different detection strategies by different subjects. The b values and R^2 values for subject 3, suggest a preference for cues related to the fine structure of the stimulus waveforms rather than envelopes of the stimulus waveforms. The remaining subjects used a combination of carrier and envelope-related cues, as indicated by the b and R^2 values for envelope predictors with respect to carrier predictors.

Although the majority of R^2 values in Figs. 2-2 through 2-4 are significant, none are above 0.80, and all are lower than the estimates of the proportions of predictable variance shown in Tables 2-1 and 2-2. Thus, substantially more variance in the detection patterns should be predictable. The scatter plots in Figs. 2-2 through 2-4 do not suggest a comparable nonlinear statistical model capable of explaining significantly more variance than the current linear model. Nevertheless, based on the overall pattern of results, it was concluded that the linear model was not necessarily a good fit for these data. This suggests that the separation of carrier and fine structure for the purposes of computing individual detection statistics (i.e., a statistic for carrier and a statistic for fine structure), and the subsequent recombination of those individual detection statistics to compute model decision variables is not a solution that will lead to acceptable model predictions for the N_0S_0 interaural configuration. One could argue that the method of separation of envelope and fine structure (e.g., the Hilbert transform) was simply inappropriate, or that the chimaeric stimulus waveforms contained interference (e.g., frequency splatter) that lowered R^2 values. However, each of these arguments also suggests that separating the stimulus waveform into envelope and fine structure is not a viable strategy for modeling the subjects' detection patterns. For example, if separate temporal processing of envelope and fine structure occurred, frequency splatter should not affect the listeners' detection patterns. In review, the results illustrated in Figs. 2-2 through 2-4 support the notion that detection models should compute detection statistics from the entire stimulus waveform, rather than separating envelope and fine structure completely. (A possible exception could be if the model computed statistics based on envelope and fine structure as a function of time, and allowed those statistics to interact temporally before computing a final decision variable. Such a computation is evaluated in Ch.3).

2.3.2.2 N_0S_π stimuli

As in the N_0S_0 interaural configuration, it was of interest to determine if the predictions for the two baseline conditions differed (E_1C_1 and E_2C_2). Several tests of correlated but non-overlapping correlations (Raghunathan *et al.*, 1996) were conducted on the corresponding R^2 values shown in Figs. 2-5 through 2-6. As with the N_0S_0 data, none of the 36 tests indicated significant ($p > 0.05$) differences between the correlations in Figs. 2-5 and 2-6.

Before moving on to the combined N_0S_π predictions in Fig. 2-7, the reader is reminded of the rather large threshold difference between subjects (some on the order of 17 dB) for this

interaural configuration. Also recall that N_0S_π noise-alone stimuli are diotic, while $N_0S_\pi T+N$ stimuli are dichotic.

Subjects 3 and 5 had the lowest threshold tone levels and showed similar trends in terms of envelope and carrier predictions (see Fig. 2-7). The linear combination of envelope and carrier failed to predict the majority of the variance using only $P(Y|N)$. Predictions using only $P(Y|T+N)$ indicated a stronger reliance on carrier, but they failed to predict more than approximately half the variance in the baseline detection patterns. Predictions for $P(Y|W)$ indicate that carrier dominated the detection process for these two subjects, but also showed a significant contribution of cues based on the signal envelope. The linear model predicted just over half the variance in the baseline detection patterns for these subjects. Slightly larger weights were fit for cues derived from the waveform carrier.

Subjects 2 and 6 were tested with threshold tone levels about 7 dB higher than subjects 3 and 5. Subject 2 showed consistent dominance of carrier-based cues over envelope-based cues. The linear model explained about 70 percent of the variance in the complete detection pattern [$P(Y|W)$] for this subject. Results for subject 6 indicated a stronger contribution of envelope over carrier with significant incremental F tests for both envelope and carrier. The linear model explained about 65 percent of the variance in the complete detection pattern for this subject.

Subjects 1 and 4 were tested with the highest threshold tone levels. Subject 1 weighted cues derived from the carrier more strongly than those derived from the envelope, but predictions using $P(Y|W)$ explained less than half the variance in the baseline detection patterns. Subject 4 used cues derived from both envelope and carrier, and the linear model was able to up to 77 percent of the variance in the baseline detection patterns.

In general, the results for the N_0S_π interaural configuration, as for the N_0S_0 interaural configuration, indicated that complex interactions occurred between signal envelope and carrier that were not captured by the linear statistical model. The model seemed to fit best for the subjects with higher thresholds, but in general predicted about 40 to 80 percent of the variance in the baseline detection patterns.

2.3.3 Comparisons between subjects

Tables 2-5 and 2-6 show intersubject r^2 values for the baseline and chimaeric detection patterns. The intersubject r^2 values were lower for the N_0S_0 intersubject configuration for this study than in previous studies (Evilsizer *et al.*, 2002; Davidson *et al.*, 2006) and ranged from 0.21 to 0.79 for $P(Y|W)$. The lower intersubject correlations suggest the use of a more diverse set of decision variables across subjects in this experiment than in previous experiments with diotic stimuli, which was likely caused by the lack of a simple energy cue and the small stimulus bandwidth. Pairs of subjects with the highest intersubject r^2 values did not necessarily share the same predictions for envelope or carrier dominance (e.g., S2 and S6 in Fig. 2-4. and Tables 2-5 and 2-6). However, given the lower intersubject correlations observed in this study as compared to previous studies (e.g., Davidson *et al.*, 2006 or Evilsizer *et al.* 2002), a strong trend of cue dominance across subjects and intersubject correlations is not necessarily implied. The lack of such a trend may be interpreted (cautiously) as indicating that separate processing of envelope and fine structure is not an ideal way of accounting for the detection patterns.

Table 2-5. Comparisons between subjects [P(Y|W)] presented in terms of r^2 .

Interaural configuration	Intersubject Comparison	P(Y W)			
		E ₁ C ₁	E ₂ C ₂	E ₁ C ₂	E ₂ C ₁
N ₀ S ₀	S1-S2	0.38*	0.57*	0.49*	0.34*
	S1-S3	0.59*	0.56*	0.45*	0.51*
	S1-S4	0.68*	0.77*	0.56*	0.54*
	S1-S5	0.36*	0.50*	0.21*	0.27*
	S1-S6	0.47*	0.62*	0.43*	0.42*
	S2-S3	0.40*	0.53*	0.40*	0.61*
	S2-S4	0.51*	0.62*	0.54*	0.60*
	S2-S5	0.59*	0.71*	0.57*	0.63*
	S2-S6	0.70*	0.71*	0.68*	0.65*
	S3-S4	0.38*	0.49*	0.49*	0.60*
	S3-S5	0.44*	0.42*	0.30*	0.46*
	S3-S6	0.40*	0.51*	0.51*	0.63*
	S4-S5	0.32*	0.58*	0.40*	0.41*
	S4-S6	0.54*	0.75*	0.67*	0.63*
	S5-S6	0.66*	0.79*	0.60*	0.65*
N ₀ S _π	S1-S2	0.05	0.00	0.00	0.01
	S1-S3	0.40*	0.18*	0.06	0.25*
	S1-S4	0.73*	0.78*	0.62*	0.45*
	S1-S5	0.16*	0.09*	0.04	0.10*
	S1-S6	0.31*	0.02	0.05	0.08*
	S2-S3	0.27*	0.28*	0.49*	0.57*
	S2-S4	0.02	0.00	0.00	0.14*
	S2-S5	0.55*	0.51*	0.60*	0.52*
	S2-S6	0.52*	0.69*	0.58*	0.52*
	S3-S4	0.34*	0.11*	0.09*	0.41*
	S3-S5	0.52*	0.27*	0.57*	0.50*
	S3-S6	0.61*	0.39*	0.53*	0.57*
	S4-S5	0.09*	0.03	0.03	0.21*
	S4-S6	0.15*	0.00	0.06	0.12*
	S5-S6	0.64*	0.56*	0.62*	0.44*

* p < 0.05

Table 2-6. Comparisons between subjects [P(Y|T+N) and P(Y|N)] presented in terms of r^2 .

Interaural configuration	Intersubject Comparison	P(Y T+N)				P(Y N)			
		E ₁ C ₁	E ₂ C ₂	E ₁ C ₂	E ₂ C ₁	E ₁ C ₁	E ₂ C ₂	E ₁ C ₂	E ₂ C ₁
N ₀ S ₀	S1-S2	0.17*	0.38*	0.29*	0.16*	0.20*	0.51*	0.55*	0.15
	S1-S3	0.14	0.49*	0.14	0.24*	0.64*	0.32*	0.56*	0.42*
	S1-S4	0.39*	0.75*	0.18*	0.32*	0.69*	0.62*	0.81*	0.40*
	S1-S5	0.17*	0.32*	0.01	0.06	0.44*	0.34*	0.34*	0.16*
	S1-S6	0.32*	0.49*	0.17*	0.14	0.33*	0.47*	0.47*	0.41*
	S2-S3	0.12	0.39*	0.13	0.49*	0.21*	0.34*	0.47*	0.36*
	S2-S4	0.24*	0.48*	0.15	0.37*	0.42*	0.50*	0.84*	0.43*
	S2-S5	0.69*	0.70*	0.35*	0.43*	0.56*	0.52*	0.65*	0.57*
	S2-S6	0.59*	0.77*	0.55*	0.51*	0.61*	0.47*	0.77*	0.51*
	S3-S4	0.01	0.44*	0.01	0.28*	0.40*	0.16*	0.56*	0.30*
	S3-S5	0.35*	0.29*	0.08	0.12	0.53*	0.07	0.23*	0.34*
	S3-S6	0.08	0.55*	0.07	0.52*	0.34*	0.11	0.37*	0.41*
	S4-S5	0.08	0.39*	0.04	0.15	0.50*	0.45*	0.54*	0.19*
	S4-S6	0.30*	0.66*	0.25*	0.57*	0.47*	0.62*	0.69*	0.33*
	S5-S6	0.51*	0.65*	0.30*	0.31*	0.76*	0.74*	0.77*	0.78*
N ₀ S _π	S1-S2	0.05	0.00	0.00	0.01	0.33*	0.41*	0.37*	0.36*
	S1-S3	0.40*	0.18*	0.06	0.25*	0.04	0.04	0.01	0.02
	S1-S4	0.73*	0.78*	0.62*	0.45*	0.64*	0.61*	0.78*	0.39*
	S1-S5	0.16*	0.09*	0.04	0.10*	0.04	0.00	0.04	0.01
	S1-S6	0.41*	0.03	0.18*	0.13*	0.17*	0.1	0.17*	0.01
	S2-S3	0.27*	0.28*	0.49*	0.57*	0.00	0.01	0.01	0.01
	S2-S4	0.02	0.00	0.00	0.14*	0.38*	0.40*	0.48*	0.13
	S2-S5	0.55*	0.51*	0.60*	0.52*	0.17*	0.05	0.18*	0.03
	S2-S6	0.42*	0.59*	0.28*	0.48*	0.04	0.17*	0.11	0.00
	S3-S4	0.34*	0.11*	0.09*	0.41*	0.00	0.09	0.07	0.01
	S3-S5	0.52*	0.27*	0.57*	0.50*	0.15	0.01	0.01	0.07
	S3-S6	0.60*	0.46*	0.44*	0.61*	0.19*	0.15	0.01	0.12
	S4-S5	0.09*	0.03	0.03	0.21*	0.13	0.06	0.16	0.00
	S4-S6	0.23*	0.00	0.25*	0.16*	0.02	0.08	0.13	0.01
	S5-S6	0.59*	0.39*	0.34*	0.40*	0.09	0.06	0.02	0.00

* p < 0.05

Under the N_0S_π conditions, intersubject r^2 values were on average lower than those for N_0S_0 conditions, and ranged from 0.00 to 0.78 for $P(Y|W)$. Subjects with similar threshold tone levels had more similar detection patterns. Subjects 1 and 4, 2 and 6, and 3 and 5 had the highest intersubject correlations, which were significant ($p < 0.05$) for all conditions for $P(Y|W)$, $P(Y|T+N)$ and $P(Y|N)$. These subjects also had the closest thresholds, showing a likely dependence of threshold on strategy. Comparing Fig. 2-7 and Tables 2-5 and 2-6 for the pairs of subjects with the largest intersubject correlations (and the closest thresholds) did not reveal any clear trend of envelope or carrier dominance, suggesting once more that the linear combination of envelope and carrier does perform well predicting the detection patterns.

2.3.4 Comparisons between interaural configurations

Tables 2-7 and 2-8 show correlations (in terms of r^2) between $P(Y|W)$, and $P(Y|T+N)$ and $P(Y|N)$ respectively. The subjects with the highest thresholds (S1 and S4) had the highest correlations between detection patterns from the two interaural configurations. Closer inspection of Table 2-8 reveals that the sources of the correlations between the two interaural configurations were entirely from responses to noise-alone stimuli, $P(Y|N)$. Subjects 1 and 4 show dramatically high r^2 values between $P(Y|N)$ values from the two interaural configurations, ranging from 0.90 to 0.95. Recall that noise-alone stimuli in the N_0S_π condition are the same as those from the N_0S_0 condition.

Table 2-7. Comparisons between interaural configurations [$P(Y|W)$] presented in terms of r^2 .

Subject	P(Y W)			
	E_1C_1	E_2C_2	E_1C_2	E_2C_1
S1	0.74*	0.69*	0.50*	0.43*
S2	0.07	0.01	0.00	0.15*
S3	0.27*	0.30*	0.27*	0.45*
S4	0.67*	0.72*	0.63*	0.54*
S5	0.00	0.06	0.02	0.09*
S6	0.08	0.03	0.29	0.18
S_{avg}	0.39*	0.47*	0.44*	0.49*

* $p < 0.05$

Table 2-8. Comparisons between interaural configurations [$P(Y|T+N)$ and $P(Y|N)$] presented in terms of r^2 .

Subject	P(Y T+N)				P(Y N)			
	E_1C_1	E_2C_2	E_1C_2	E_2C_1	E_1C_1	E_2C_2	E_1C_2	E_2C_1
S1	0.10	0.18*	0.12	0.10	0.95*	0.95*	0.90*	0.93*
S2	0.06	0.00	0.03	0.07	0.47*	0.48*	0.59*	0.42*
S3	0.01	0.02	0.02	0.05	0.10	0.06	0.00	0.01
S4	0.18*	0.29*	0.13	0.05	0.91*	0.95*	0.93*	0.93*
S5	0.00	0.00	0.00	0.02	0.29*	0.29*	0.35*	0.22*
S6	0.00	0.09	0.00	0.00	0.00	0.10	0.08	0.00
S_{avg}	0.04	0.04	0.01	0.05	0.26*	0.25*	0.37*	0.15

* $p < 0.05$

Such high r^2 values suggest that S1 and S4 were attempting to use the same detection strategy for the two interaural configurations, explaining the high thresholds for the N_0S_π stimuli. The subjects with the lowest thresholds (S3 and S5) and intermediate thresholds (S2 and S6) had much lower correlations between detection patterns from the two interaural configurations. The noise-alone intersubject r^2 values have implications for the types of detection models used to explain the detection patterns, which will be outlined in the following section.

2.3.5. Implications for computational modeling

2.3.5.1 Comparisons between detection patterns estimated with baseline and chimaeric stimuli

Although results varied across listeners for both N_0S_0 and N_0S_π stimuli, several important modeling implications are embedded within the data. The first pertains to the separation of envelope and carrier. Part of the motivation for this study was to collapse the entire detection process into decision variables based on envelope and carrier without concern for temporal interactions of envelope and carrier. The goal was to quantify how well the statistical models explain the baseline detection patterns. Large proportions of the predictable variance remain unexplained by the linear statistical model, which implies that understanding the temporal interaction of stimulus envelopes with stimulus carriers is paramount for modeling these data. Moreover, previous studies attempting to explain detection patterns with computational models have omitted peripheral filtering and nonlinearities under the assumption that these do not contribute to the detection process (e.g., Isabelle, 1995; Davidson *et al.*, 2006). The results of the present study suggest that such dynamic interactions may be needed to explain the detection process. Further, the results of this study also suggest that extraction of the complex analytic signal should not be used to separately process envelope and fine-structure unless some interaction between the two occurs before decision variables are computed. The fact that corresponding $P(Y|W)$ values estimated from sets of stimuli with either the same envelopes or the same carriers differed suggests a form of interaction between the two that has not previously been employed for the purposes of modeling data collected with reproducible maskers.

Several candidate models remain in contention for both diotic and dichotic signal detection, and each will be tested in future studies. These models are worth briefly mentioning here. In general, each incorporates some sort of dynamic interaction of envelope and carrier, and each computes the decision variable from the entire stimulus waveform (rather than stripping the stimulus envelope or fine structure apart for separate analyses). An example of a diotic model that remains under consideration is the multiple-detector model (e.g., Gilkey and Robinson, 1986), which uses monaural banks of filters that are weighted and combined linearly to produce a decision variable. This model may provide sufficient interaction of envelope and carrier at the peripheral processing stages.

With respect to binaural models, equalization-cancellation-style models with realistic peripheral processing stages (e.g., Breebaart *et al.*, 2001a) should remain under consideration. Cross-correlation-style models (e.g., Colburn, 1977) with realistic peripheral processing should also remain under consideration, given that these models operate on the entire stimulus waveform, rather than on envelope or fine structure alone.

2.3.5.2 Comparisons between interaural configurations

The fact that reliable detection patterns are obtained for N_0S_π noise-alone stimuli (e.g., Evilsizer *et al.*, 2002; Isabelle, 1995; Siegel and Colburn, 1989) has at least two implications worth discussing. First: The intersubject variability in detection patterns is not likely to be caused by internal noise processes. As the number of trials of each reproducible stimulus is

increased, the internal noise present in the response probability for that waveform is averaged out. Such averaging should tend to always increase intersubject correlations with increased numbers of trials, which is not observed experimentally. Second: If independent internal noise processes dominate over external noise at each ear (additive noise), any symmetrical binaural processing would not result in a stable detection pattern. A multiplicative internal noise source may produce a stable pattern if no normalization or cancellation occurs during binaural processing. The response on each trial would simply be based on interaural differences that result from the internal noise processes. Over large numbers of trials, such noise-generated interaural differences would produce “flat” detection patterns with no reliable differences in detection probabilities from noise to noise. One mechanism that would generate reliable detection patterns for noise-alone stimuli is a static frequency mismatch, or a static interaural delay or attenuation. Such a mechanism would be stable over time, and would generate a specific detection pattern based on processing asymmetry. The magnitudes and types of plausible processing asymmetries will be examined in future work.

2.4 SUMMARY AND FUTURE DIRECTIONS

This experiment investigated the roles of cues based on stimulus envelope and carrier for tone-in-noise detection. A simple linear model was unable to explain all of the predictable variance in the detection patterns, yielding R^2 values between 0.31 and 0.80 for N_0S_0 stimuli and between 0.53 and 0.77 for N_0S_π stimuli. No clear trends in the data indicated that higher-order interactions would yield better predictions. Overall, the pattern of correlations between detection patterns estimated with chimaeric and baseline stimuli suggested a temporal interaction of signal envelope and carrier. The decision variable was most likely a result of this interaction, rather than a result of separately processing signal envelope and fine structure, computing decision variables, and subsequently recombining decision variables derived from signal envelope and carrier.

These results were consistent with the findings of van de Par and Kohlrausch (1998) that showed no remarkable difference in thresholds when ITDs or ILDs were eliminated in an N_0S_π detection task. In future studies, several diotic and dichotic computational models incorporating dynamic envelope and carrier interactions will be tested using the detection patterns estimated in this experiment.

An evaluation of models for diotic and dichotic detection in reproducible noises

ABSTRACT

Several psychophysical models for masked detection were evaluated using reproducible noises. The data were hit and false-alarm rates estimated in four separate studies. Models were tested with both N_0S_0 and N_0S_π stimuli at several stimulus bandwidths. A linear combination of the stimulus energy at the output of several critical-band filters was the best predictor of diotic data. The decision variables of other more complicated temporal models, including the Dau *et al.* (1996a) model and the Breebaart *et al.* (2001a) model, were only weakly correlated to the data for cases in which predictions were significant. A model that temporally combined ITD and ILD processing best explained tone-plus-noise responses to the N_0S_π stimuli, but offered no predictions for noise-alone trials.

3.1 INTRODUCTION

The traditional goal of psychophysical masked-detection experiments has been to characterize threshold signal levels as functions of physical parameters of the stimuli (e.g., signal frequency, noise bandwidth, interaural phase difference of the signal, etc.; for a review, see Durlach and Colburn, 1978). These thresholds were estimated using masking waveforms that were drawn without replacement from an infinite set on each trial, such that a new sample of masking noise was always presented. More recently, a number of studies have sought to collect data using reproducible maskers (e.g., Pfafflin and Matthews, 1966; Gilkey *et al.*, 1985; Siegel and Colburn, 1989; Isabelle and Colburn, 1991; Isabelle, 1995; Evilsizer *et al.*, 2002; Davidson *et al.* 2006). The goal of studies using reproducible maskers is to characterize detection performance for each stimulus waveform in a small set (e.g., 10 – 30), rather than describing a single threshold estimated over an infinite set of waveforms. Such data present a rigorous test for models of masked detection, because the model must not only predict average threshold to be considered successful, but it must also accurately predict detection statistics estimated for individual waveforms. As shown here and in other work, models that are capable of accurately predicting average thresholds may fail at predicting responses to individual waveforms (e.g., Isabelle, 1995).

Before explaining the details of the models tested in this work, a brief review of the target data and the methods used to collect those data is presented. Data sets from 4 studies, which shared similar experimental methods, are used in this work for modeling purposes. First, an approximate threshold was estimated using a 2-interval adaptive track, or other similar task, drawing from an infinite set of masker waveforms without replacement. Then, a fixed-level, single-interval experiment was performed with feedback, again drawing from an infinite set of masker waveforms. Subjects were asked on each trial to report if the tone was present. Feedback was then removed and signal levels were adjusted until subjects' performance was stable (customarily by monitoring detectability, d' and bias, β ; Macmillan and Creelman, 1991). Finally, the closed set of reproducible maskers replaced the infinite set used above. Maskers were drawn with replacement in a random order until each tone plus noise (T+N) and each noise alone (N) stimulus was presented multiple times (50-100). Upon completion of the experiment, hit rate, or probability of a "yes" response when the tone was present [$P(Y|T+N)$], and false-

alarm rate, or the probability of a “yes” response when the tone was *not* present [$P(Y|N)$], were tabulated. The resulting sets of hit and false-alarm rates are called detection patterns [$P(Y|W)$], or the probability of a “yes” response for any given stimulus waveform].

Hit rates and false-alarm rates from Isabelle (1995), Evilsizer *et al.* (2002), and the Appendix and Ch. 2 of this thesis served as the target data for all modeling exercises presented in this study. These data were selected because collectively, they established a set of detection patterns estimated under both N_0S_0 and N_0S_π interaural configurations, with several different noise bandwidths (50, 100, 115, and 2900 Hz), and all used a tone frequency of 500 Hz. The Isabelle (1995) data (henceforth study 1) was collected under the N_0S_π configuration only, with a noise bandwidth of 1/3 octave (approx 115 Hz) at 500 Hz. The Evilsizer *et al.* (2002) data (henceforth, study 2) were collected using N_0S_0 and N_0S_π interaural configurations, and 100-Hz (450 – 550 Hz) and 2900-Hz (100 – 3000 Hz) noise bandwidths. Corresponding stimuli from each of the bandwidths share the same frequency components in the 100-Hz region surrounding the tone frequency. The data from the Appendix of this thesis (henceforth, study 3) were collected under both N_0S_0 and N_0S_π conditions and had 50-Hz noise bandwidths. Within each interaural configuration, there were four “cue” conditions: random noise, random energy (RNRE); low noise, random energy (LNRE); random noise equal, energy (RNEE); and low noise, equal energy (LNEE). Stimuli for the low-noise conditions were produced using a modified version of the low-noise noise algorithm (Pumplin, 1985) described in the Appendix. T+N stimuli in the equal-energy conditions had overall stimulus energies normalized to the average value of all T+N stimulus energies within the low-noise and random-noise conditions, while N stimuli in the equal-energy conditions were normalized to the average value of all N stimuli within each condition. The data from Ch. 2 of this thesis (henceforth, study 4) were collected under both N_0S_0 and N_0S_π conditions and had 50-Hz noise bandwidths. There were again four stimulus conditions within each interaural configuration. The conditions were denoted E_1C_1 , E_2C_2 , E_1C_2 , and E_2C_1 ; with E denoting envelope and C denoting carrier. Corresponding stimuli within the E_1C_1 and E_1C_2 conditions and within the E_2C_1 and E_2C_2 conditions shared the same temporal-envelopes. Similarly, corresponding stimuli within the E_1C_1 and E_2C_1 conditions and within the E_1C_2 and E_2C_2 conditions shared the same carriers (i.e., had the same zero crossings). The methods section in Ch. 2 provides specific details regarding stimulus construction, but it is worth reminding the reader that the energies of T+N and N waveforms were equalized for all N_0S_0 stimuli in study 4, thus eliminating cues related to overall energy.

Several studies have examined the abilities of different models to predict detection patterns for both diotic and dichotic stimuli. Detection patterns estimated in diotic or monaural conditions are best predicted by the multiple-detector model (MD), as shown in Ahumada and Lovell (1971), Gilkey *et al.* (1986), and Davidson *et al.* (2006). This model uses the weighted sum of energies at the outputs of several auditory filters surrounding the tone frequency as a decision variable. The MD model accounted for up to 90 percent of the variance in one listener’s responses in Ahumada and Lovell (1971) and up to 72 percent of the variance in one listener’s responses in Gilkey *et al.* (1986). Predictions have also been made for study 2 using the MD model in Davidson *et al.* (2006), accounting for 80 to 90 percent of the variance in the average subject’s responses, depending on bandwidth and interaural configuration (monaural or diotic). The MD model is an extension of Fletcher’s (1940) proposal that energy at the output of the critical band (or auditory filter) centered at the tone frequency could explain threshold for tone-in-noise detection tasks. Davidson *et al.* (2006) showed that the critical-band (CB) model predicts 64 to 82 percent of the variance in their average subject’s responses. Two temporal

models were also considered: a modified version of the Richards (1992) envelope-slope (ES) model (Zhang 2004) and the phase-opponency (PO) model (Carney *et al.*, 2002). Davidson *et al.* (2006) showed that the ES and PO models predicted about 60 percent of the variance in narrowband and wideband detection patterns. These models have not been previously tested using detection patterns estimated from stimuli with energy equalized across stimulus waveforms.

Predictions for N_0S_π detection patterns have been less successful than those for N_0S_0 detection patterns. Isabelle (1995) and Colburn *et al.* (1997) analyzed several different decision variables for explaining their N_0S_π detection patterns. Colburn *et al.* (1997) considered the equalization-cancellation model (EC), and both normalized (NCC) and unnormalized cross-correlation models. They found that the EC decision variable is more dependent on stimulus energy than on external interaural differences or interaural differences resulting from the time and amplitude jitter used to establish model threshold. The dichotic detection patterns are not well-explained by an energy model, so the EC model was rejected as a suitable predictor. They found that the unnormalized cross-correlation model was too dependent on masker waveform rather than on the addition of the tone to the masker waveform for tone-plus-noise stimuli. Decision variables for an unnormalized cross-correlation model were almost identical regardless of signal presence (that is, hit rates and false-alarm rates were too similar). Colburn *et al.* (1997) found that the NCC model is equivalent to the EC model when using multiplicative time and amplitude jitter, such that the decision variable was again heavily dependent on the energy in each waveform. Isabelle (1995) showed that the variation in the NCC decision variable based on the tone waveforms was too weak with respect to the dependence of the NCC decision variable on stimulus energy to predict his data.

Isabelle (1995) was able to explain at most about 50 percent of the variance in his N_0S_π data and the Isabelle and Colburn (1991) data using signal energy (as a substitute for the EC, and NCC models), standard deviations of interaural-time differences (ITDs) and interaural level-differences (ILDs), and decision variables computed using various combinations of ITDs and ILDs. Isabelle's modeling strategy was the inspiration for the experiments presented in Ch. 2 of this thesis. The experimental results of the present study suggested that separately processing ITDs (based on fine structure) and ILDs (based on envelope) would not result in a decision variable that adequately explains the data for dichotic or diotic stimuli.

The models tested in this study were selected because they have been used to predict reproducible noise data successfully in the past (i.e., Fletcher, 1940; Ahumada and Lovell, 1971; Gilkey and Robinson, 1986), because they have been used with some success to predict thresholds for a broad spectrum of psychophysical-detection tasks (i.e., Dau *et al.*, 1996a, b; Breebaart, *et al.*, 2001a, b, c), because they are straightforward adaptations of observed physiological phenomena (i.e., McAlpine *et al.*, 2001; Marquardt and McAlpine, 2001), or because they use a processing strategy that involves a complex temporal interaction between stimulus envelope and fine structure (i.e., Goupell, 2005).

One final strategy, an interaural mismatch model (MM), was motivated by the following observation: Suppose that the binaural decision device was symmetrical and introduced no interaural differences as a consequence of its processing scheme. Under this case, a binaural system operating on *interaural differences* or *normalized cross correlation* produces essentially no decision variables for N_0S_π noise-alone stimuli (recall that the noise-alone case is diotic) *except* directly as a result of some internal noise process (if the noise is added after the binaural processor), or the way in which the internal noise decorrelates the right and left noise-alone

waveforms (if noise is added before the binaural processor). If an additive or multiplicative Gaussian noise source was responsible for producing interaural differences for noise-alone stimuli, it would, over the course of many trials, not yield detection statistics that vary consistently across noises, and yet, consistent detection patterns for noise-alone trials are observed in the data. One possible explanation for interaural differences not driven by external noise could be that slight interaural mismatches in binaural processing contribute to the decision variable (personal communication with Joseph Hall III). Such a suggestion is provocative because diotic and dichotic detection patterns are correlated for wideband stimuli (and not narrowband stimuli), indicating that either the mismatch could be exploited in both diotic and dichotic detection or that it is bandwidth dependent.

A brief overview of each model featured in this study is given here. (Specific model structures will be described in the methods section.) The general modeling approach is divided into two distinct sections: predictions for the N_0S_0 interaural configuration and predictions for the N_0S_π interaural configuration. The CB, MD, Dau *et al.* (1996a) model (DA) and Breebaart *et al.* (2001a) models (BR) were applied to all the diotic data (sets 2, 3 and 4). The Dau model (Dau *et al.*, 1996a,b) has been used to predict thresholds in a number of different monaural (or diotic) psychophysical tasks including detection of tones in random and frozen noise as a function of temporal position, duration, and frequency of the tone, as well as forward and backward masking tasks. This model's decision variable is computed primarily from the stimulus envelope. The Breebaart model (Breebaart *et al.*, 2001a,b,c) has also been used to accurately predict results for a number of psychophysical tasks such as masking-level differences in a multitude of different interaural configurations, as functions of tone frequency, noise bandwidth, masker energy, as well as tasks such as interaural correlation discrimination, effects of masker fringe, and binaural forward masking. Monaurally, this model is also envelope dependent, although it employs a different detection strategy than the Dau model (described in detail in the methods section). Thus the relation of the Dau and Breebaart decision variables was of interest, in addition to the fact that neither of these models were designed to incorporate temporal-fine structure information at the monaural level.

The models considered for N_0S_π detection patterns are briefly summarized below. Several of the decision variables (standard deviations of ITD, ILD, and combinations thereof) from Isabelle (1995) were re-examined with the newer data from studies 2, 3 and 4. These decision variables were supplemented with decision variables from Goupell (2005) and Goupell and Hartmann (2005). The Goupell decision variables were adapted from the Isabelle decision variables to include two distinct groups that make use of both ITD and ILD: "Separate centers" models in which integration over time occurs separately for the decision variable based on ITD and ILD, and "auditory image" models, in which ITDs and ILDs are allowed to interact in some way as a function of time. The results from Ch. 2 suggest that because the Isabelle decision variables do not allow envelope and fine-structure to interact temporally, they will not be capable of predicting the detection patterns; thus, it is of interest to determine the effectiveness of the Goupell "auditory image" decision variables that allow for this interaction. A variant of the Marquardt and McAlpine (2001) model for masked detection was also tested; this model has been shown to successfully predict masked-detection thresholds using only 4 binaural delay channels (henceforth referred to as the 4-channel model, FC). This model was inspired by the findings of McAlpine *et al.* (2001) who reported that recordings from delay-sensitive neurons in the guinea pig inferior colliculus were centered around 45° , regardless of the neurons' best frequencies. A model based on binaurally-mismatching frequency channels (i.e., channels tuned

to different characteristic frequencies) and specific interaural delays and attenuations was also tested (MM). The binaural counterpart of the Breebaart model was also tested; this model makes use of temporal fine structure in the binaural processor.

3.2. METHODS

3.2.1 General modeling strategy

The models included in the following analyses are of varying complexity, ranging from simple decision variables (e.g., RMS energy) to complicated, multi-stage psychophysical models (e.g., the Breebaart model). Some of the published versions of these models make explicit assumptions about internal noise, while others do not include internal noise. In general, internal noise is used to calibrate each model such that the model's overall threshold becomes reasonable for some given task. For the purposes of this analysis, incorporating internal noise would reduce the correlation of the model decision variables to the various detection patterns. (This is true in all cases except for certain model responses to N_0S_{π} noise-alone stimuli for which, as explained above, internal noise itself could be responsible for generating the noise-alone model responses. Because such responses are a result of a zero-mean noise process, they cannot be correlated with the detection patterns.) Since the goal of this study was to show the best possible correlations of each model to the data, internal noise was not included here. As a consequence of not including internal noise, individual model thresholds, which would be universally underestimated, are not reported. With the exception of the MM model, each of the models tested here has been shown in the previous literature to accurately predict thresholds using randomly-generated noises when incorporating internal noise. Because each model was tested at each individual subject's threshold without internal noise, model performance (i.e., d') varied across models and subjects, and in general produced d' values greater than unity. This variability in performance and lack of internal noise has a net effect of artificially increasing r^2 values for $P(Y|W)$ when model d' 's are large, due to separation of the distributions of $P(Y|T+N)$ and $P(Y|N)$. Thus, modeling analyses were confined to $P(Y|T+N)$ and $P(Y|N)$. The effect of analyzing hit and false-alarm rates separately is to lower the proportions of variance explained with respect to the variance that might be explained in $P(Y|W)$. The smaller values are due in part to the reduction of the number of waveforms included in the computation of r^2 and in part to the separation of the relationship of model d' and the proportion of variance explained.

The strategy used to evaluate all models was to establish a decision variable for each stimulus waveform within each study. The decision variables were then correlated to the z -scores of the listeners' $P(Y|T+N)$ or $P(Y|N)$.

The r^2 metric, or the square of the Pearson product-moment correlation, was used to quantify model predictions and can be interpreted as the percent of the variance in each detection pattern explained by each model. It is important to establish an upper limit of expected performance for any given prediction (V_p). Model results are therefore presented in duplicate; first in terms of r^2 and then in terms of an estimate of the percent of *predictable* variance (r_p^2) explained, computed as the ratio of r^2 over V_p . Isabelle (1995) described that the reasonable upper limit for predicting his data (study 1) was an r^2 of about 0.88. Evilsizer *et al.* (2002) report first-half, last-half correlations that yield predictable variances (V_p) from 0.80 to 0.97 depending on subject. Predictable variances are reported for studies 3 and 4 in the Appendix and in Ch. 2 of this dissertation, respectively, and ranged from 0.18 to 0.99.

3.2.2 Individual model implementations

3.2.2.1 Diotic models

3.2.2.1.1 Critical-band model

A block diagram of the critical-band model is shown in Fig. 3-1. This model was based on Fletcher's (1940) suggestion that detection can be explained by the energy at the output of an auditory filter centered at the tone frequency. The model decision variable was the RMS output of a 4th-order gamma-tone filter centered at 500 Hz. The decision variable increased upon addition of the target tone. The equivalent rectangular bandwidth (ERB) of the filter was set at 75 Hz to correspond to the estimate of Glasberg and Moore (1990). The model decision variable was given by

$$F(j) = \sqrt{\frac{1}{\tau} \int_0^{\tau} f(j,t)^2 dt} \quad \text{Eq. (1)}$$

where f is the output of the gammatone filter, t is time, and τ is the duration of the stimulus waveform j .

3.2.2.1.2 Multiple-detector model

This model was based on a linear combination of the RMS output of several 4th-order gammatone filters (Fig. 3-1). The results of Davidson *et al.* (2006) showed that filters exceeding the bandwidth of the stimulus noise do not significantly increase the predictive power of the model. Therefore, center frequencies were selected that spanned 275 to 725 Hz (in 75-Hz increments) for the 2900-Hz condition of study 2, and 425 to 575 Hz for the 100-Hz condition of study 2. The MD model was not used to predict the data from studies 3 and 4, for which the masker bandwidth was only 50 Hz. The filter bandwidth was held constant at 75 Hz to match the methods of Davidson *et al.* (2006). A block diagram of the MD model is shown in Fig. 3-1.

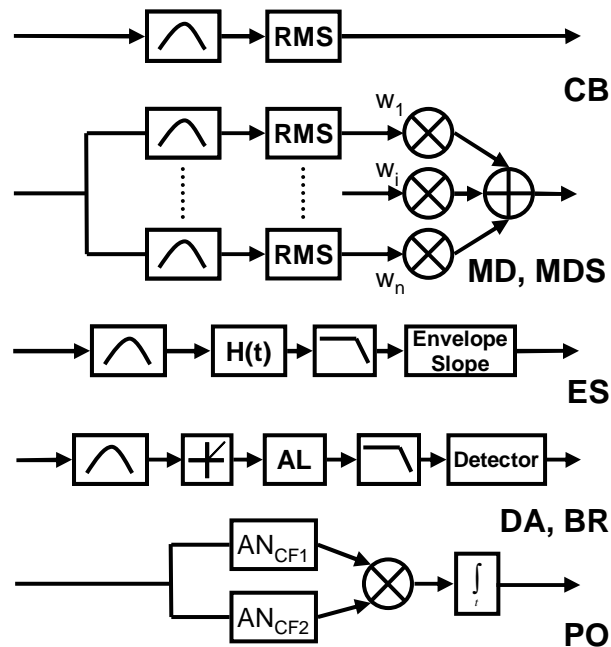


Figure 3-1. Block diagrams of the models used to predict detection patterns estimated under the N_0S_0 conditions. The models listed from top to bottom are: CB, critical band; MD, multiple detector with fit weights; MDS, multiple detector with sub-optimal weights; ES, envelope slope; DA, Dau *et al.*; BR, Breebaart *et al.*; PO, phase opponency. $H(t)$ denotes the Hilbert transform used to recover the absolute value of the complex-analytic signal. AL denotes the adaptation loops as described in Dau *et al.* (1996a). AN denotes the Heinz *et al.* (2001) auditory-nerve model.

The weights (w_i) for the linear combination were established using two separate methods. In the first method (MD), the weights were fit to the individual subject's detection patterns using the reproducible stimuli from each study. The MATLAB function `fminsearch` was used to minimize the quantity of one minus the correlation coefficient of the linear combination of the RMS filter outputs and the z -scores of $P(Y|T+N)$ and $P(Y|N)$ for each subject in each condition in studies 1 and 2. Thus, the MD model uses a fit to the subjects' data, rather than a decision-theoretic weighting strategy. In the second method (MDS), a decision-theoretic sub-optimal weighting scheme was used rather than fitting the weights to the listeners' data, in order to compare the resulting weight profiles across frequency and the variance of the detection patterns explained by each method. Individual weights were computed using 1000 repetitions of randomly-created (i.e., not reproducible) noise. Tones were added and weights were computed as

$$w(i) = \frac{\overline{F(i, m)_{(T+N)}} - \overline{F(i, m)_{(N)}}}{\frac{\text{var}(F(i, m)_{(T+N)}) + \text{var}(F(i, m)_{(N)})}{2}} \quad \text{Eq. (2)}$$

where F is the root-mean-squared filter output for frequency channel i and random-noise repetition m for T+N or N stimuli, and the means and variances were computed across each repetition (m) within frequency channel (i). (Note that this method would be optimal if the covariance of each channel were accounted for in Eq. 2. The models decision variables were given by

$$M(j) = \sum_i F(i, j)w(i), \quad \text{Eq. (3)}$$

using the weights computed with either method above.

3.2.2.1.3 Envelope-slope model

The envelope-slope model is a metric for quantifying fluctuation in the stimulus envelope (ES, Fig. 3-1). Methods were matched to Davidson *et al.* (2006). The model decision variable was computed as

$$E(j) = \frac{\sum_t |x[t - \Delta t, j] - x[t, j]|}{\sum_t x[t, j]} \quad \text{Eq. (4)}$$

where $x[t, j]$ is the Hilbert envelope of the output of a 4th-order gammatone filter centered at 500 Hz, with a 75-Hz ERB for stimulus waveform j . To ensure that all fine-structure was removed from the stimulus waveform, $x[t, j]$ was filtered with a 10th-order maximally flat IIR filter with a cut-off frequency of 250 Hz before being processed with Eq. (1). The statistic was normalized as suggested by Zhang (2004) to remove effects of energy and duration. Upon addition of the tone to the noise waveform the stimulus envelope flattens. As such, the decision variable decreases with increasing tone level.

3.2.2.1.4 Dau model

The Dau model is significantly more complex than the previously described decision variables (DA, Fig. 3-1). The model consists of a third-order gammatone filter centered at the tone frequency (500 Hz) using Glasberg and Moore (1990) filter bandwidths (1 ERB is approximately 75 Hz at a center frequency of 500 Hz). The output is half-wave rectified and

passed on to a series of adaptation loops (Dau *et al.* 1996a). The adaptation loops are designed to simulate adaptation found in auditory-nerve responses by processing fast stimulus fluctuations almost linearly and compressing slowly-fluctuating stimuli. The output of the adaptation loops is low-pass filtered with a time constant of 20 msec (8 Hz), effectively removing fine structure and leaving envelope information. The output at this stage is referred to as the “internal representation” of the model. The internal representation is passed to an optimal detector. The optimal detector uses a template derived from the normalized difference between the mean of 500 T+N internal representations subtracted from the mean of 500 N internal representations. Such a large number of noises were used to simulate extensive training. The templates were computed using randomly generated noise with a signal added at 10 dB above each listener’s threshold. The optimal detector first subtracts the noise-alone template from each of the T+N and N internal representations corresponding to the reproducible stimuli. The mean scalar product of the normalized difference template and the difference between the noise-alone template and the internal representation of each reproducible stimulus is then computed as a function of time. The model was originally designed to pick the interval (from a 2-interval task) with the larger scalar product as containing the tone. For the purposes of this study, which focuses on single-interval tasks, the scalar product itself was used as the decision variable. This process is summarized with

$$D(j) = \frac{1}{\tau} \int_0^{\tau} [\varphi_j(j, t) - \gamma_N(t)] \frac{[\gamma_{T+N}(t) - \gamma_N(t)]}{\text{RMS}[\gamma_{T+N}(t) - \gamma_N(t)]} dt \quad \text{Eq. (5)}$$

where D is the Dau decision variable, φ_j is the internal representation of the current stimulus waveform j , γ_{T+N} is the mean of 500 internal representations of T+N stimulus waveforms (the T+N template), γ_N is the mean of 500 internal representations of N stimuli (the N template), τ is the duration of the stimulus waveform, and RMS is the root-mean-squared function.

3.2.2.1.5 Breebaart Model

The diotic version of the Breebaart model is shown in Fig. 3-1 (BR). This model is similar to the Dau model but has the following differences: The Breebaart model was implemented as a bank of processors with different center frequencies. Filters were implemented with a spacing of 2 filters per ERB over the same bandwidths as the MD and MDS models. The low-pass filter from the Dau model was replaced with a double-sided exponential window with time constants of 10 ms each. The structure of the decision device is described in detail in Breebaart *et al.* (2001a), and is composed of a sub-optimally weighted combination of internal representations at different frequency channels, which are then summed as a function of time and frequency. This model also makes use of both T+N and N templates. The templates were established as the means of 50 internal representations³ of randomly-generated T+N and N waveforms at each listener’s threshold. The detector first computes the decision variable B according to

$$B(j) = \int_i \int_{\tau} \frac{\mu(i, t)}{\sigma^2(i, t)} U(j, i, t) di dt. \quad \text{Eq. (6)}$$

³ This number was reduced from 500 for practical considerations. The sensitivity of model decision variables on the number of internal representations was not great; results were stable for 20 or more repetitions.

The quantity $U(j, i, t)$ is the difference between the internal representation of the reproducible waveform j , and the N template for each frequency channel i . This $U(j, i, t)$ is weighted by the difference between the T+N and N templates (μ) over the variance of the N templates (σ^2).

3.2.2.1.6 Phase-opponency model

The phase-opponency model was computed as described in Davidson *et al.* (2006) and was based on the model described by Carney *et al.* (2002) (PO, Fig. 3-1). Two Heinz *et al.* (2001) model auditory-nerve fibers with spontaneous rates of 50 spikes/sec converged upon a coincidence detector of the type described in Colburn (1977). The fibers' center frequencies were selected such that their phase responses differed by 180° at the tone frequency (which occurred for the two center frequencies of 459 and 542 Hz). The count at the output of the coincidence detector was used as the model decision variable as described by

$$G(j) = n_{fib}^2 T_{CW} \int_0^\tau \kappa_{459}(j, t) \kappa_{542}(j, t) dt \quad \text{Eq. (7)}$$

where n_{fib} is the number of auditory-nerve fiber inputs at each center frequency, T_{CW} is the time window for coincidence detection, t is time, τ is the duration of the stimulus, κ is the output of the Heinz *et al.* (2001) auditory-nerve model at each of the two center frequencies. The model decision variable (G) was computed for each reproducible stimulus j . As the level of the tone is increased, the count at the output of the coincidence detector is reduced because the two model fibers progress to firing perfectly out of phase. Ten model fibers were used with a coincidence window of 20 μ s. As in Davidson *et al.* (2006), the onsets and offsets of the auditory-nerve fiber responses were truncated because they exceeded realistic levels and did not produce decision variables correlated to the psychophysical data. Due to the use of relatively short-duration stimuli in the present study, only the first and last 25 ms of the responses were truncated.

3.2.2.2 Dichotic models

3.2.2.2.1 Isabelle (1995) and Goupell (2005) decision variables

Isabelle (1995) used several decision variables that were based either on fine structure, (i.e., ITDs), or on envelope (i.e., ILDs). ITDs and ILDs are given by

$$\Psi_{\Delta\Phi}(j, t) = \frac{\phi_L(j, t) - \phi_R(j, t)}{\omega} \quad \text{Eq. (8)}$$

$$\Psi_{\Delta L}(j, t) = 20 \log \frac{A_L(j, t)}{A_R(j, t)} \quad \text{Eq. (9)}$$

where $\phi(j, t)$ is the instantaneous phase from the complex analytic signal for the right or left stimulus waveforms, ω is the center frequency of the noise band, $A(j, t)$ is the envelope of the complex analytic signal for either the right or left stimulus waveform, and j is the index of each reproducible stimulus waveform. The complex analytic signals were computed using the Hilbert transform. It should be noted that since internal noise was not used in the present implementation, the resulting decision variables described in this section will be identically 0 for noise-alone (and diotic) stimuli. Therefore, predictions were not computed for $P(Y|N)$. A selection of several decision variables featured in Isabelle (1995) and Goupell (2005) are shown below: These included the variances of ITD and ILD computed for each reproducible stimulus as defined by

$$s(j)_{\Delta\Phi}^2 = \int_0^{\tau} (\Psi_{\Delta\Phi}(j,t) - \overline{\Psi_{\Delta\Phi}})^2 dt \quad \text{Eq. (10)}$$

$$s(j)_{\Delta L}^2 = \int_0^{\tau} (\Psi_{\Delta L}(j,t) - \overline{\Psi_{\Delta L}})^2 dt . \quad \text{Eq. (11)}$$

A weighted combination of the standard deviations of ITD and ILD [Eq. (12)] and a combination of the average values of ITD and ILD [Eq. (13)] were also explored, as defined by

$$W_{ST}(j, a) = a s_{\Delta\Phi}(j) + (1 - a) s_{\Delta L}(j), \quad \text{and} \quad \text{Eq. (12)}$$

$$W_{AV}(j, a) = \frac{1}{\tau} \int_0^{\tau} a |\Psi_{\Delta\Phi}(j,t)| dt + \frac{1}{\tau} \int_0^{\tau} (1 - a) |\Psi_{\Delta L}(j,t)| dt \quad \text{Eq. (13)}$$

respectively, where a is a weight determined by minimizing the sum of squared errors between W and $z\{P(Y|T+N)\}$ for each condition and subject in each study. Note that the decision variables described by Eqs. (12-13) were called “separate centers” models by Goupell (2005), because the standard deviations (or average absolute values) of ITD and ILD were computed before the weighted combination of ITD and ILD was computed. The four metrics described in Eqs. (10-13) were compared to the “auditory image” decision variables of Goupell (2005), which included the standard deviation of a temporal combination of ITD and ILD [Eq. (14)] as well as a temporal combination of average values of ITD and ILD [Eq. (15)] defined by

$$X_{ST}(j, b) = s_t [b \Psi_{\Delta\Phi}(j,t) + (1 - b) \Psi_{\Delta L}(j,t)] \quad \text{and} \quad \text{Eq. (14)}$$

$$X_{AV}(j, b) = \frac{1}{\tau} \int_0^{\tau} |b \Psi_{\Delta\Phi}(j,t) + (1 - b) \Psi_{\Delta L}(j,t)| dt \quad \text{Eq. (15)}$$

respectively, where b is a weight computed in the same manner as in Eqs. (12 and 13). Goupell (2005) called the decision variables computed in Eqs. (14-15) “auditory image” models because ITD and ILD were combined before computing standard deviations or summing over time.

Finally, a lateral-position model was considered (Hafter, 1971) which is based on a combination of ITD and ILD using a trading ratio of 20 $\mu\text{s}/\text{dB}$ and is defined by

$$L(j, a) = \frac{1}{\tau} \int_0^{\tau} |\Psi_{\Delta\Phi}(j,t) + a \Psi_{\Delta L}(j,t)| dt . \quad \text{Eq. (16)}$$

where τ is the duration of the stimulus and a is the trading ratio. These models [Eqs. (10-16)] were of particular interest because they allow for the distinct interaction of statistics based on envelope and carrier as a function of time. The lateral position model is similar to Eqs. (14-15) except a constant trading ratio was used for all computations.

3.2.2.2.2 Four-channel model

The general structure of the four-channel model (FC; Marquardt and McAlpine, 2001) is shown in Fig. 3-2. The right and left stimulus waveforms were processed using the Heinz *et al.* (2001) auditory-nerve model. The output of each filter was passed to a delay line with a delay of 45° (corresponding to 250 μs at 500 Hz) on each side. The delayed stimulus from the ipsilateral

side and delayed stimulus from the contralateral side converged onto binaural coincidence detectors where both the unnormalized cross correlation and the binaural cancellation (difference) were computed for each channel (in order to approximate EE and EI neurons, respectively). The cells were thus tuned to $\pm 45^\circ$ and $\pm 135^\circ$, spanning the entire range of possible interaural phase differences at 500 Hz in relative increments of 90° . The outputs of the four binaural channels were sub-optimally weighted and summed using the same strategy as used for the MDS model. In summary,

$$F_C(j) = \{w_1[\kappa(j,t)_R - \kappa(j,t-D)_L]\} + [w_2\kappa(j,t)_R\kappa(j,t-D)_L] + [w_3\kappa(j,t-D)_R\kappa(j,t)_L] + \{w_4[\kappa(j,t)_L - \kappa(j,t-D)_R]\}, \text{ Eq. (17)}$$

where $F_C(j)$ is the model decision variable for the reproducible stimulus (j), $\kappa(j,t,D)$ is the output of the Heinz *et al.* (2001) auditory-nerve model delayed by D seconds (45°)

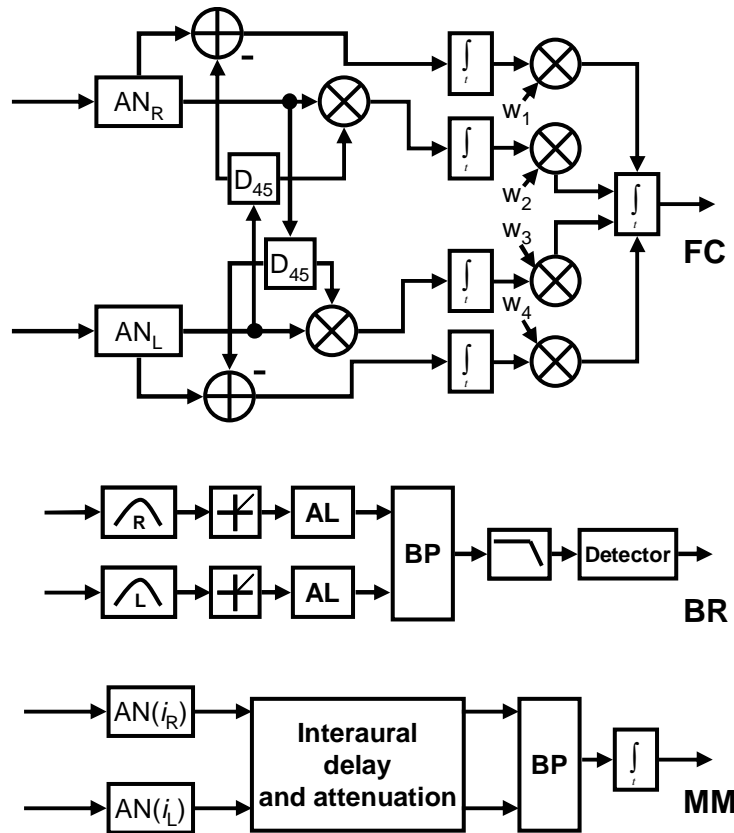


Figure 3-2. Block diagrams of the models used to predict detection patterns estimated under the N_0S_π interaural configuration. The models listed from top to bottom are: FC, four-channel; BR, Breebaart model. AN denotes the Heinz *et al.* (2001) auditory-nerve model. D denotes a delay block computed at the center frequency of each model auditory-nerve fiber. AL denotes the adaptation loops as described in Dau *et al.* (1996a). BP denotes a binaural processor.

or 250 μs) and w is the sub-optimal weight computed for each delay channel as shown in Fig. 3-2.

3.2.2.2.3 Breebaart model

The binaural version of the Breebaart model is an extension of the monaural model described above. A simplified block diagram is shown in Fig. 3-2 (BR). The output of the adaptation loops from the ipsilateral and contralateral sides are passed to a binaural processor that simulates an excitatory-inhibitory interaction. The processor was originally designed with a series of attenuation taps and delays as described in Breebaart *et al.* (2001). The model selects only the single delay and attenuation line that shows the greatest change in output between T+N and N stimuli. The zero delay and attenuation channel always shows the largest change in output for N_0S_π stimuli, thus allowing the model here to be collapsed to only the zero-delay, zero-attenuation channel. The binaural processor is described by

$$E(j,i,t) = a \log b [(\varphi_L(j,i,t) - \varphi_R(j,i,t)) + 1]^2 \quad \text{Eq. (18)}$$

for N_0S_π stimuli, where $\varphi(j,i,t)$ describes the output of the adaptation loops for reproducible stimulus j , frequency channel i , at time t . $E(j,i,t)$ is then filtered with a double-exponential window with a time constant of 30 ms per exponential. The filtered $E(j,i,t)$ is then scaled, compressed with a logarithm and then scaled again as shown in Eq. (18), with $a = 0.1$ and $b = 0.00002$. The two scale factors were calibrated by setting the model threshold to predict N_0S_π and N_pS_π detection tasks as described in Breebaart *et al.* (2001). The detector operates similarly for the monaural and binaural models as described in Eq. (6). However, for the binaural case, the temporally-weighted internal representation of each waveform is simply summed over time and frequency, and the templates are computed using the compressed and filtered output of the binaural processor (as opposed to computing a difference between T+N and N templates and using the filtered output of the adaptation loops as in the monaural model; recall that the N template is identically zero for the binaural model).

3.2.2.2.4 Mismatch model

The mismatch model (MM) is an adaptation of the standard binaural delay line approach proposed by Jeffress (1948) and the idea of attenuation taps as used in Breebaart *et al.* (2001). This model is essentially a normalized cross-correlation model with inputs that are free to vary in frequency, delay, and attenuation (Fig. 3-2). The model parameters were selected using a search procedure designed to maximize the correlation between z scores of individual subject's $P(Y|T+N)$ and $P(Y|N)$ values (individually) and the model decision variables. The parameter space was selected based on the distributions of Zurek (1991) with spacing to allow the search to be completed in a reasonable amount of time (1 day). There were a total of 4 free parameters that were varied systematically: the right and left auditory-nerve center frequencies, the interaural phase delay, and the interaural level difference. The right and left auditory-nerve center frequencies ranged from 300 to 700 Hz with 50 Hz spacing, while interaural phase delays ranged from -400 to 400 μs in 100 μs increments and interaural level differences ranged from -10 to 10 dB in steps of 5 dB. Colburn *et al.* (1997) showed that the normalized (and unnormalized) cross correlation model produced decision variables for individual noise waveforms that were too correlated to overall energy to explain $P(Y|T+N)$ and $P(Y|N)$. However, those computations assumed no periphery and symmetric processing (with no interaural delays or attenuations). In this modeling exercise both normalized (MMn) and unnormalized (MMc) cross correlations, as well as an excitatory-inhibitory interaction (MMe) are used (separately) as binaural processors

operating on the output of the Heinz *et al.* (2001) auditory-nerve model driving functions. The goal was to determine if any of the best-fit parameters remained constant across different stimulus conditions for each subject, to determine the difference between the best-fit parameters that were fit to $P(Y|T+N)$ and $P(Y|N)$, and to determine if the resulting decision variables were correlated to the listeners' detection patterns.

3.3 Results and Discussion

3.3.1 Diotic models

Predictions for each model are shown in Figs. 3-3 and 3-4 for $P(Y|T+N)$ and $P(Y|N)$ respectively. The critical r^2 value for a significant prediction is 0.16, as denoted by the horizontal-dashed line. These predictions are re-plotted in Figs. 3-5 and 3-6 in terms of r_p^2 , or the square of the correlation coefficient (r^2) normalized by the proportion of predictable variance (V) for each subject within each condition in each study. Although the r_p^2 predictions are provided, they are not a substrate for performing statistical tests, and most of the discussion will be oriented to the unnormalized r^2 values. Each subject in each study is denoted using a different symbol. Identical symbols do not correspond to the same subjects across studies (but do correspond to the same subjects within studies). The CB model is discussed first, followed by the MD model, after which results from the temporal models (ES, DA, BR, and PO) are presented for conditions where stimulus energies were equalized.

3.3.1.1 Individual model results

3.3.1.1.1 Critical-band model

The critical band model was the simplest model tested in this study and in general, made significant predictions for all subjects but S4 in study 2 for $P(Y|T+N)$ and significant predictions for all subjects but S3 and S4 in study 3 for $P(Y|N)$. Recall that energy was equalized among T+N and N trials separately for the RNEE and LNRE conditions of study 3 and overall energies were equalized for all stimuli in study 4. Inspection of Figs. 3-3 and 3-4 shows some significant predictions for the energy model under these equal-energy conditions. These predictions *appear* significant because no internal noise was used in the simulations. One might suspect that the peripheral filter included in this study recovered energy differences from stimulus-to-stimulus. However, careful inspection of the decision variables calculated for the CB model in EE cases of study 2 reveals a maximum difference of only 0.65 dB SPL between the stimulus waveform with the highest level and the stimulus waveform with the lowest level (when compared to differences on the order of 8 dB SPL for RE stimuli). The largest difference between levels at the output of the gammatone filter for the stimuli in study 4 was about 1 dB. In order for the CB model to explain these results the listeners would have had to reliably measure the output of a critical band filter with a resolution of about 0.04 dB (to correctly order 25 T+N or N stimulus waveforms in terms of level) in the presence of internal noise with an effective variance of approximately 1 dB across noises (estimated assuming the internal-to-external noise ratio is approximately 1 for the data from study 2 in the conditions where the data is correlated to the CB model). Thus, the CB model failed in the equal energy cases but was significantly correlated to subjects' detection patterns when energy cues were available. This finding is in agreement with the results from the Appendix of this thesis, and with the results of Richards (1992).

3.3.1.1.2 Multiple-detector model

The reader is first reminded that the MD model was only applied to the results for study 2 (which had super-critical noise bandwidths). The MD model has been used to predict these data previously (Davidson *et al.*, 2006), but only predictions for $P(Y|W)$ are shown in that study.

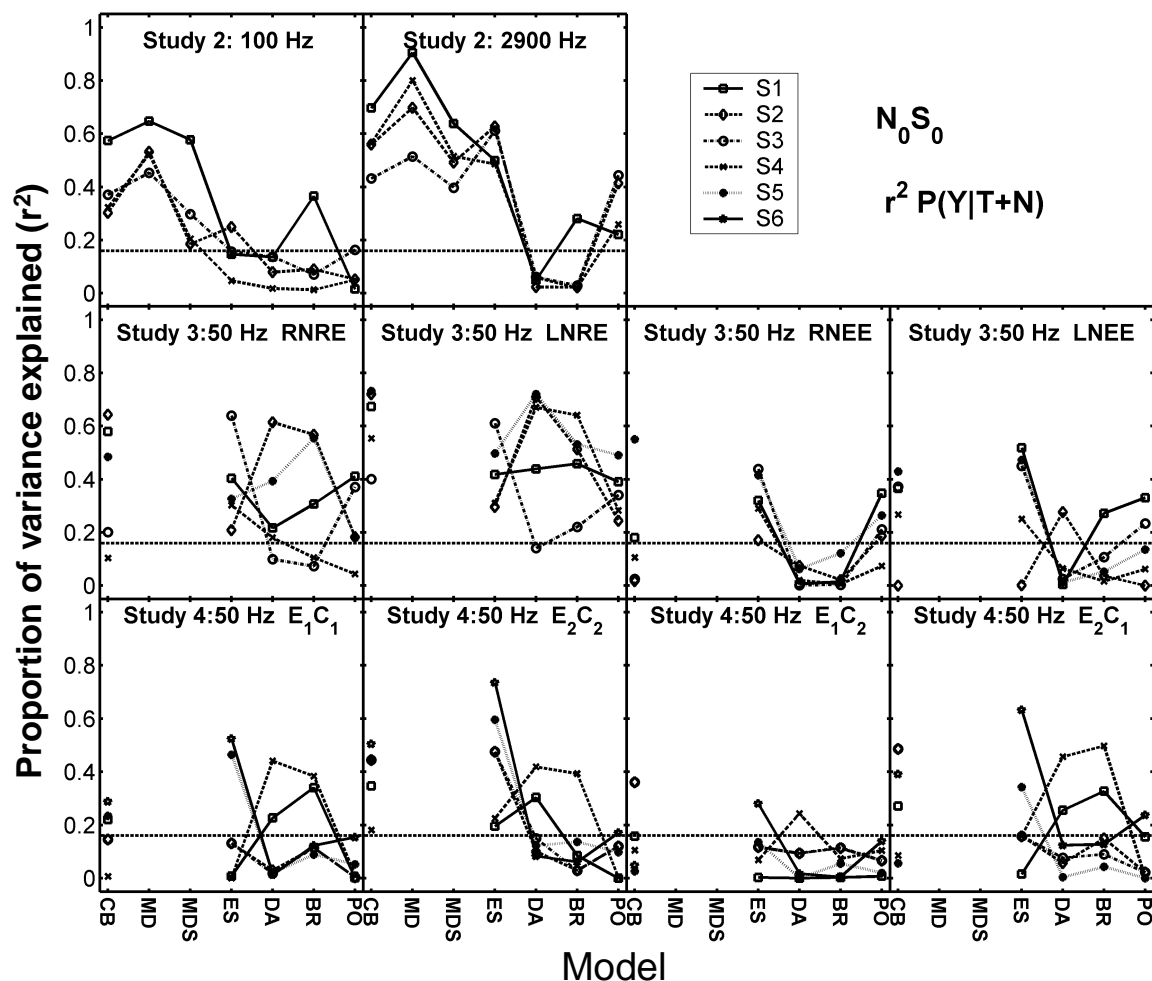


Figure 3-3. Proportion of variance explained in the z scores of $P(Y|T+N)$ by the N_0S_0 model predictions. Results were quantified using the square of the correlation coefficient (r^2). Model abbreviations from left to right: CB, critical band; MD, multiple detector; MDS, multiple detector with sub-optimal weights; ES envelope slope; DA, Dau; BR, Breebaart; and PO, the Phase opponency. Different subjects are indicated with different symbols connected across models to facilitate intersubject comparisons. Note that subjects sharing the same number do not correspond across studies.

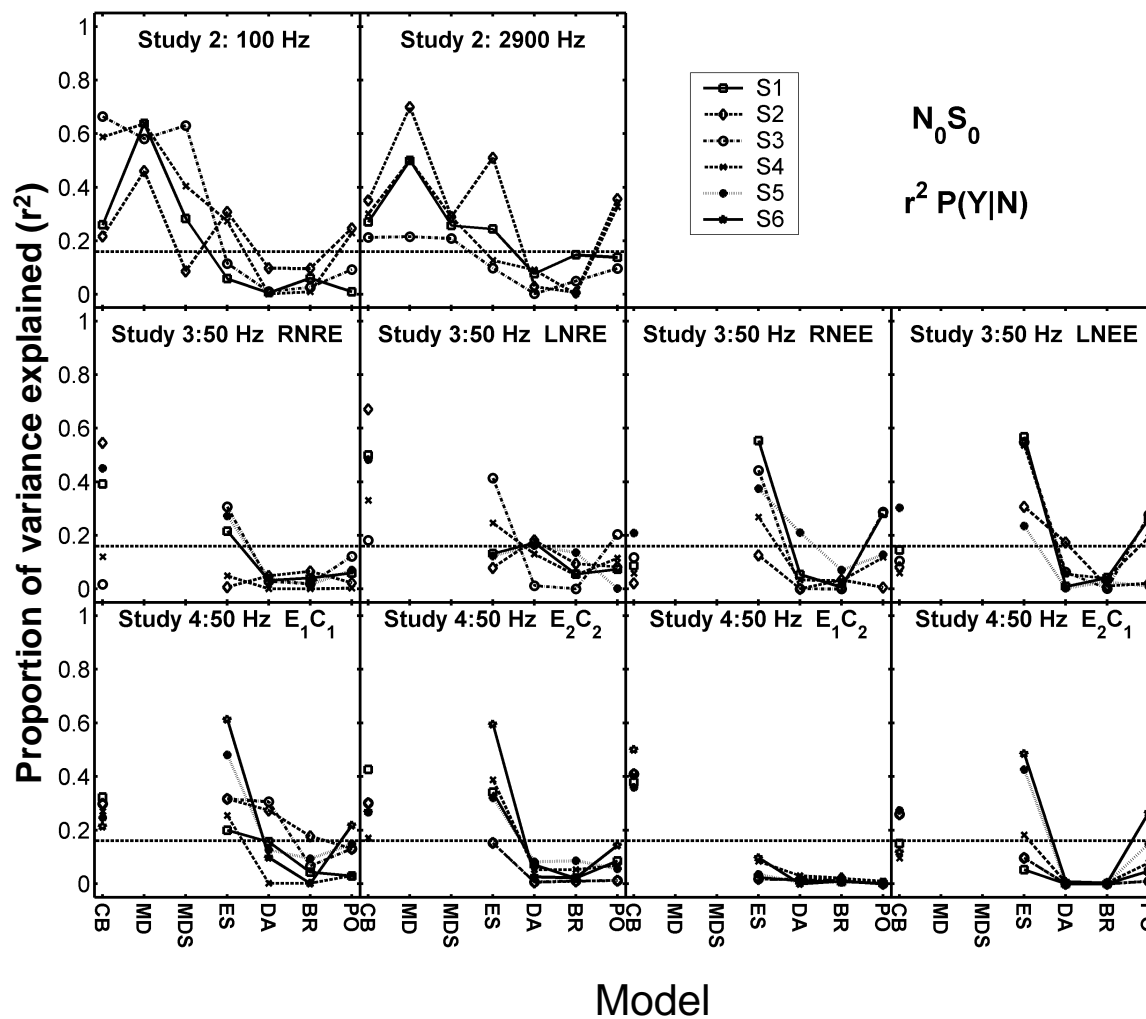


Figure 3-4. Proportion of variance explained in the z scores of $P(Y|N)$ by the N_0S_0 model predictions. Results were quantified using the square of the correlation coefficient (r^2). Model abbreviations are as in Fig. 3-3. Note that subjects sharing the same number do not correspond across studies.

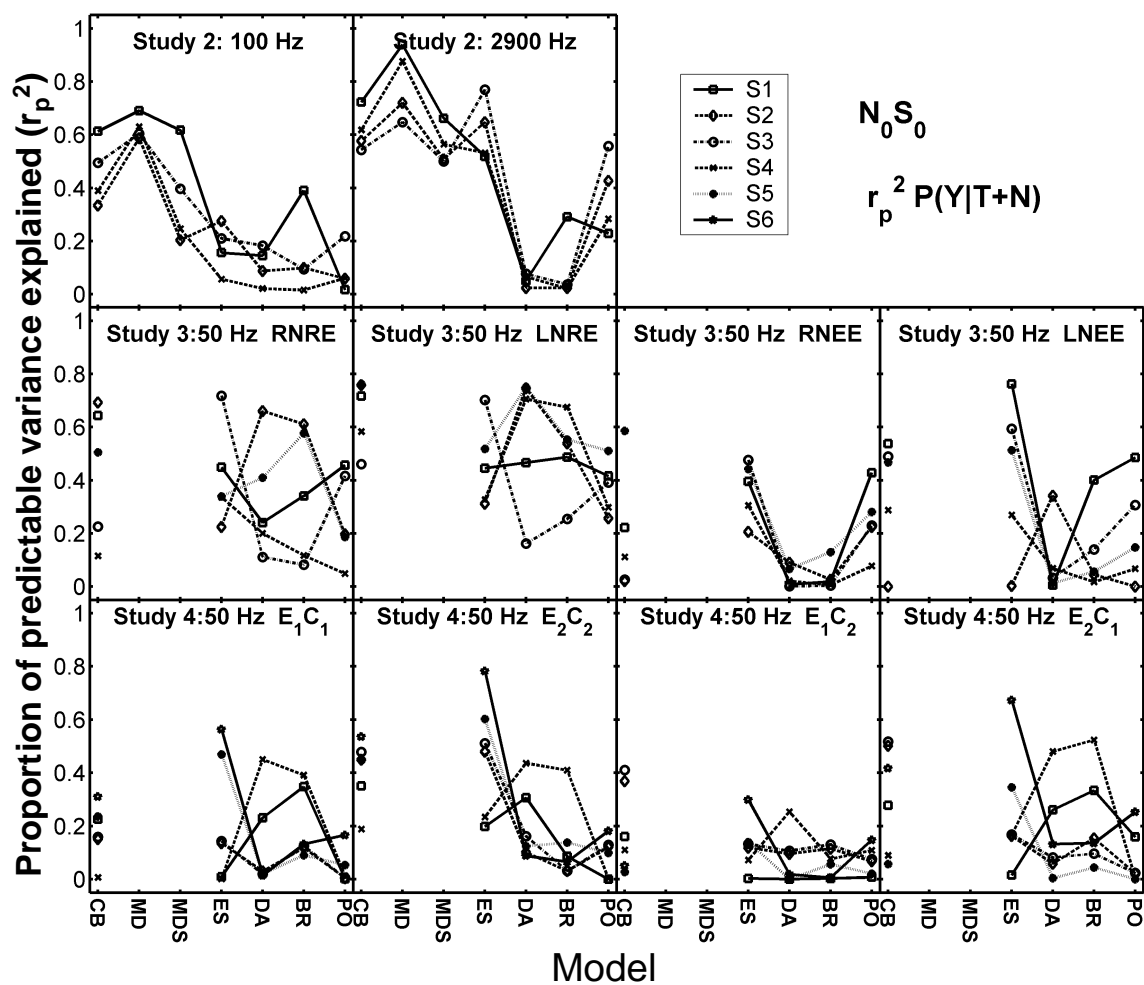


Figure 3-5. Proportion of variance explained in the z scores of $P(Y|T+N)$ by the $N_0 S_0$ model predictions. Results were quantified using the square of the correlation coefficient normalized by the proportion of predictable variance ($r_p^2 = r^2/V$). Model abbreviations from left to right are the same as in Fig. 3-3. Note that subjects sharing the same number do not correspond across studies.

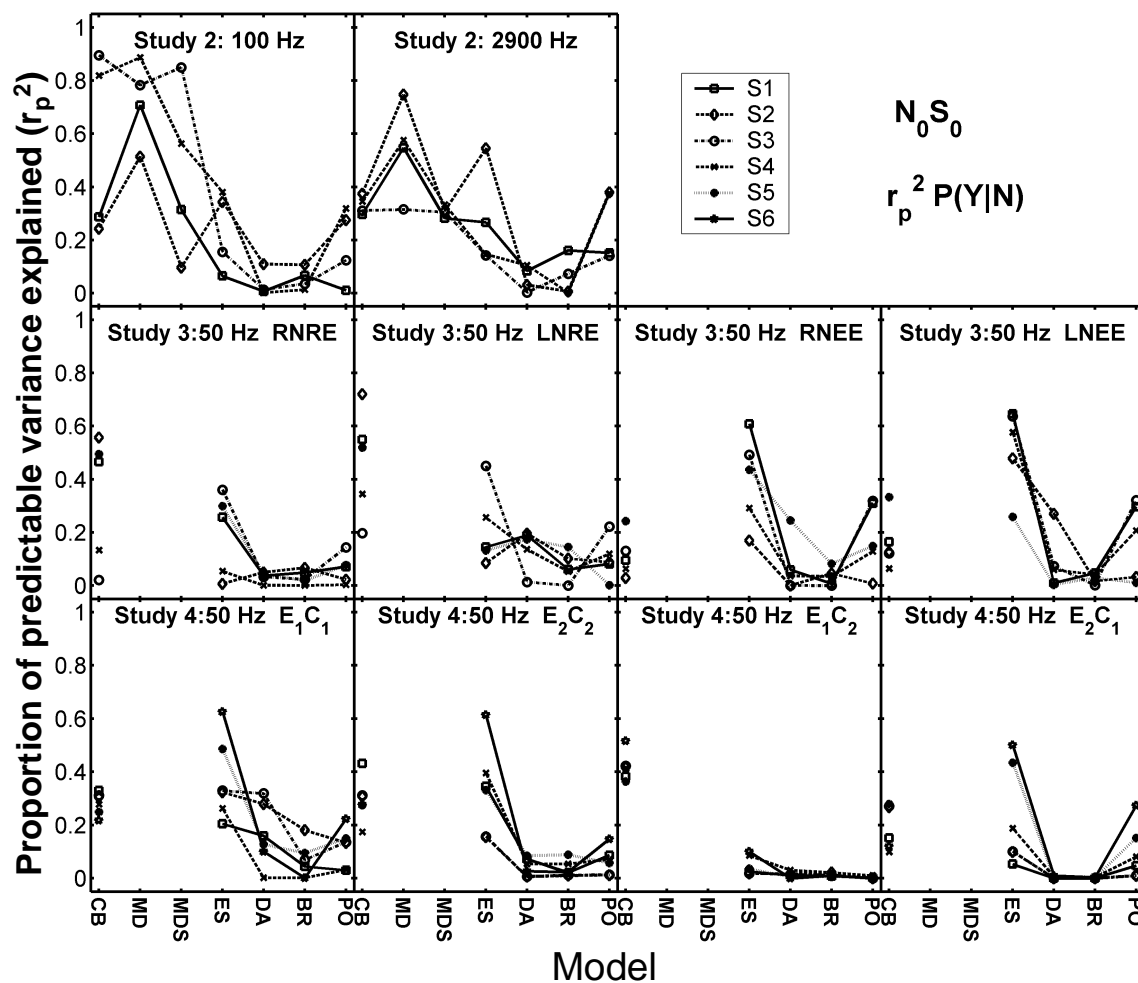


Figure 3-6. Proportion of variance explained in the z scores of $P(Y|N)$ by the $N_0 S_0$ model predictions. Results were quantified using the square of the correlation coefficient normalized by the proportion of predictable variance ($r_p^2 = r^2/V$). Model abbreviations from left to right are the same as in Fig. 3-3. Note that subjects sharing the same number do not correspond across studies.

Here, individual predictions for $P(Y|T+N)$ and $P(Y|N)$ are presented, along with predictions that were derived using sub-optimal weights. Weights resulting from the fit to the data (MD model) and from the sub-optimal computation (MDS model) are shown in Fig. 3-7. Note that the negative weights found above and below the target frequency were not present in the sub-optimal weighting scheme. The weighting strategy used by S3 in the 100-Hz condition was close to that of the sub-optimal scheme (recall that the sub-optimal weights attempt to maximize d' rather than increasing r^2). Inspection of Figs. 3-3 through 3-6 shows that predictions of the MD and MDS models for $P(Y|T+N)$ and $P(Y|N)$ are similar for that listener. Detection patterns produced by listeners that tended to have more negative peripheral weights for the MD model were not well predicted by the MDS model (as one might expect) and in some cases, the MDS model predictions were even worse than the CB model predictions (S2 and S4 for example). The MD model accounted for more variance in the subjects' detection patterns than any other model tested here. Davidson *et al.* (2006) showed that the MD model weights significantly raised r^2 values for the CB model greater than would be expected by simply adding free parameters. Here it is of interest to determine whether the MD model

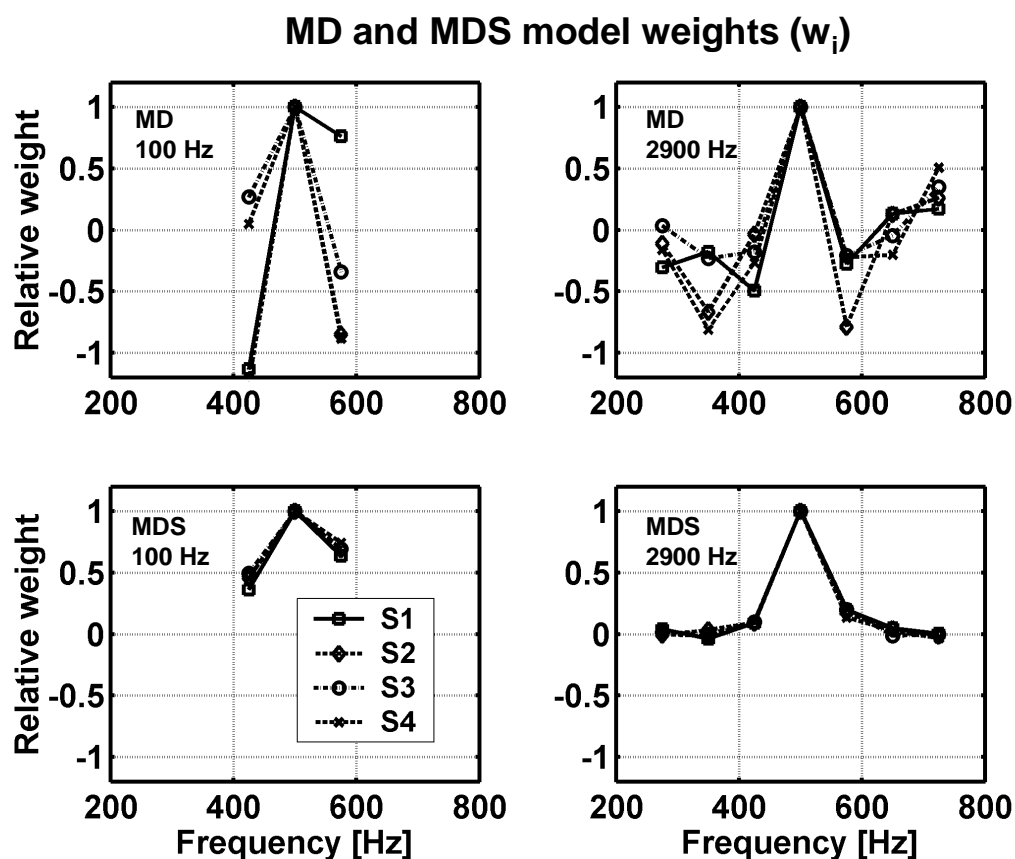


Figure 3-7. Weights computed for MD and MDS models for the 100-Hz and 2900-Hz data in study 2. Weights are shown for the 4 subjects. Note that weights in each condition were normalized to the maximum weight (occurring at 500 Hz). These weights correspond to the w_i for the MD and MDS models in Fig. 3-1 and Eq. (3).

predictions were significantly better than the MDS model predictions. For all subjects but S3 [see Fig. 3-4, $P(Y|N)$], the MD model made better predictions than the MDS model. Tests of significant differences between non-independent correlations were computed for each subject in each bandwidth to test the hypothesis that the MD model predicts detection patterns significantly better than the MDS model. Results indicate that the MD model predicted significantly ($p < 0.05$) more variance in $P(Y|T+N)$ for S2, S3 and S4 in the 100-Hz bandwidth, and S1 and S4 in the 2900-Hz bandwidth; and more variance in $P(Y|N)$ for S1 and S2 in the 100-Hz bandwidth and S2 in the 2900-Hz bandwidth. Thus, the MD weighting strategy does, for some subjects, make significantly better predictions than the MDS weighting strategy.

3.3.1.1.3 Envelope-slope model

The ES model is a temporal model based entirely on the stimulus envelope. It was of interest to test this model under equal energy-conditions in comparison to conditions with energy differences from stimulus waveform to stimulus waveform. It was also of interest to see how this model compared to other (temporal) models in cases in which energy differences were not present across the reproducible stimuli. In general, the ES model predictions were mediocre, with only 36 of the 52 predictions reaching significance for $P(Y|T+N)$ (Fig. 3-3) and only 20 of the 52 predictions reaching significance for $P(Y|N)$ (Fig. 3-5). Predictions were highly variable across subjects in all studies, except for $P(Y|T+N)$ in the wideband condition of study 2. Despite this model's simplicity, it was able to account for more variance in some of the subject's detection patterns than some of the more complicated temporal models (i.e., DA, BR, PO, see below), particularly in the equal-energy cases of studies 3 and 4 for $P(Y|N)$ (see Figs. 3-4 and 3-6). Note that the ES model predictions shown here explain less variance than those in Davidson *et al.* (2006), due the use of separate predictions for $P(Y|T+N)$ and $P(Y|N)$ in the present study.

3.3.1.1.4 Dau model

Like the ES model, the Dau model relies primarily on the temporal envelope of the stimulus waveform, but the Dau model does allow some fine structure to pass on to the decision device. This model uses a distinct template-matching strategy, such that stimulus waveforms more similar to the internal template (computed as normalized version of the difference between a T+N and a N template) will be selected as containing the tone. Figure 3-8 shows representative T+N and N templates (top panel) and the normalized difference template (bottom panel) for the 100-Hz condition of study 2. Each trace in the top panel shows the output of the adaptation loops averaged over 500 stimulus waveforms. It is clear that the averaging brings out some fine-structure information related to the tone frequency in the T+N template. This information is effectively increased by the normalization process with respect to that of the difference of the two templates (bottom panel). The model decision variable is computed as the mean (over time) product of the internal representation of an individual stimulus and the normalized difference template. This difference is largest at the onset of the noise waveform due to the lack of compression in the adaptation loops for stimuli with fast changes in SPL (i.e., at the onset, whereas the latter portion of the difference is compressed). The operation of the detector indicates that the Dau decision variables should be at least partially correlated to overall energy, and thus should be impacted by energy-equalization of the stimuli (which will be discussed in detail in the following sections). The co-variation in time of the fine-structure present in the stimulus waveform and in the internal template also contributes to the decision variable.

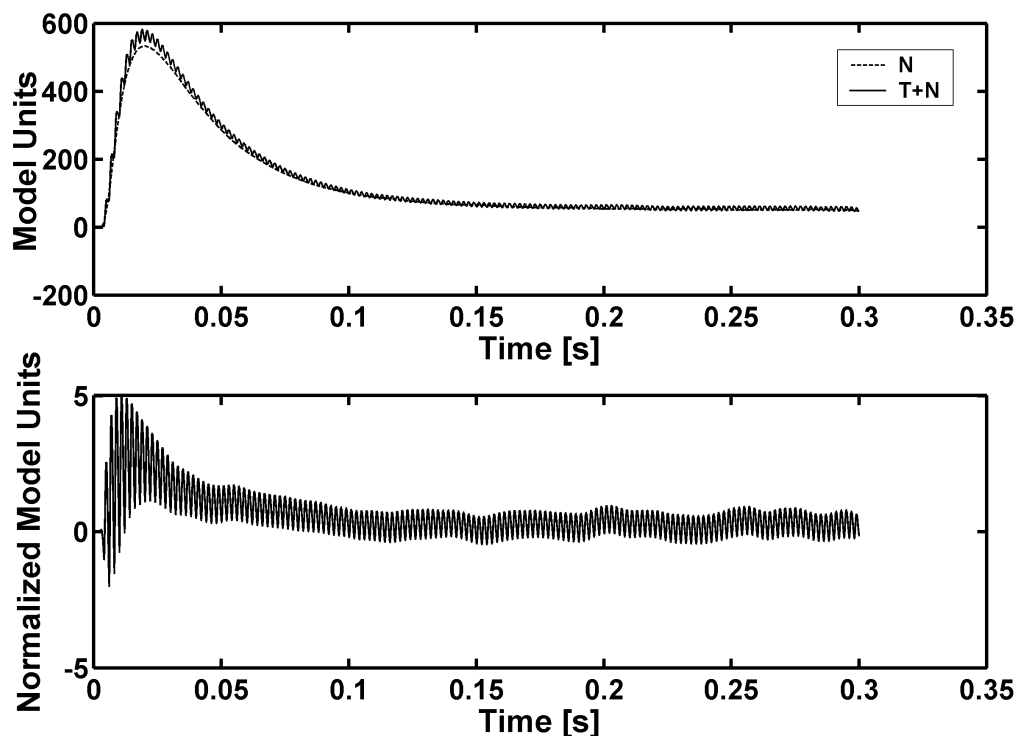


Figure 3-8. Templates used in the DA model decision device. The T+N and N templates were computed as the mean of 500 internal representations of T+N and N stimuli respectively. The lower panel shows the normalized difference between the two templates in the upper panel.

In general, the predictions for the Dau model were not significantly correlated to the subjects' detection patterns. Only 16 of the 54 $P(Y|T+N)$ predictions and only 6 of the 54 $P(Y|N)$ predictions reached significance. This model will be revisited below in the context of its energy dependence.

3.3.1.1.5 Breebaart model

The peripheral processing in the Breebaart model is similar to that of the Dau model. Any difference between the model predictions of the Breebaart and Dau models must result from the differences in the decision devices (including template mechanisms). Representative templates of the Breebaart model are shown in Fig. 3-9 for the 100-Hz condition of study 2 for the 3 frequency channels used. Recall that the noise-alone template was subtracted from the internal representation for each T+N stimulus waveform as a measure of the "distance" from the noise-alone stimulus. The frequency weighing computed in Eq. (6) is shown in Fig. 3-9. The wideband weights are similar to those shown in Fig. 3-7 for the MDS model, while the narrowband Breebaart weights are similar to the MDS weights. This model produced only 16 significant predictions of the 54 made for $P(Y|T+N)$ and only 1 significant prediction of the 54 made for $P(Y|N)$. Thus, this model's decision device seems to be incapable of predicting the subjects' detection patterns.

3.3.1.1.6 Phase-opponency model

The PO model's ability to predict detection patterns was on a par with the ES model, with 23 of the 54 $P(Y|T+N)$ predictions reaching significance and only 11 of the

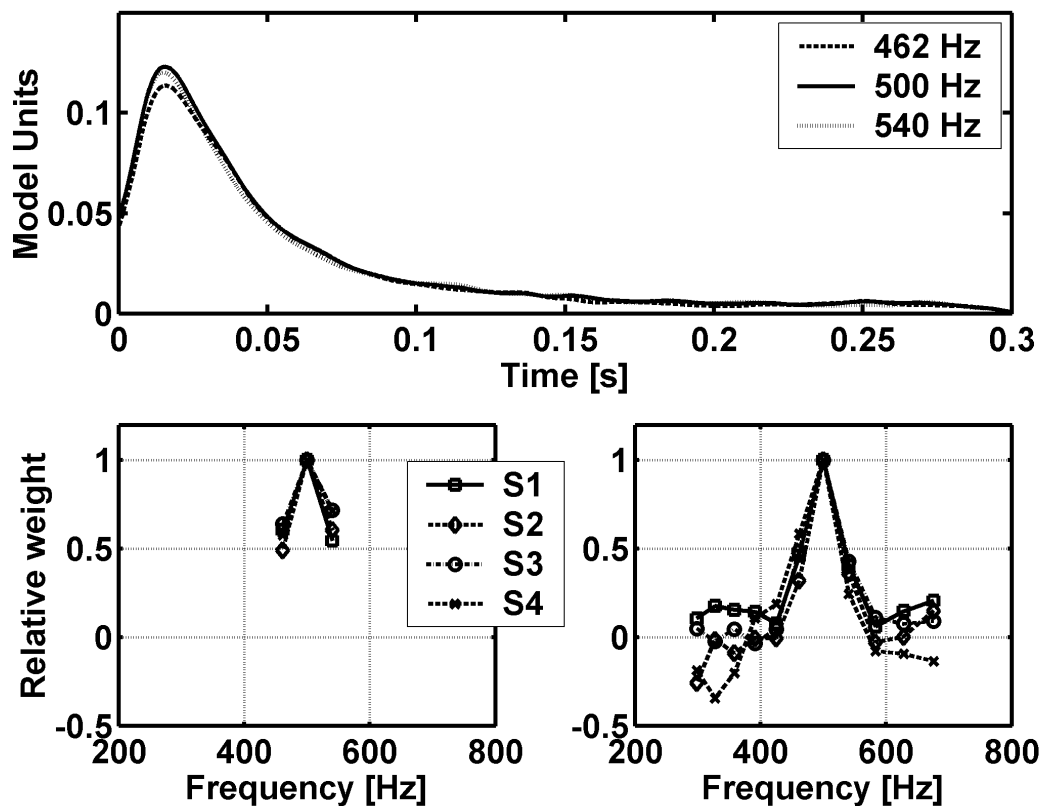


Figure 3-9. Internal noise-alone templates (top panel) for the different frequency channels in the narrowband condition of study 2, and frequency weighting (bottom two panels) imposed by the Breebaart *et al.* (2001) model on the internal representations of the stimulus waveforms. The frequency weights were summed over time and normalized to the peak value for the 100 and 2900-Hz conditions of study 2 for the left and right panels respectively.

54 P(Y|N) predictions reaching significance. This model performed no better under cases for which energy was equalized than in cases with energy cues present. Its relation to envelope, fine-structure, and energy will be examined below, along with the other models tested.

3.3.1.2. Comparisons between models

Although none of the models was able to predict a significant proportion of the variance in subjects' detection patterns in all conditions, it is still useful to determine how similar or different each models' decision variables are to any other model's decision variables. Because the models were "run" at different tone levels for each subject, the decision variables varied slightly. In order to simplify comparisons between models, and because the signal-to-noise ratios were within 3 dB for all subjects within each study, a single signal-to-noise ratio (the median signal-to-noise ratio for each study) was selected. Inter-model correlations are presented in terms of r^2 in Tables 3-1 and 3-2 for P(Y|T+N) and P(Y|N), respectively, for the each of the models in Fig. 3-1. Blank values indicate conditions for which models did not apply.

Tables 3-1 and 3-2 show that the CB model was significantly correlated to each of the diotic models tested in this study. The most highly correlated models were the CB and MDS models with r^2 values on the order of 0.91 to 0.98. These models were also significantly

correlated to the MD model, which is not surprising given that each of the three used energy at the output of one or more critical bands as decision variables. The CB and ES models were also significantly, albeit weakly, correlated. The CB, MD and MDS models were all significantly correlated to the Dau and the BR models (for conditions that had differences in level across stimulus waveforms), as expected given the Dau and BR models' envelope dependences and detection strategies. Finally, the Dau and BR models were correlated for every condition in every study. The PO and ES models were correlated for all studies as well.

3.3.1.3 Effect of stimulus energy

The contribution of stimulus energy to each of the model decision variables was tested with a multiple regression approach, using two models to predict the listeners' data. The CB model was always one of the 2 predictor models. An incremental F test was used to determine if the addition of the second model (when added as a predictor to the CB model), significantly increased the proportion of predicted variance. This procedure was equivalent to testing the significance of the partial correlation coefficient or the significance of the slope of a predictor variable in a multiple regression analysis. Results are briefly summarized in the text below in terms of the increase in R^2 (the proportion of variance explained by a multiple regression including the CB model as a predictor) achieved by adding the second model to the CB model for both $P(Y|T+N)$ and $P(Y|N)$. Of all the diotic models tested [312 tests were run in total for $P(Y|T+N)$ and 312 for $P(Y|N)$; 6 models x 1 study with 4 subjects and 2 conditions x 1 study with 5 subjects and 4 conditions and 1 study with 6 subjects and 4 conditions], only the MD, ES, and PO models yielded significantly better predictions when added to the energy model as a second predictor, confirming that the variance explained by the MDS, Dau, and BR models "overlaps" with the variance already explained by the CB model. Significant increases in R^2 values by the addition of the MD model as a predictor were in the range of 0.10 to 0.33 for T+N stimuli and 0.10 to 0.36 for N stimuli, depending on subject. Significant increases by adding the ES model were in the range of 0.10 to 0.32 for T+N stimuli and of 0.10 to 0.52 for N stimuli, depending on subject. Significant increases resulting from adding the PO model were in the range of 0.08 to 0.46 for T+N stimuli and 0.16 to 0.21 for N stimuli.

3.3.1.4. Between-condition comparisons of model decision variables for study 4

Correlations were computed in terms of r^2 between model decision variables computed in the E_1C_1 , E_2C_2 , E_1C_2 , and E_2C_1 conditions to test whether each model operated primarily on envelope, fine structure, or both features of the stimulus waveforms. Results are presented in Table 3-3. Cues dominated by envelope will show high correlations between the "E" conditions with the same subscripts. Cues dominated by fine structure will show high correlations between "C" conditions with the same subscripts. Almost every model relied more heavily on stimulus features related to the envelope for T+N stimuli. The comparison was less clear for noise-alone stimuli, for which the distortion control algorithm (see Ch. 2 Methods) may have allowed some correlation between envelope and fine structure. (This correlation was removed in the analyses of Ch. 2 by the statistical blocking procedure). Inspection of Table 3-3 indicates that, with the exception of energy-based models (which were not included because all waveforms had the same overall level), the ES, DA, BR models all rely heavily on envelope, while the PO model relies on both envelope and fine structure.

Model Comparison	Study 2		Study 3				Study 4			
			RNRE	LNRE	RNEE	LNEE	E ₁ C ₁	E ₂ C ₂	E ₂ C ₁	E ₁ C ₂
CB - MD	0.78*	0.70*								
CB - MDO	0.95*	0.98*								
CB - ES	0.33*	0.56*	0.15	0.34*	0.32*	0.86*	0.22*	0.51*	0.07	0.12
CB - DA	0.00	0.09	0.60*	0.77*	0.00	0.13	0.04	0.00	0.17*	0.01
CB - BR	0.10	0.28*	0.90*	0.54*	0.18*	0.03	0.01	0.00	0.11	0.09
CB - PO	0.01	0.2*	0.22*	0.34*	0.33*	0.28*	0.23*	0.3*	0.14	0.10
MD - MDO	0.79*	0.65*								
MD - ES	0.34*	0.55*	0.15	0.34*	0.32*	0.86*	0.22*	0.51*	0.07	0.12
MD - DA	0.00	0.07	0.60*	0.77*	0.00	0.13	0.04	0.00	0.17*	0.01
MD - BR	0.07	0.3*	0.90*	0.54*	0.18*	0.03	0.01	0.00	0.11	0.09
MD - PO	0.00	0.21*	0.22*	0.34*	0.32*	0.28*	0.23*	0.3*	0.14	0.10
MDO- ES	0.21*	0.5*	0.15	0.34*	0.32*	0.86*	0.22*	0.51*	0.07	0.12
MDO- DA	0.00	0.08	0.60*	0.77*	0.00	0.13	0.04	0.00	0.17*	0.01
MDO- BR	0.09	0.25*	0.90*	0.54*	0.18*	0.03	0.01	0.00	0.11	0.09
MDO- PO	0.00	0.19*	0.22*	0.34*	0.32*	0.28*	0.23*	0.3*	0.14	0.10
ES - DA	0.00	0.05	0.04	0.36*	0.04	0.12	0.11	0.00	0.16*	0.00
ES - BR	0.01	0.33*	0.31*	0.15	0.04	0.05	0.03	0.00	0.03	0.03
ES - PO	0.34*	0.51*	0.49*	0.73*	0.54*	0.46*	0.49*	0.41*	0.46*	0.55*
DA - BR	0.31*	0.38*	0.64*	0.9*	0.37*	0.57*	0.86*	0.81*	0.74*	0.89*
DA - PO	0.00	0.13	0.10	0.46*	0.00	0.01	0.26*	0.24*	0.36*	0.22*
BR - PO	0.05	0.2*	0.40*	0.27*	0.08	0.09	0.18*	0.26*	0.22*	0.09

* $p < 0.05$

Table 3-1. Correlations between model decision variables for T+N trials for all N₀S₀ data. Correlations are given in units of r^2 . Model abbreviations are the same as listed in Fig. 3-1. Significant correlation ($p < 0.05$) are marked with *.

S

Model Comparison	Study 2		Study 3				Study 4			
			RNRE	LNRE	RNEE	LNEE	E ₁ C ₁	E ₂ C ₂	E ₂ C ₁	E ₁ C ₂
CB - MD	0.57*	0.56*								
CB - MDO	0.91*	0.97*								
CB - ES	0.20*	0.44*	0.00	0.00	0.17*	0.35*	0.30*	0.23*	0.06	0.01
CB - DA	0.40*	0.34*	0.46*	0.94*	0.03	0.04	0.00	0.02	0.05	0.05
CB - BR	0.41*	0.08	0.83*	0.91*	0.00	0.15	0.01	0.03	0.10	0.13
CB - PO	0.11	0.37*	0.00	0.02	0.10	0.01	0.03	0.03	0.00	0.00
MD - MDO	0.60*	0.50*								
MD - ES	0.22*	0.35*	0.00	0.00	0.17*	0.35*	0.3*	0.23*	0.06	0.01
MD - DA	0.35*	0.35*	0.46*	0.94*	0.03	0.04	0.00	0.02	0.05	0.05
MD - BR	0.20*	0.20*	0.83*	0.91*	0.00	0.15	0.01	0.03	0.10	0.13
MD - PO	0.05	0.33*	0.00	0.02	0.10	0.01	0.03	0.03	0.00	0.00
MDO- ES	0.05	0.37*	0.00	0.00	0.17*	0.35*	0.3*	0.23*	0.06	0.01
MDO- DA	0.33*	0.27*	0.46*	0.94*	0.03	0.04	0.00	0.02	0.05	0.05
MDO- BR	0.41*	0.05	0.83*	0.91*	0.00	0.15	0.01	0.03	0.10	0.13
MDO- PO	0.01	0.31*	0.00	0.02	0.10	0.01	0.03	0.03	0.00	0.00
ES - DA	0.20*	0.22*	0.00	0.00	0.02	0.02	0.01	0.00	0.00	0.03
ES - BR	0.03	0.32*	0.07	0.00	0.02	0.16*	0.04	0.03	0.01	0.02
ES - PO	0.78*	0.58*	0.61*	0.43*	0.63*	0.45*	0.58*	0.29*	0.56*	0.39*
DA - BR	0.35*	0.47*	0.63*	0.96*	0.73*	0.09	0.81*	0.78*	0.69*	0.88*
DA - PO	0.12	0.35*	0.04	0.03	0.12	0.00	0.23*	0.01	0.05	0.02
BR - PO	0.02	0.34*	0.01	0.04	0.00	0.02	0.34*	0.02	0.11	0.06

* $p < 0.05$

Table 3-2. Correlations between model decision variables for N trials for all N₀S₀ data. Correlations are given in units of r^2 . Model abbreviations are the same as listed in Fig. 3-1. Significant correlation ($p < 0.05$) are marked with *.

Table 3-3. Correlations between N_0S_0 model decision variables (in terms of r^2) for the various conditions of study 4, separated by T+N and N stimuli, and model. Model abbreviations are as in Fig. 3-1. Significant correlations are shown with *. Note that energy-based models were omitted from this analysis; all stimulus waveforms had the same overall level.

Model		P(Y T+N)			P(Y N)		
		E_2C_2	E_1C_2	E_2C_1	E_2C_2	E_1C_2	E_2C_1
ES	E_1C_1	0.04	0.95*	0.02	0.18*	0.93*	0.33*
	E_2C_2		0.01	0.98*		0.28*	0.90*
	E_1C_2			0.01			0.40*
DA	E_1C_1	0.00	0.91*	0.00	0.11	0.69*	0.17*
	E_2C_2		0.02	0.92*		0.30	0.83*
	E_1C_2			0.01			0.18*
BR	E_1C_1	0.00	0.96*	0.01	0.20*	0.89*	0.18*
	E_2C_2		0.01	0.92*		0.25*	0.92*
	E_1C_2			0.01			0.19*
PO	E_1C_1	0.00	0.60*	0.00	0.59*	0.71*	0.75*
	E_2C_2		0.13	0.94*		0.80*	0.83*
	E_1C_2			0.11			0.66*

* $p < 0.05$

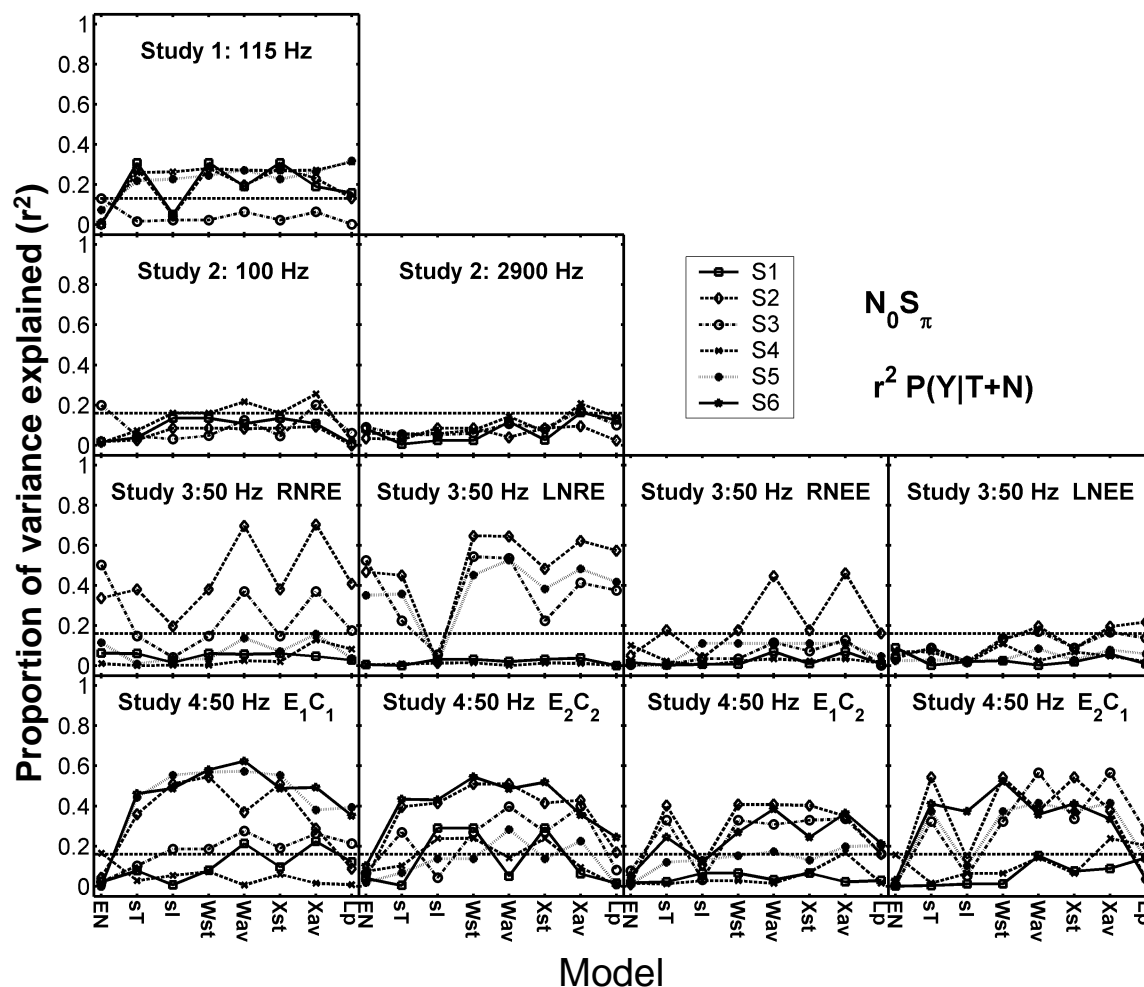


Figure 3-10. Proportion of variance explained by the $N_0 S_\pi$ model predictions for selected Isabelle (1995) and Goupell (2005) decision variables for z scores of $P(Y|T+N)$. EN is the RMS energy of the right stimulus waveform, sT is the standard deviation of ITDs, sI is the standard deviation of ILDs, Wst is a linear combination of the standard deviations ITDs and ILDs, Wav is a linear combination of the average value of ITDs and ILDs, Xst is the standard deviation of a linear combination of ITDs and ILDs, Xav is the average value of a linear combination of ITDs and ILDs, and Lp is a lateral position model relating ITDs and ILDs with a trading ratio. Models are described in Eq. (10-16). Studies 1-4 correspond to rows 1-4, with each column representing a different condition within each study. Different subjects are indicated with different symbols connected across models to facilitate comparisons between models. Note that subject identification numbers only correspond to the same subjects within studies. Model predictions were quantified with the square of the correlation coefficient (r^2). Predictions above the horizontal-dashed line are significant ($p < 0.05$).

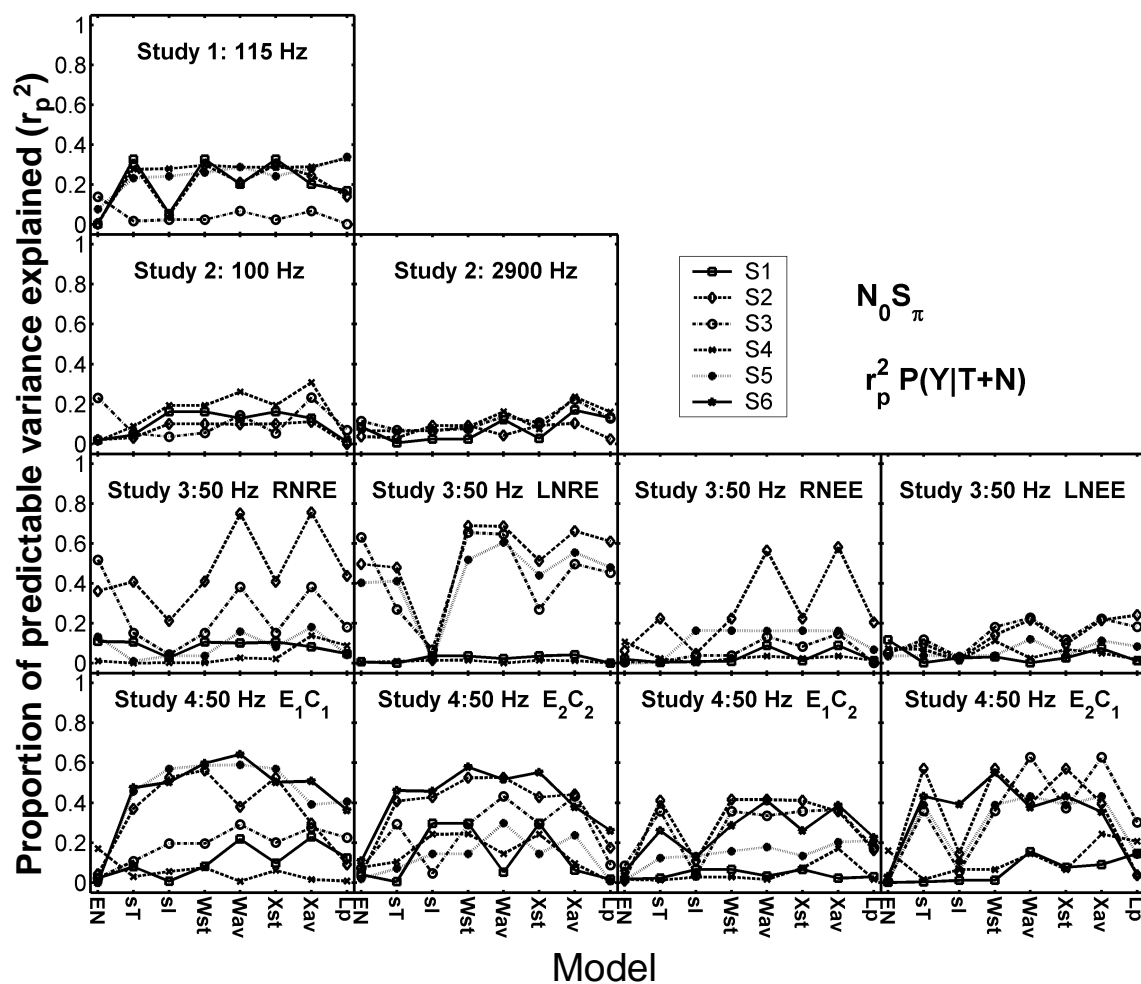


Figure 3-11. Same as Fig. 3-10 except predictions were quantified using the square of the correlation coefficient normalized by the proportion of predictable variance ($r_p^2 = r^2/V$). Model abbreviations from left to right are the same as in Fig. 3-10. Note that subjects sharing the same number do not correspond across studies.

3.3.2 Dichotic models

Predictions for each model are shown in two sets of figures. Figure 3-10 shows results for $P(Y|T+N)$ for the Isabelle (1995) and Goupell (2005) decision variables. The critical r^2 value for reaching a significant prediction is 0.14 for study 1 and 0.16 for studies 2, 3 and 4, as denoted by the horizontal-dashed line. As was done for the diotic model results, the dichotic predictions are re-plotted in Fig. 3-11 in terms of r_p^2 , or the square of the correlation coefficient (r^2) normalized by the proportion of predictable variance (V) for each subject within each condition in each study. Note that r_p^2 results are not suitable for statistical testing. Additionally, if only a small proportion of the variance in the detection pattern is estimated to be predictable, r_p^2 may exceed 1. The reader is reminded that each subject in each study is denoted using a different symbol and that identical symbols do not correspond to the same subjects across studies (but do correspond to the same subjects within studies).

3.3.2.1 Individual model results

3.3.2.1.1 Isabelle (1995) and Goupell (2005) decision variables

The Isabelle (1995) and Goupell (2005) decision variables are based on interaural differences calculated directly from the stimulus waveforms. No predictions are shown for $P(Y|N)$ because the N stimuli were diotic (i.e., had no interaural differences) and internal noise was not included in this analysis, resulting in decision variables that were identically zero. Figures 3-10 and 3-11 show predictions for r^2 and r_p^2 respectively⁴. The EN decision variable (RMS energy of the right stimulus waveform) performed poorly in general, with only 8 of the 57 predictions reaching significance. These findings are consistent with those of the original Isabelle (1995) study, and indicate that the cue used by listeners to perform the detection task was not correlated to energy. Standard deviations of ITDs and ILDs (sT and sI) performed better, with 20 and 11 of the 57 predictions reaching significance respectively. Linear combinations of the standard deviations or average values of ITDs and ILDs performed best, consistent with the results of Ch. 2, suggesting that both envelope and fine-structure contribute to the detection process. Of the 57 predictions, 23 using the Wst decision variable reached significance and 29 using the Wav decision variable reached significance. Predictions for models that *first* computed decision variables *and then* combined across ITD and ILD processors (Wst and Wav) accounted for about the same amount of variance in $P(Y|T+N)$ as those that combined ITDs and ILDs as a function of time *before* decision variables were computed (Xst, Xav, and Lp). The Xst, Xav and Lp decision variables made 26, 36, and 22 significant predictions respectively for the 57 comparisons performed. Summarizing, the “separate centers” (Wst and Wav) and “auditory-image” (Xst and Xav) decision variables had about the same predictive power. It is of interest to view the weights placed on ITD or ILD decision variables to examine possible trends across subjects, and to determine if any relationship exists between threshold tone level and weight selection. Figure 3-12 shows weights (a and b) used in Eqs. 12-15, organized by model and subject for the four studies. Recall that weights were bounded by 0 and 1, with 1 indicating

⁴ These decision variables were also computed using 4th-order gammatone filters centered at 500 Hz. However, in all cases but the 2900-Hz case, predictions were poorer when peripheral filtering was used, and therefore, those results were not included in this document. These decision variables were also tested using the Heinz *et al.* (2001) and Zilany *et al.* (2006) auditory-nerve models. Poor results (i.e., worse than those achieved with no peripheral processing) were also encountered using the auditory-nerve models as a peripheral processing stage, but this was likely due to the fact that these decision variables rely on the complex-analytic signal, which is not well defined for the output of the auditory-nerve models (the outputs of which have nonzero dc components). Therefore, the Heinz *et al.* (2001) predictions are not shown.

reliance on ITD and 0 indicating reliance on ILD. Figure 3-12 shows that the favorable results of these models were largely due to their ability to select the better decision variable from sT or sI. Certain subjects, however, did use a true weighted combination of the two decision variables (S4 in study 1; S2, S3, and S5 in study 3; and S3 in study 4). In almost all cases, these subjects were subjects with lower thresholds. Subjects with higher thresholds were fit more reliably with weights of 0 or 1, indicating that they relied on only ITDs or ILDs.

3.3.2.1.2 Four-channel model

Results for the four-channel model are shown in terms of the square of the correlation coefficient (r^2) in Figs. 3-13 and 3-14 for $P(Y|T+N)$ and $P(Y|N)$ respectively, and estimates of the proportion of predictable variance explained (r_p^2) are shown in Figs. 3-15 and 3-16 for $P(Y|T+N)$ and $P(Y|N)$ respectively. Two versions of this model were implemented. The basic structure of both versions is shown in Fig. 3-2 (FC). The first (FCc) used a cross correlation (product) of the inputs for the “peaker” channels (weighted by w_2 and w_3 in Fig. 3-2), while the second used a normalized cross correlation (Colburn *et al.* 1997) for the “peaker” channels. Weights derived for each of the four channels are shown in Fig. 3-17. The weights were normalized separately for w_2 and w_3 and for by w_1 and w_4 for display purposes only (in order to eliminate the difference in absolute value of the output from two types of channels). As expected, very few of the weights took on a value of 0, as each channel was tuned in increments of 90° of interaural phase. This model performed poorly in general, with predictions reaching significance for only 3 of the 114 comparisons (including both FCc and FCn) for $P(Y|T+N)$. This model performed better for noise-alone stimuli; 22 of the 57 comparisons were significant for FCc and 14 of the 57 comparisons were significant for FCn.

3.3.2.1.3 Breebaart Model

Like the Isabelle and Goupell decision variables, the binaural version of the Breebaart model produced decision variables that were identically 0 for noise-alone stimuli due to the subtraction mechanism in Eq. 18, and was therefore omitted from Figs. 3-14 and 3-16 (recall the noise-alone template for this model would also be identically zero). Figure 3-18 shows representative temporal and spectral weights for the binaural version of the model computed for the two bandwidths of study 2. Note that the onset was weighted more heavily than the steady-state portion of the stimulus due to the action of the adaptation loops. This model produced significant predictions for 24 of the 57 comparisons to $P(Y|T+N)$, performing on a level comparable to Wst or Wav, despite its much more complex structure.

3.3.2.1.4 Mismatch model

The mismatch model predictions are shown in Figs 3-13 through 3-16 for $P(Y|T+N)$ and $P(Y|N)$. Model parameters were fit separately to $P(Y|T+N)$ or $P(Y|N)$. The resulting decision variables were compared to both $P(Y|T+N)$ and $P(Y|N)$, regardless of the condition of fit [i.e., the decision variables resulting from a fit to $P(Y|T+N)$ were also used to make predictions for $P(Y|N)$, and vice versa]. This model was implemented with three different binaural processors: a standard binaural cross correlation (MMc), a normalized cross correlation (MMn), and a subtraction mechanism (MMe). Correlations were compared for the various fit conditions to determine whether a single set of parameters was capable of predicting a significant proportion of the variance of both $P(Y|T+N)$ and $P(Y|N)$. The maximum number of significant predictions possible within $P(Y|T+N)$ or $P(Y|N)$ is 57. The MMc model fit to $P(Y|T+N)$ produced 52 significant

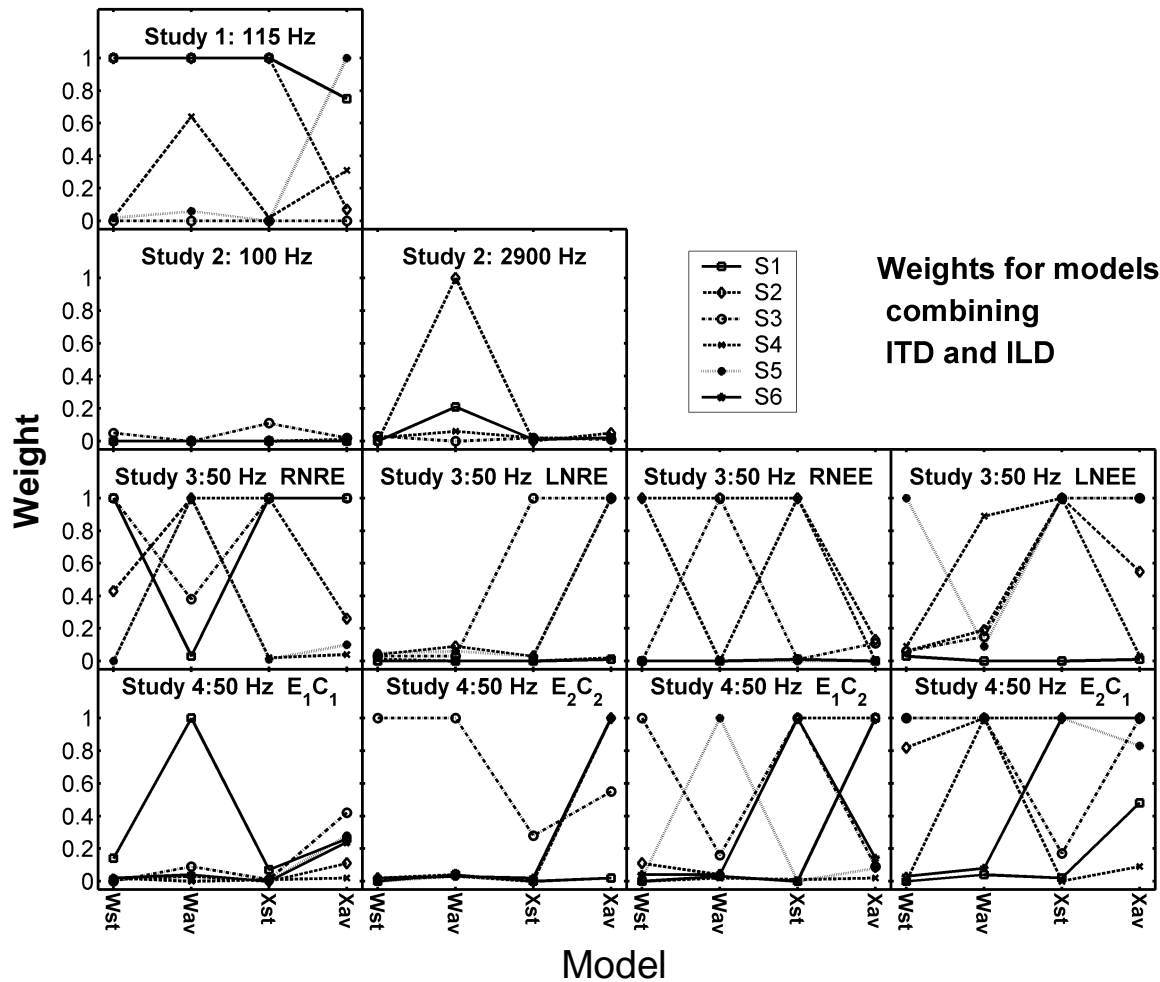


Figure 3-12. Model weights for decision variables based on both ITDs and ILDs. A weight approaching 1 indicates reliance on ITD and a weight approaching 0 indicates reliance on ILD. Note that subjects sharing the same number do not correspond across studies. Lines connecting subjects are present for comparison purposes only.

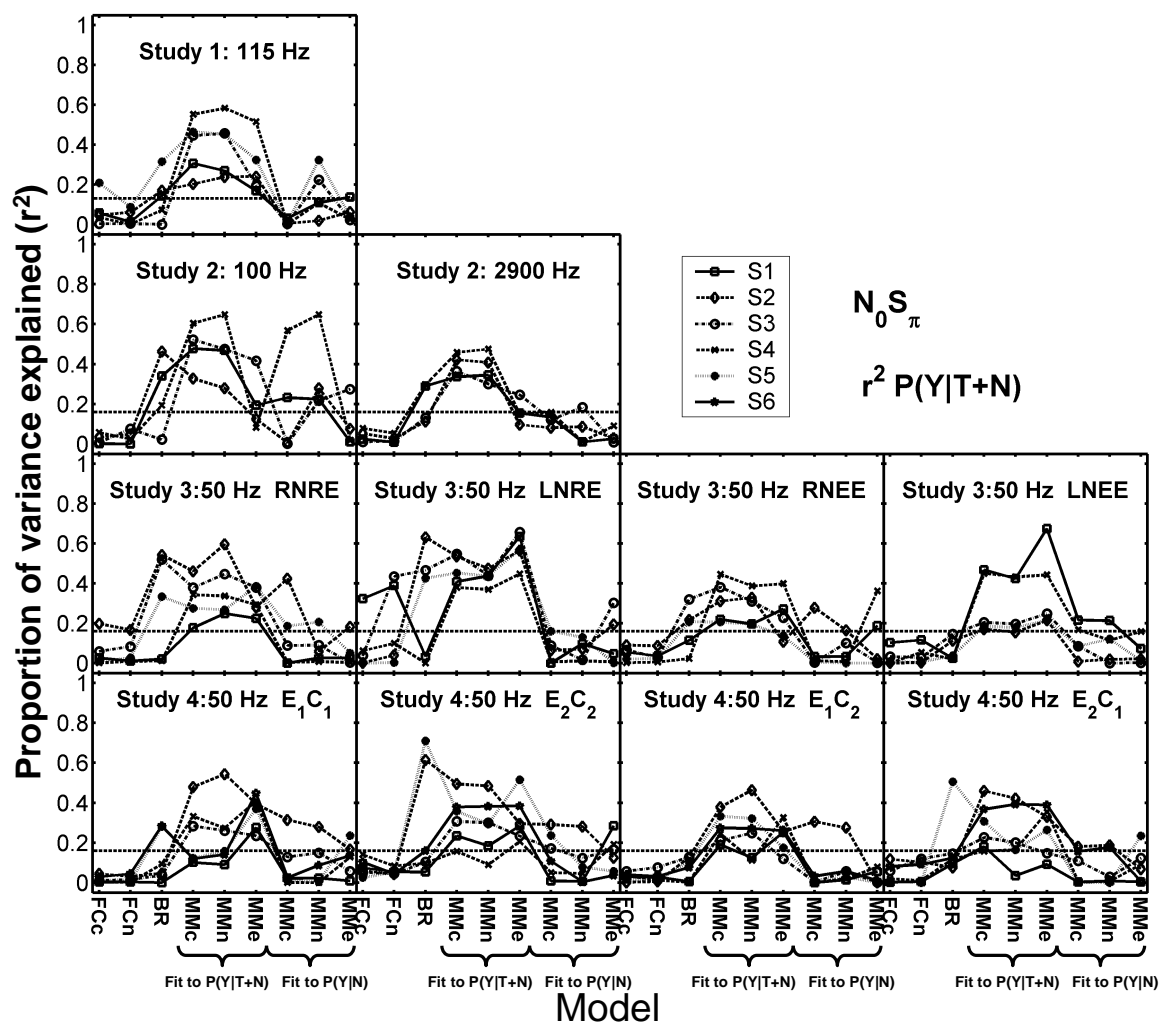


Figure 3-13. Proportion of variance explained by $N_0 S_\pi$ model predictions for z scores of $P(Y|T+N)$ quantified using the square of the correlation coefficient (r^2). Model abbreviations from left to right are: FcC, four-channel model using cross correlation for simulated peaker channels; FcN, four-channel model using a normalized cross correlation for peaker channels, BR, the Breebaart (2001) model. The mismatch (MM) models included (c) cross-correlation, (n) normalized-cross-correlation, and (e) subtraction mechanisms. The left 3 MM models were fit to $P(Y|T+N)$ data, while the right 3 were fit to $P(Y|N)$ data. Different subjects are indicated with different symbols connected across models to facilitate intersubject comparisons. Note that subjects sharing the same number do not correspond across studies.

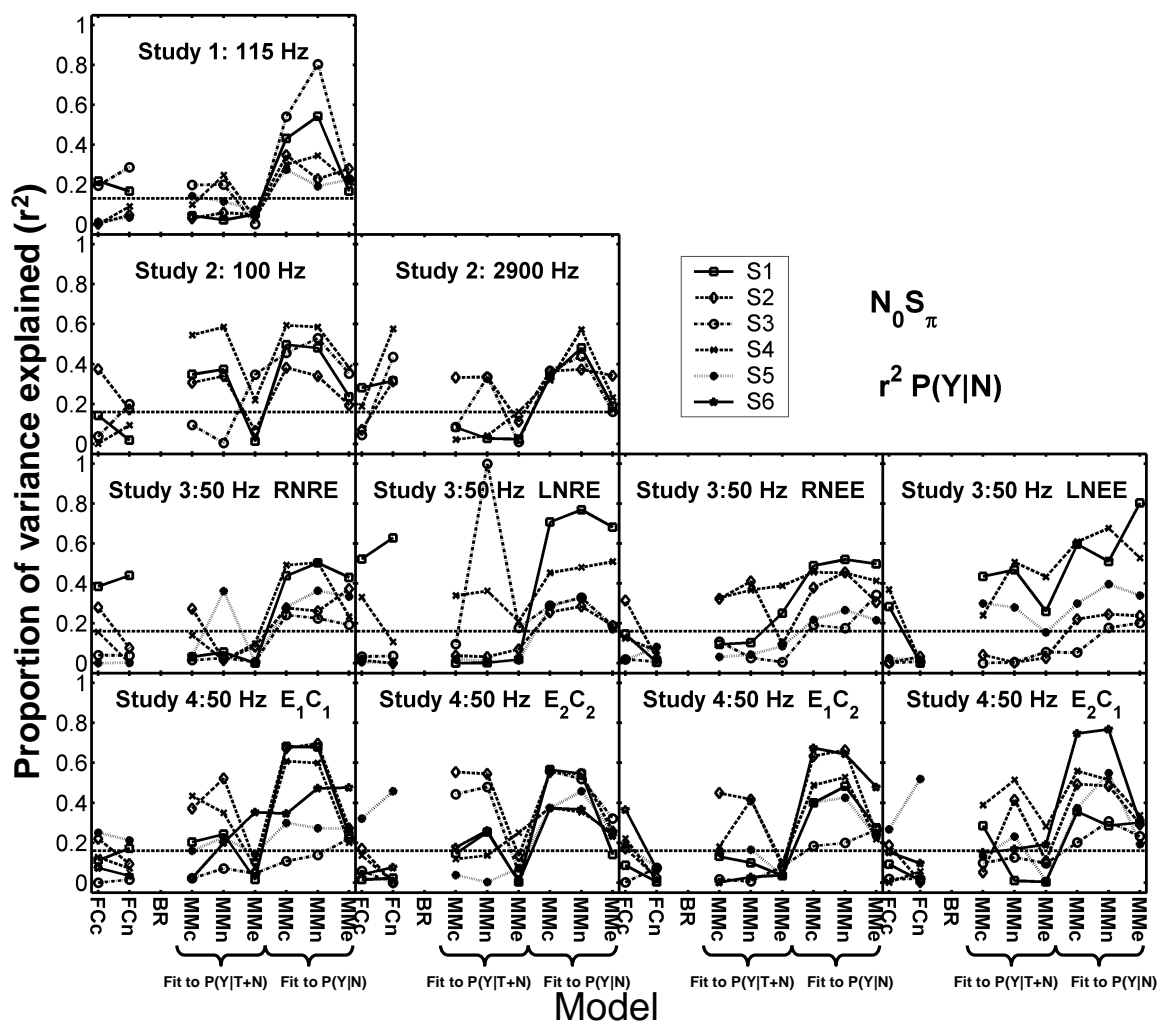


Figure 3-14. Proportion of variance explained by $N_0 S_\pi$ model predictions for z scores of $P(Y|N)$ quantified using the square of the correlation coefficient (r^2). Model abbreviations are as in Fig. 13. Note that subjects sharing the same number do not correspond across studies. Note that BR model predictions were identically 0, and therefore were omitted.

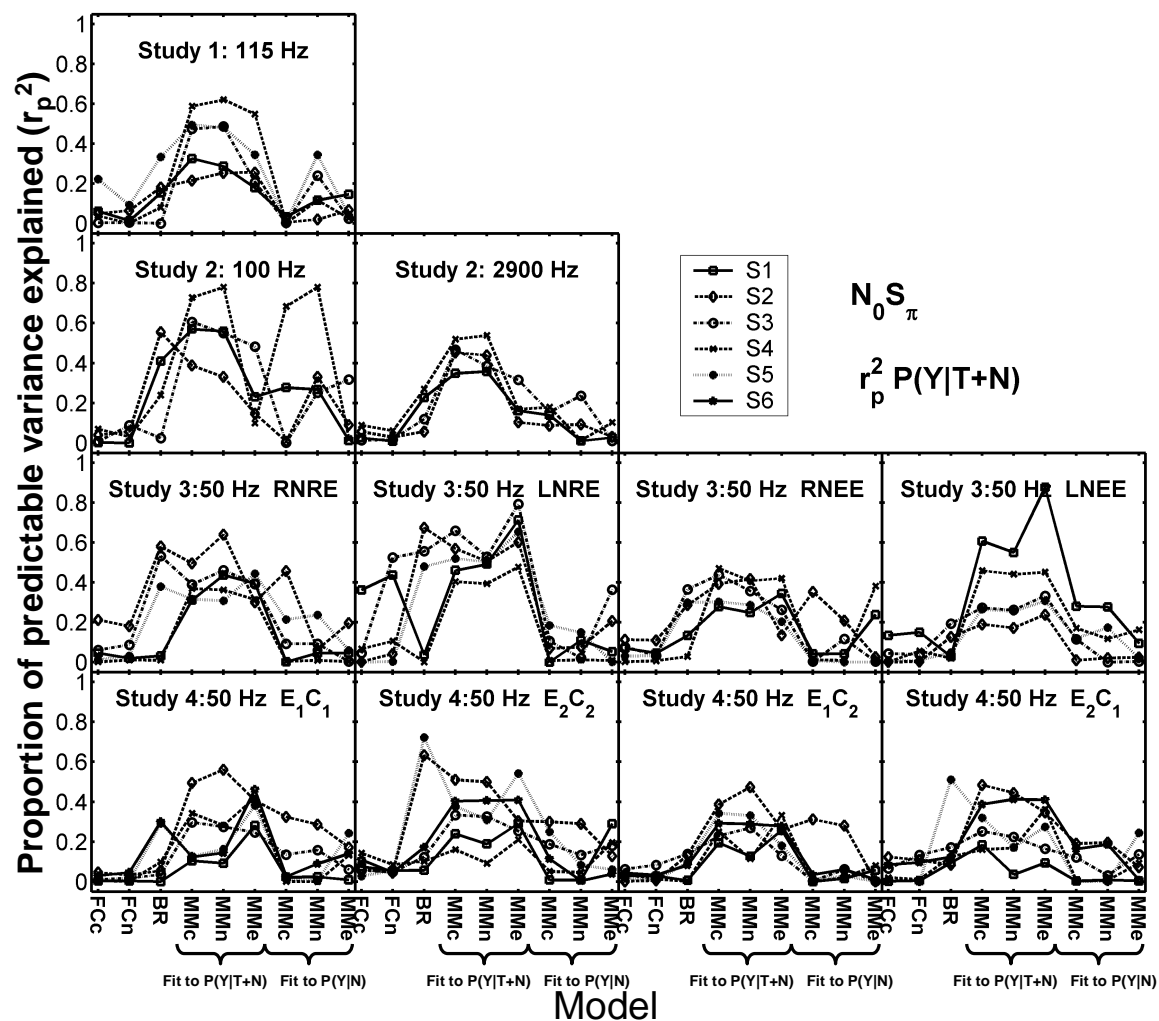


Figure 3-15. Proportion of variance explained by $N_0 S_\pi$ model predictions for $P(Y|T+N)$. Predictions were quantified using the square of the correlation coefficient normalized by the proportion of predictable variance ($r_p^2 = r^2/V$). Model abbreviations from left to right are the same as in Fig. 3-13. Note that subjects sharing the same number do not correspond across studies.

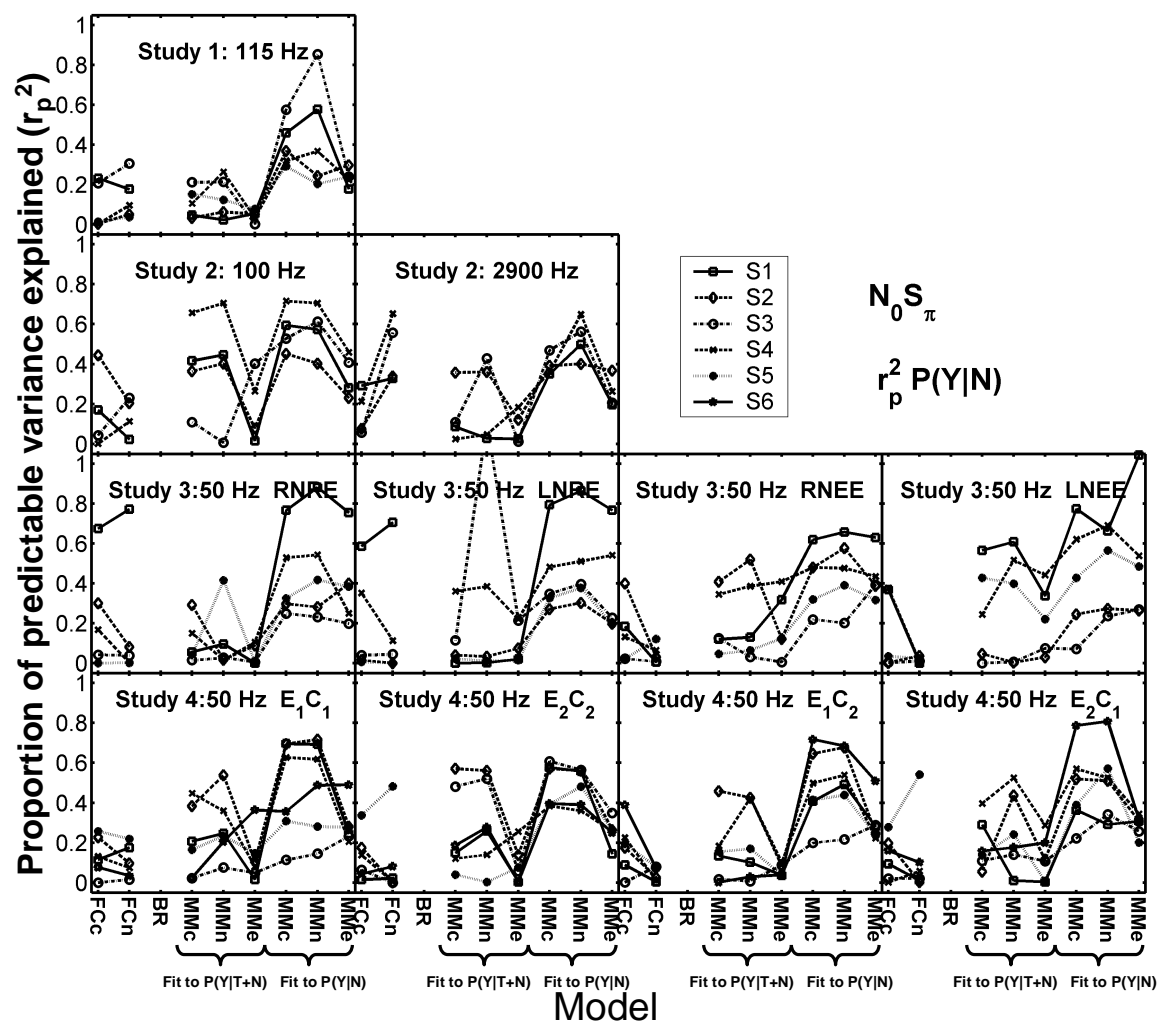


Figure 3-16. Proportion of variance explained by $N_0 S_\pi$ model predictions for $P(Y|N)$. Predictions were quantified using the square of the correlation coefficient normalized by the proportion of predictable variance ($r_p^2 = r^2/V$). Model abbreviations from left to right are the same as in Fig. 3-13. Note that subjects sharing the same number do not correspond across studies. Note that BR model predictions were identically 0, and therefore were omitted.

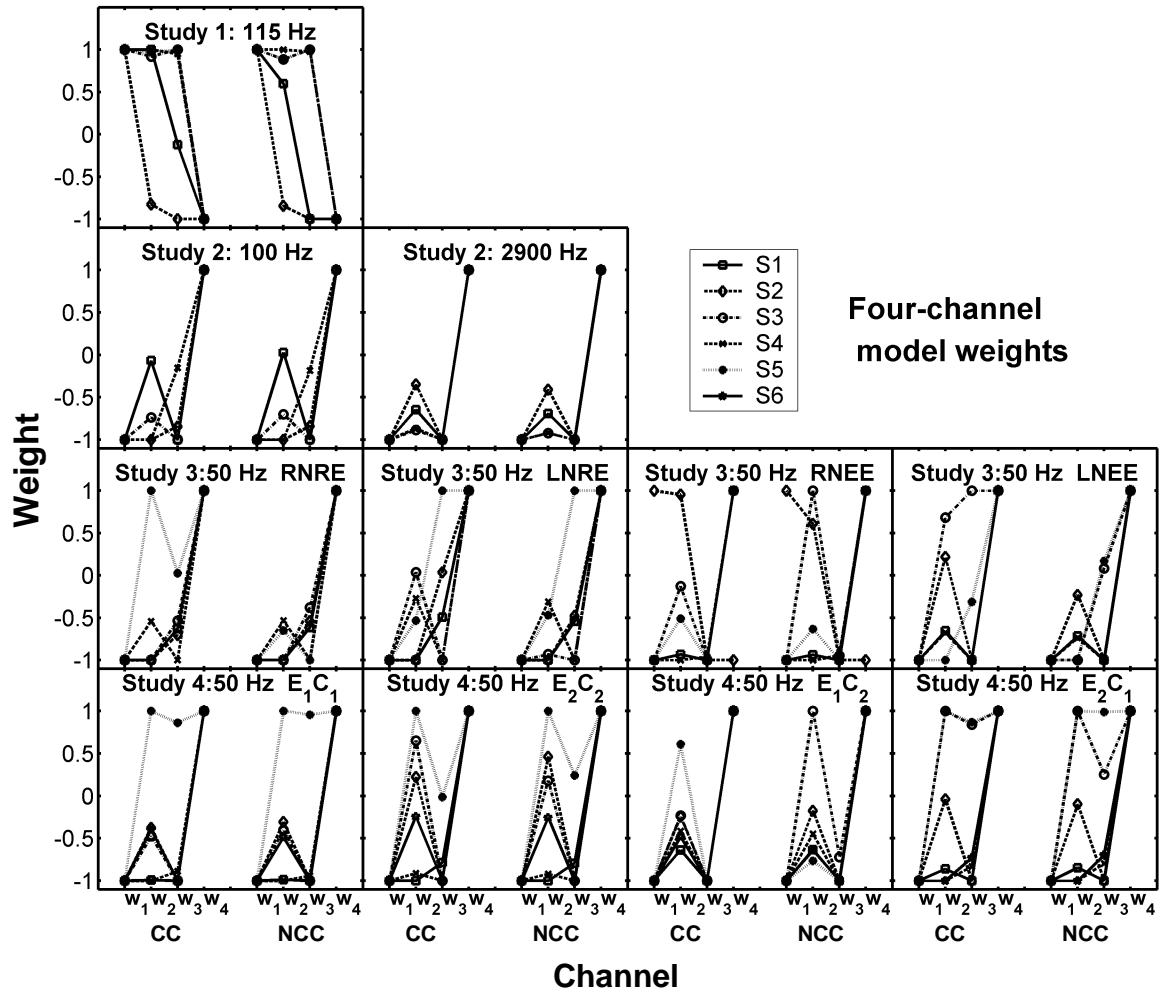


Figure 3-17. Weights derived for the individual channels of the four-channel model. Weights are shown for all channels using a cross correlation operation (CC) in channels 2 and 3, as well as for a using a normalized cross correlation for channels 2 and 3. Weights are virtually the same using either mechanism. Note that very few channels have weight values of 0.

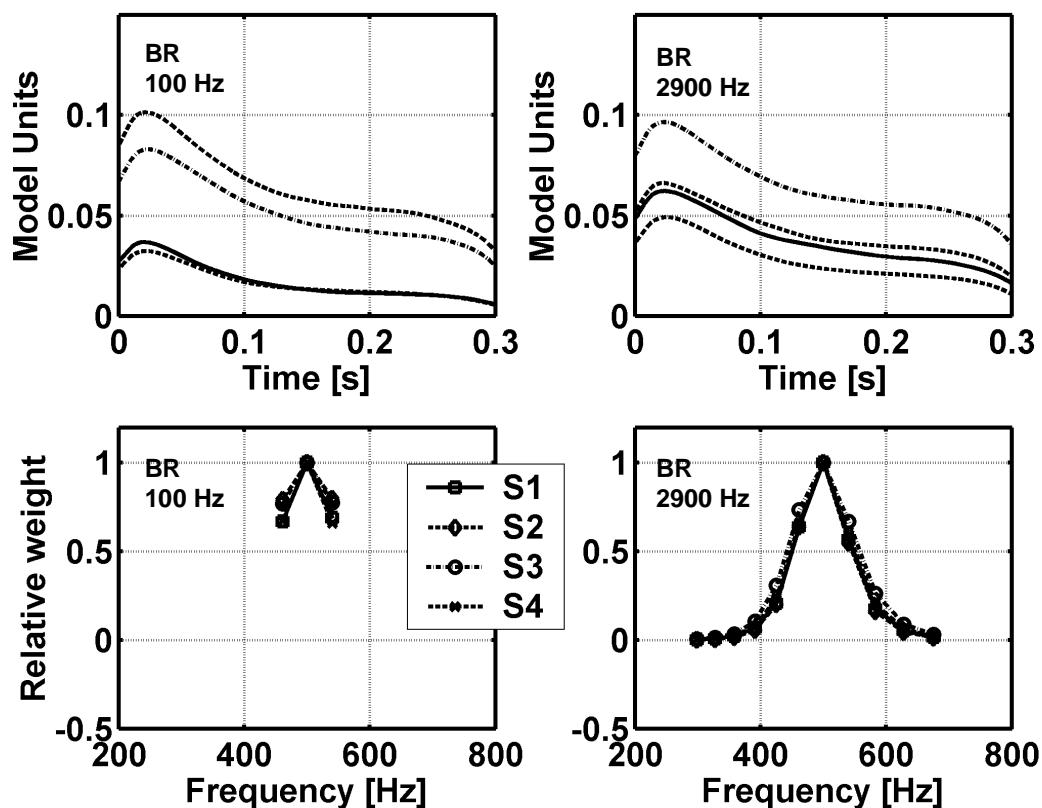


Figure 3-18. Representative Temporal weights (top) and spectral weights (bottom) from the Breebaart model. Note that the onset of the stimulus is weighted more heavily than the steady-state portion of the stimulus.

predictions for $P(Y|T+N)$ and 23 significant predictions for $P(Y|N)$. The same model fit to $P(Y|N)$ yielded only 14 significant predictions for $P(Y|T+N)$ and 55 significant predictions for $P(Y|N)$. The MMn model fit to $P(Y|T+N)$ produced 46 significant predictions for $P(Y|T+N)$ and 31 significant predictions for $P(Y|N)$. The same model fit to $P(Y|N)$ produced 15 significant predictions for $P(Y|T+N)$ and 56 significant predictions for $P(Y|N)$. The MMe model fit to $P(Y|T+N)$ produced 47 significant predictions for $P(Y|T+N)$ and only 13 significant predictions for $P(Y|N)$. The same model fit to $P(Y|N)$ produced only 13 significant predictions for $P(Y|T+N)$ and 56 significant predictions for $P(Y|N)$. It is unreasonable to assume that subjects could employ a different “mismatch” for T+N and N stimuli. The pattern of fit results described above suggest that the best fit parameters differ for $P(Y|T+N)$ and $P(Y|N)$, indicating that a single mismatch channel was not likely responsible for the subjects responses. The parameters are presented in Table 3-4 for only the study-condition-subject combinations that yielded significant predictions for both $P(Y|T+N)$ and $P(Y|N)$ for both fit conditions. Under the mismatch hypothesis, one would expect that the values fit for $P(Y|T+N)$ would be the same as those for $P(Y|N)$. In general, the values differ depending on fit condition with no discernable pattern. One caveat could be that the specific resolution used to search the parameter space was not sufficiently fine. Ongoing simulations will include a more exhaustive search using finer resolutions for auditory-nerve model center frequencies, delays, and attenuations.

Table 3-4. MM model parameters fit to $P(Y|T+N)$ and $P(Y|N)$. Model abbreviations are the same as in Fig. 3-13.

Study	Condition	Subject	Model	Fit to $P(Y T+N)$				Fit to $P(Y N)$			
				CF(R)	CF(L)	Delay	Atten	CF(R)	CF(L)	Delay	Atten
1	115 Hz	3	MMn	700	400	400	0	650	650	-100	-5
2	100 Hz	1	MMc	550	700	-300	-5	500	550	-300	10
		1	MMn	600	700	200	-10	500	550	-300	-10
3	2900 Hz	2	MMn	300	700	-400	5	300	700	-400	0
		3	Mme	350	300	-400	-10	300	350	400	5
		4	MMc	300	400	-400	-10	300	350	-400	-10
		4	MMn	300	350	400	-5	300	350	400	0
		3	MMn	500	450	-100	5	500	500	-200	10
		2	MMc	550	600	400	5	500	500	-400	-10
	LNRE	5	MMn	350	550	0	0	400	550	-300	-5
		3	Mme	400	300	300	0	300	350	400	5
	RNEE	1	Mme	400	550	400	-10	350	700	-300	0
		2	MMc	450	600	100	10	450	700	400	-5
	LNEE	2	MMn	650	700	100	0	600	600	-400	-5
		4	Mme	550	300	400	0	300	550	200	-5
		1	MMc	400	400	-100	10	600	600	0	0
		1	MMn	450	400	400	-10	350	400	200	10
4	E_1C_1	4	MMc	600	650	100	-10	350	700	-200	-10
		2	MMc	350	650	-400	0	500	550	-300	5
	E_2C_2	2	MMn	300	550	300	0	400	600	200	-5
		2	MMc	600	500	0	0	500	600	0	-5
	E_1C_2	2	MMn	600	500	0	-10	500	600	0	-5
		3	MMc	300	550	-400	0	300	450	100	5
		5	MMe	500	500	100	5	500	500	-200	10
	E_2C_1	2	MMc	450	300	100	-10	400	700	-400	10
		2	MMn	500	300	300	10	400	700	-400	10
		2	MMn	350	550	100	5	350	450	-400	-10
		5	MMe	400	400	-300	0	450	700	400	10
		6	MMn	300	400	-400	-10	450	600	400	-5

3.3.2.2 Comparisons between models

Between-model comparisons are shown in Tables 3-5 and 3-6 for $P(Y|T+N)$ and $P(Y|N)$ for the EN, Wav, Xav, FCn, BR, MMnd [MMn fit to $P(Y|T+N)$], and the MMnf [MMn fit to $P(Y|N)$] models. The signal-to-noise ratios of the listeners with the lowest thresholds (in each study) were used for comparisons in Table 3-5, and the signal-to-noise ratios of the listeners with the highest thresholds were used for comparisons in Table 3-6. Table 3-5 is discussed first. The following model decision variables were correlated: Wav and Xav, Wav and BR, and Wav and MMnd. Stimulus energy (EN) was modestly, albeit significantly correlated to Wav, Xav, and the BR models under some conditions. The results of Table 3-6 are similar, but as expected for the subjects with higher thresholds, the correlations of many of the decision variables to energy were stronger at the higher signal-to-noise ratio. Note also that Wav and Xav were slightly less correlated at the higher signal-to-noise ratios.

Colburn *et al.* (1997) and Isabelle (1995) discounted models based on cross-correlation, normalized cross-correlation, and equalization-cancellation based upon their relationships to stimulus energy. Tables 3-5 and 3-6 showed only moderate correlations between the MM, BR, and FC models to energy. (Note that the EE conditions of study 3 had equal energies, and all conditions of study 4 had nearly equal energies, correlations of the MM, BR, and FC models were expected to be near zero.) Because energy was not correlated to the subjects' detection patterns, the failure of these models is partially explained by their moderate correlations to stimulus energy.

3.3.2.3 Between-condition comparisons of model decision variables for study 4

Tables 3-7 and 3-8 show between-condition comparisons of the 7 representative model decision variables applied to detection patterns of study 4. The results in Table 3-7 were computed using the lowest signal-to-noise ratio in study 4, and the results in Table 3-8 were computed using the highest signal-to-noise ratios in the study 4. Models relying on envelope will show high correlations between "E" conditions sharing the same subscript. Models relying on fine-structure will show high correlations between "C" conditions sharing the same subscripts. At lower signal-to-noise ratios (Table 3-7) it is apparent that overall energy is highly correlated to the envelope. Such a high correlation occurs because no peripheral filtering was used and the overall energies of each stimulus waveform in each condition were identical before the tone waveform was added. Wav and Xav relate primarily to the fine structure of the stimulus waveforms but less so at higher signal-to-noise ratios (see Table 3-8). The Breebaart model was correlated more strongly to fine structure than to envelope, as temporal smoothing occurs after the binaural interaction. The four-channel and mismatch models did not produce strong patterns of envelope or carrier dominance.

P(Y T+N)											
Model Comparison	Study 1	Study 2		Study 3				Study 4			
	115 Hz	100 Hz	2900 Hz	RNRE	LNRE	RNEE	LNEE	E ₁ C ₁	E ₂ C ₂	E ₂ C ₁	E ₁ C ₂
EN - Wav	0.30*	0.43*	0.30*	0.30*	0.79*	0.04	0.00	0.05	0.15	0.08	0.01
EN - Xav	0.30*	0.43*	0.32*	0.20*	0.57*	0.04	0.00	0.01	0.12	0.34*	0.01
EN - FCn	0.23*	0.12	0.08	0.29*	0.00	0.08	0.01	0.00	0.04	0.04	0.00
EN - BR	0.06	0.20*	0.10	0.24*	0.63*	0.05	0.01	0.04	0.01	0.01	0.01
EN - MMnd	0.02	0.14	0.10	0.03	0.71*	0.03	0.01	0.09	0.02	0.02	0.00
EN - MMnf	0.24*	0.13	0.03	0.00	0.11	0.03	0.10	0.24*	0.00	0.00	0.02
Wav - Xav	1.00*	1.00*	0.78*	0.91*	0.92*	1.00*	0.94*	0.93*	1.00*	0.83*	1.00*
Wav - FCn	0.07	0.01	0.03	0.14	0.03	0.18*	0.09	0.01	0.14	0.03	0.15
Wav - BR	0.18*	0.42*	0.29*	0.52*	0.88*	0.03	0.64*	0.14	0.52*	0.56*	0.39*
Wav - MMnd	0.18*	0.28*	0.15	0.00	0.60*	0.43*	0.01	0.03	0.03	0.43*	0.13
Wav - MMnf	0.05	0.10	0.03	0.01	0.11	0.09	0.03	0.03	0.00	0.52*	0.09
Xav - FCn	0.08	0.01	0.02	0.10	0.03	0.18*	0.10	0.03	0.13	0.01	0.15
Xav - BR	0.16*	0.42*	0.18*	0.48*	0.88*	0.03	0.54*	0.11	0.52*	0.35*	0.39*
Xav - MMnd	0.18*	0.28*	0.14	0.00	0.40*	0.43*	0.00	0.02	0.03	0.30*	0.13
Xav - MMnf	0.05	0.10	0.09	0.01	0.10	0.09	0.05	0.02	0.00	0.32*	0.09
FCn - BR	0.05	0.00	0.01	0.01	0.02	0.00	0.02	0.05	0.07	0.11	0.00
FCn - MMnd	0.00	0.00	0.01	0.03	0.05	0.07	0.12	0.01	0.09	0.02	0.27*
FCn - MMnf	0.13	0.03	0.05	0.02	0.28*	0.07	0.08	0.01	0.17*	0.08	0.06
BR - MMnd	0.06	0.21*	0.34*	0.08	0.43*	0.01	0.00	0.01	0.03	0.24*	0.00
BR - MMnf	0.01	0.03	0.00	0.05	0.08	0.14	0.01	0.00	0.10	0.70*	0.01
MMnd- MMnf	0.31*	0.46*	0.00	0.88*	0.20*	0.11	0.06	0.38*	0.47*	0.38*	0.13

P(Y N)											
FCn - MMnd	0.01	0.24*	0.01	0.00	0.69*	0.09	0.32*	0.00	0.01	0.01	0.01
FCn - MMnf	0.47*	0.01	0.33*	0.10	0.01	0.02	0.04	0.11	0.01	0.02	0.00
MMnd- MMnf	0.23*	0.40*	0.02	0.81*	0.00	0.13	0.61*	0.50*	0.77*	0.10	0.07

* $p < 0.05$

Table 3-5. Correlations between model decision variables for T+N trials (top) and N trials (bottom) for N₀S_π data. Correlations are given in units of r^2 for the lowest listener signal-to-noise ratio of each study. Model abbreviations are the same as listed in Fig. 3-13. MMnd indicates the model was fit to P(Y|T+N) MMnf indicates the model was fit to P(Y|N). Significant correlation ($p < 0.05$) are marked with *.

P(Y T+N)											
Model Comparison	Study 1	Study 2		Study 3				Study 4			
	115 Hz	100 Hz	2900 Hz	RNRE	LNRE	RNEE	LNEE	E ₁ C ₁	E ₂ C ₂	E ₂ C ₁	E ₁ C ₂
EN - Wav	0.48*	0.35*	0.45*	0.48*	0.38*	0.06	0.02	0.03	0.22*	0.12	0.10
EN - Xav	0.25*	0.43*	0.39*	0.29*	0.21*	0.06	0.03	0.00	0.56*	0.21*	0.06
EN - FCn	0.31*	0.01	0.12	0.00	0.06	0.01	0.16*	0.00	0.03	0.02	0.00
EN - BR	0.19*	0.34*	0.10	0.30*	0.39*	0.21*	0.01	0.03	0.04	0.01	0.01
EN - MMnd	0.13	0.24*	0.06	0.07	0.19*	0.05	0.02	0.16*	0.02	0.04	0.03
EN - MMnf	0.08	0.21*	0.00	0.18*	0.05	0.16*	0.01	0.00	0.01	0.01	0.23*
Wav - Xav	0.87*	0.75*	0.84*	0.83*	0.25*	1.00*	0.31*	0.95*	0.00	0.67*	0.63*
Wav - FCn	0.59*	0.05	0.06	0.13	0.01	0.22*	0.15	0.09	0.02	0.34*	0.08
Wav - BR	0.37*	0.46*	0.37*	0.45*	0.02	0.04	0.01	0.17*	0.34*	0.17*	0.14
Wav - MMnd	0.31*	0.06	0.07	0.13	0.18*	0.40*	0.00	0.02	0.02	0.21*	0.18*
Wav - MMnf	0.01	0.15	0.01	0.16*	0.03	0.01	0.01	0.01	0.24*	0.03	0.16*
Xav - FCn	0.39*	0.04	0.09	0.18*	0.00	0.22*	0.05	0.10	0.09	0.14	0.00
Xav - BR	0.25*	0.49*	0.24*	0.55*	0.45*	0.04	0.16*	0.17*	0.02	0.44*	0.46*
Xav - MMnd	0.29*	0.19*	0.09	0.06	0.23*	0.40*	0.06	0.00	0.00	0.21*	0.10
Xav - MMnf	0.00	0.15	0.03	0.23*	0.01	0.01	0.03	0.01	0.00	0.00	0.09
FCn - BR	0.47*	0.07	0.10	0.15	0.01	0.12	0.00	0.11	0.02	0.19*	0.04
FCn - MMnd	0.12	0.01	0.03	0.02	0.10	0.09	0.00	0.04	0.10	0.03	0.11
FCn - MMnf	0.04	0.03	0.01	0.01	0.37*	0.11	0.00	0.00	0.19*	0.04	0.22*
BR - MMnd	0.01	0.06	0.02	0.07	0.25*	0.10	0.00	0.03	0.07	0.11	0.01
BR - MMnf	0.15	0.03	0.00	0.31*	0.04	0.01	0.03	0.08	0.10	0.00	0.00
MMnd- MMnf	0.06	0.18*	0.35*	0.00	0.20*	0.03	0.78*	0.23*	0.16*	0.24*	0.69*

P(Y N)											
Model Comparison	Study 1	Study 2	Study 3	Study 4	Study 5	Study 6	Study 7	Study 8	Study 9	Study 10	Study 11
FCn - MMnd	0.17*	0.02	0.38*	0.50*	0.26*	0.09	0.00	0.00	0.00	0.01	0.10
FCn - MMnf	0.47*	0.57*	1.00*	0.77*	0.55*	0.00	0.01	0.19*	0.01	0.01	0.03
MMnd- MMnf	0.07	0.00	0.39*	0.19*	0.01	0.14	0.77*	0.39*	0.61*	0.10	0.01

* $p < 0.05$

Table 3-6. Correlations between model decision variables for T+N trials (top) and N trials (bottom) for N₀S_n data. Correlations are given in units of r^2 for the highest listener signal-to-noise ratio of each study. Model abbreviations are the same as listed in Fig. 3-13. MMnd indicates the model was fit to P(Y|T+N) MMnf indicates the model was fit to P(Y|N). Significant correlation ($p < 0.05$) are marked with *.

3.4 Summary and conclusions

Results for diotic modeling exercises indicated that none of the temporal models were able to explain detection patterns as well as the linear combination of energy at the output of several critical bands. The relatively complicated Dau *et al.* (1996a) and Breebaart *et al.* (2001) models performed more poorly than the standard critical-band model at predicting the subjects' detection patterns, and the variance explained by these models was attributable to their correlation to the critical band model. The best performing dichotic model tested here was the average value of a linear combination of ITDs and ILDs (Isabelle, 1995), predicting up to about 70 percent of the variance in some subjects' detection patterns. The following chapter highlights suggestions for future work, including alternate implementations of the Breebaart model with an emphasis on the specific use of the model's template mechanism. The analysis described in Ch. 2 is repeated in Ch. 4, except subjects' detection patterns were replaced with decision variables produced by the models described in this chapter.

Table 3-7. Correlations between dichotic model decision variables (in terms of r^2) for the various conditions of study 4, separated by T+N and N stimuli, and model. These predictions were calculated at the signal-to-noise ratio of the subject with the lowest threshold. Model abbreviations are as in Fig. 3-1. Significant correlations are shown with *.

Model		P(Y T+N)			P(Y N)		
		E ₂ C ₂	E ₁ C ₂	E ₂ C ₁	E ₂ C ₂	E ₁ C ₂	E ₂ C ₁
EN	E ₁ C ₁	0.27*	1.00	0.27*			
	E ₂ C ₂		0.27*	1.00			
	E ₁ C ₂			0.27			
Wav	E ₁ C ₁	0.04	0.07	0.94*			
	E ₂ C ₂		0.98*	0.04			
	E ₁ C ₂			0.05			
Xav	E ₁ C ₁	0.03	0.02	0.98*			
	E ₂ C ₂		0.83*	0.04			
	E ₁ C ₂			0.02			
FCn	E ₁ C ₁	0.00	0.33*	0.36*	0.05	0.90*	0.06
	E ₂ C ₂		0.14	0.37*		0.05	0.92*
	E ₁ C ₂			0.15			0.03
BR	E ₁ C ₁	0.43*	0.43*	0.60*			
	E ₂ C ₂		0.76*	0.27*			
	E ₁ C ₂			0.26			
MMnd	E ₁ C ₁	0.00	0.02	0.21*	0.07	0.05	0.08
	E ₂ C ₂		0.37*	0.07		0.05	0.00
	E ₁ C ₂			0.15			0.05
MMnf	E ₁ C ₁	0.12	0.00	0.02	0.21*	0.11	0.00
	E ₂ C ₂		0.01	0.01		0.01	0.00
	E ₁ C ₂			0.08			0.00

* $p < 0.05$

Table 3-8. Correlations between dichotic model decision variables (in terms of r^2) for the various conditions of study 4, separated by T+N and N stimuli, and model. These predictions were calculated at the signal-to-noise ratio of the subject with the highest threshold. Model abbreviations are as in Figs. 3-3 and 3-13. Significant correlations are shown with *.

Model		P(Y T+N)			P(Y N)		
		E_2C_2	E_1C_2	E_2C_1	E_2C_2	E_1C_2	E_2C_1
EN	E_1C_1	0.27*	1.00*	0.27*			
	E_2C_2		0.27*	1.00*			
	E_1C_2			0.27*			
Wav	E_1C_1	0.05	0.01	0.63*			
	E_2C_2		0.50*	0.01			
	E_1C_2			0.09			
Xav	E_1C_1	0.06	0.01	0.94*			
	E_2C_2		0.02	0.02			
	E_1C_2			0.00			
FCn	E_1C_1	0.10	0.35*	0.03	0.05	0.90*	0.06
	E_2C_2		0.29*	0.14		0.05	0.92*
	E_1C_2			0.02			0.03
BR	E_1C_1	0.23*	0.17*	0.61*			
	E_2C_2		0.89*	0.14			
	E_1C_2			0.12			
MMnd	E_1C_1	0.00	0.01	0.21*	0.06	0.00	0.08
	E_2C_2		0.01	0.00		0.01	0.09
	E_1C_2			0.20*			0.14
MMnf	E_1C_1	0.01	0.00	0.46*	0.14	0.30*	0.31*
	E_2C_2		0.00	0.03		0.16*	0.27*
	E_1C_2			0.00			0.17*

* $p < 0.05$

CHAPTER 4

General discussion and summary

This chapter summarizes the general findings of the experiments in the Appendix and Ch. 2 and the modeling efforts described in Ch.3. The analysis of the relative contributions of envelope and fine-structure cues from Ch. 2 is repeated using the model decision variables described in Ch.3 (rather than subjects' detection patterns). Specific psychophysical experiments and modeling exercises are then proposed, with special attention paid to the utility of template mechanisms for detection of tones in noise.

4.1 Summary

The goal of this dissertation was to characterize the stimulus features upon which detection of tones in noise is based. A main result of the experiment described in the Appendix was in agreement with previous research, (e.g., Fletcher, 1940; Richards and Nekrich, 1993) indicating that if overall energy differences are present between noise-alone and tone-plus noise stimuli in a diotic detection task, subjects tend to use those energy differences to perform the task. This finding was verified using a basic energy model that was able to predict up to 85 percent of the variance in subjects' detection patterns. Seemingly in conflict with this result, sets of stimuli with corresponding waveforms having different overall energies but identical temporal structures (and therefore relative magnitudes and phase spectra) produced significantly correlated detection patterns. This fact indicates that although much of the variance in listeners' detection patterns can be explained by energy, the spectral or temporal structure of the stimulus waveform did indeed factor into subjects' decisions. Implications of this result for physiological experiments, along with suggestions for future psychophysical experiments, are discussed below.

The results of the experiment described in Ch. 2 suggest that listeners' decision variables must be based on both envelope and fine-structure in order to predict detection patterns estimated in the N_0S_0 or N_0S_π stimulus configurations. Generally speaking, significantly more variance in subjects' detection patterns was predicted when responses based on both stimulus envelopes and carriers were included as predictors. Further, these stimulus features interacted in a way such that a linear combination of decision variables separately derived from envelope and fine structure did not predict all of the predictable variance in listeners' detection patterns.

The results of Ch.3 showed that several diotic models that successfully predict thresholds for tone-in-noise detection tasks cannot explain diotic detection patterns. Further, none of the temporal models examined in this work were able to predict significant proportions of variances in all of the subjects' data, including cases in which energy cues were made unreliable (and thus listeners were forced to rely on cues other than overall energy). A model based on a linear combination of energies at the output of several filters surrounding the target frequency (MD; Gilkey *et al.*, 1986) best predicted the data with level variations from noise to noise. A model based on envelope fluctuation (ES; Richards, 1992; Zhang, 2004) best predicted detection patterns estimated from stimuli with no level variations; the model predicted more variance in listeners' detection patterns than either the Dau *et al.* (1996a) or Breebaart *et al.* (2001a) models.

Implementations of several models of binaural unmasking were also tested in Ch. 3. The best performing models used a linear combination of ITD and ILD information (Xav and Wav; Goupell, 2005; Isabelle, 1995). The binaural version of the Breebaart *et al.* (2001a) model performed on a level comparable to that of Xav and Wav. As in the diotic condition, none of the models tested were able to make significant predictions for every subject in every condition of the studies tested. The combined results of Ch. 2 and 4 lead to interesting observations about

both diotic and dichotic detection models that rely on templates to compute decision variables. These observations are discussed the context of future modeling exercises below.

4.2 An analysis of model decision variables using the methods of Ch. 2

Some of the models explored in Ch. 3 used interactions of cues derived from envelope and fine structure to compute decision variables. It was of interest to determine if the nature of these interactions was appropriate by comparing them to the interactions of listeners presented in Ch. 2. Table 4-1 shows the results of applying the analysis described in Ch. 2 to the model decision variables described in Ch.3 (in lieu of the subjects' detection patterns). Briefly summarizing, stimuli from 4 conditions (E_1C_1 , E_2C_2 , E_1C_2 , and E_2C_1) shared either envelopes (E) or carriers (C); subscripts shared between conditions indicate the particular waveform component shared between conditions). Model detection patterns were computed for each of the 4 conditions. A linear regression was performed to use model detection patterns from the E_1C_2 and E_2C_1 conditions to predict either E_1C_1 (top row for each model, note that variability linearly associated with E_2C_2 was removed from the analysis for this condition) or E_2C_2 (bottom row for each model, note that variability linearly associated with E_1C_1 was removed from the analysis for this condition) detection patterns.

Results are presented for simulations run at the median signal-to-noise ratio of the subjects in Ch. 2 for the N_0S_0 condition (Table 4-1, N_0S_0). Model abbreviations are as in Fig. 3-3. All the N_0S_0 models tested relied more heavily on envelope than carrier ($R^2_E > R^2_C$) with the exception of the PO model, which made use of both envelope and fine structure. In general, the patterns of model interactions between envelope and fine structure (i.e., R^2 values) were in stark contrast to the results of the analysis presented in Figs. 2-2 and 2-5, indicating that subjects relied roughly equally on envelope and carrier. The only notable exception was for the PO model, which predicted a more equal utilization of envelope and fine-structure cues, but captured at most only about 60 percent of the variance in subjects' detection patterns.

Results are also presented for simulations run at the highest and lowest signal-to-noise ratios of the subjects in Ch. 2 for the N_0S_π condition (Table 4-1, N_0S_π : High SNR and N_0S_π Low SNR, respectively). Model abbreviations are the same as in Fig. 3-13. [Note that some models did not produce decision variables for the noise-alone conditions and thus analyses for those models were omitted from the table.] Each of the models tested here was dominated by the carrier of the waveform (with the exception of stimulus energy), as would be expected for conventional models of binaural detection at low stimulus frequencies. The pattern of interactions differs from the results of Ch 3, which indicated a more equal reliance on envelope and carrier.

Some models considered in Ch.3 actually predicted *too much* (up to 100 percent) of the variance in the model E_1C_1 and E_2C_2 detection patterns when fit with a linear model (e.g., all but the PO model for N_0S_0 conditions, and the Wav and Xav models at low signal-to-noise ratios for N_0S_π conditions). The results of Ch. 2 indicate that a linear combination of cues derived from envelope and carrier should not account for all of the predictable variance in the E_1C_1 and E_2C_2 detection patterns. These results support the conclusion that models explaining detection in reproducible noises must rely on a combination of temporal envelope and fine structure cues, and that the reliance on envelope and fine structure is likely a necessary, but not sufficient condition for predicting the listeners' detection patterns.

N_0S_0

Model	P(Y W)					P(Y T+N)					P(Y N)				
	b_E	R^2_E	b_C	R^2_C	R^2_{EC}	b_E	R^2_E	b_C	R^2_C	R^2_{EC}	b_E	R^2_E	b_C	R^2_C	R^2_{EC}
ES	1.00*	0.94*	0.45*	0.16*	0.95*	1.04*	0.96*	0.95*	0.19*	0.98*	0.89*	0.93*	0.53*	0.36*	0.96*
DA	1.06*	0.94*	0.49*	0.18*	0.96*	1.03*	0.98*	0.54*	0.14	0.99*	1.08*	0.90*	0.70*	0.25*	0.94*
	1.27*	0.87*	0.87*	0.23*	0.95*	1.33*	0.92*	0.86*	0.25*	0.95*	1.30*	0.81*	0.89*	0.27*	0.95*
BR	1.78*	0.92*	0.49	0.01	0.92*	1.90*	0.96*	-0.31	0.03	0.96*	1.66*	0.87*	1.08*	0.00	0.90*
	1.80*	0.92*	0.50*	0.03	0.93*	1.78*	0.92*	-0.47	0.04	0.93*	1.78*	0.91*	0.85*	0.07	0.94*
PO	1.04*	0.61*	1.25*	0.23*	0.79*	1.19*	0.70*	1.33*	0.05	0.77*	0.81*	0.28*	1.22*	0.40*	0.70*
	0.96*	0.84*	0.48*	0.41*	0.92*	1.14*	0.94*	0.32*	0.36*	0.96*	0.80*	0.59*	0.55*	0.52*	0.85*

 N_0S_π : High SNR

Model	P(Y W)					P(Y T+N)					P(Y N)				
	b_E	R^2_E	b_C	R^2_C	R^2_{EC}	b_E	R^2_E	b_C	R^2_C	R^2_{EC}	b_E	R^2_E	b_C	R^2_C	R^2_{EC}
EN						1.00*	1.00*	0.15	0.03	1.00*					
sT						1.00*	1.00*	0.00	0.10	1.00*					
						0.33	0.00	0.99*	1.00*	1.00*					
sl						0.32	0.00	1.00*	1.00*	1.00*					
						1.03*	0.99*	0.06	0.10	0.99*					
Wav						1.01*	1.00*	0.03	0.20*	1.00*					
						0.34*	0.02	1.00*	1.00*	1.00*					
Xav						0.58*	0.05	1.00*	1.00*	1.00*					
						0.85*	0.34*	0.86*	0.01	0.91*					
FCn	0.28*	0.27*	0.14	0.01	0.29*	0.27*	0.27*	0.09	0.00	0.28*	0.82*	0.89*	0.71*	0.02	0.98*
	0.35*	0.09*	0.36*	0.25*	0.35*	0.35	0.11	0.30*	0.21*	0.31*	1.03*	0.92*	1.00*	0.00	0.99*
BR						-0.12	0.01	0.62*	0.55*	0.55*					
						-0.03	0.00	0.69*	0.87*	0.87*					
MMnd	0.19	0.11*	0.78*	0.64*	0.66*	0.09	0.05	0.87*	0.65*	0.65*	0.65*	0.51*	0.49*	0.30*	0.62*
	0.24	0.15*	0.78*	0.63*	0.65*	-0.07	0.00	0.69*	0.51*	0.52*	0.69*	0.68*	0.42*	0.43*	0.76*
MMnf	0.43*	0.00	0.81*	0.73*	0.78*	0.25	0.13	0.78*	0.74*	0.75*	0.58*	0.45*	0.74*	0.48*	0.79*
	0.34*	0.03	0.97*	0.81*	0.84*	0.28	0.07	1.06*	0.86*	0.88*	0.75*	0.58*	0.51*	0.33*	0.80*

 N_0S_π : Low SNR

Model	P(Y W)					P(Y T+N)					P(Y N)				
	b_E	R^2_E	b_C	R^2_C	R^2_{EC}	b_E	R^2_E	b_C	R^2_C	R^2_{EC}	b_E	R^2_E	b_C	R^2_C	R^2_{EC}
EN						1.00*	1.00*	0.74	0.00	1.00*					
sT						1.00*	1.00*	0.02	0.01	1.00*					
						-0.06	0.03	0.95*	0.97*	0.97*					
sl						-0.25	0.17*	0.87*	0.94*	0.94*					
						0.51*	0.39*	-0.56	0.01	0.41*					
Wav						1.01*	0.95*	-0.08	0.29*	0.96*					
						0.18	0.08	0.99*	0.99*	0.99*					
Xav						0.55	0.06	0.91*	0.99*	0.99*					
						0.32	0.01	0.97*	0.99*	0.99*					
FCn	0.31*	0.35*	1.07*	0.49*	0.70*	0.43	0.00	0.91*	0.99*	0.99*					
	2.12*	0.51*	0.49*	0.18*	0.65*	0.31*	0.38*	1.17*	0.55*	0.78*	-8.16*	0.89*	-2.12*	0.02	0.98*
BR						2.33*	0.56*	0.53*	0.21*	0.73*	-1.84*	0.92*	-5.94*	0.00	0.99*
						0.10	0.06	0.46*	0.45*	0.47*					
MMnd	0.06	0.03	0.59*	0.45*	0.46*	0.03	0.00	0.47*	0.60*	0.60*					
	0.54*	0.29*	0.63*	0.58*	0.63*	-0.00	0.05	0.98*	0.76*	0.76*	0.01	0.00	0.12	0.08	0.08
MMnf	0.08	0.27*	0.17*	0.48*	0.49*	1.32*	0.59*	0.42*	0.53*	0.69*	0.16	0.08	0.54*	0.54*	0.55*
	-0.08	0.00	0.31*	0.26*	0.29*	0.08	0.31*	0.15*	0.51*	0.53*	0.26	0.10	-0.07	0.02	0.10

* $p < 0.05$

Table 4-1. The contribution of envelope and carrier to model decision variables computed using the analysis technique described in Ch. 2. Results are shown for the N_0S_0 condition at the subjects' median signal-to-noise ratio and for high and low signal-to-noise ratios for the N_0S_π condition. The subscripts E and C denote statistics computed for envelope and carrier respectively. Significant b values indicated that the addition of

envelope or carrier significantly increased the amount of variance explained. R^2 values are given in bold for the proportion of variance explained by envelope, carrier, and a combination of envelope and carrier. The first and second rows for each model show predictions for the E1C1 and E2C2 condition respectively (see Ch. 2 methods for details). Model abbreviations are the same as described in Figs. 3-3 and 3-13.

4.3 Implications and suggestions for future experiments

4.3.1 Psychophysical experiments

The results described in the Appendix and of Richards and Nekrich (1993) both indicate that listeners are free to use detection strategies (for diotic detection tasks) based on overall level, level-invariant information, or both, depending on the specific cues provided in the stimuli. It would be of interest to quantify the contributions of such cues in a way that (1) does not require assumptions about peripheral processing or particular decision variables, (2) allows the contribution of energy and temporal information to be specified beyond comparing thresholds computed with random maskers and (3) would allow the effect of stimulus energy on detection patterns to be quantified directly. An experiment is suggested below that is designed to meet the 3 requirements listed above.

The proposed experiment contains 3 conditions, each one with a set of reproducible maskers. All are diotic and use sub-critical noise bandwidths. Listeners would be trained using the methods described in the Appendix and in Ch. 2. In the first condition, stimuli would have overall levels that vary in the standard way; each noise-alone waveform is normalized by the RMS value of the ensemble of waveforms. When the tones are added at each listener's threshold, energy differences between the tone-plus-noise and noise-alone stimuli would be created as a function of the way the tone and noise sum temporally. In the second condition, the level variation of the noise-alone waveforms would be inverted. That is, the formerly low-level noise-alone waveforms would be scaled to become the high-level waveforms, and vice versa, but no effort would be taken to change the level resulting when the tone is added. Finally, in the third condition, the corresponding energies for tone-plus-noise waveforms would be interchanged with energies for corresponding noise-alone waveforms. In this manner, 3 detection patterns would be estimated, and by comparing across various sub-groups, different energy-related hypotheses could be tested. Order becomes paramount in these analyses, so the Spearman rank-order correlation coefficient should be used for all comparisons. For example, suppose the energy cue was dominant over all other cues. In that case, one would expect that within hit and false-alarm rates, the probabilities associated with each waveform to reverse order in the detection pattern for the second set of stimuli with respect to the detection pattern from the first set of stimuli. (e.g., the stimuli leading to high hit and false-alarm rates should now lead to low hit and false-alarm rates and vice versa, following the trend of the overall energies of each waveform). This reversal would cause the rank-order correlation between the two detection patterns to be -1. Comparing between the first and third sets, detection patterns corresponding hit and false-alarm rates should interchange on a waveform-by-waveform basis, leading to a negative but equal correlation between hit rates and false-alarm rates in the third condition with respect to hit rates and false-alarm rates from the first condition.

Such an experiment would produce a measure of the strength of the use of an energy cue, indexed by the ability to reverse the order of the probabilities in the detection patterns. Specific waveforms could be examined in the context of their probability ranks to determine any features related to the rank change (or lack of change) from detection patterns estimated in one condition or another.

The resulting correlations between detection patterns could be used as weights in a model that combines energetic and level-invariant information. Because an energy model could be applied to the data collected with the same stimuli but at two different energies (i.e., rather than equalizing overall energy), a direct test is possible of whether the successes of energy-model predictions are the result of their correlation to a yet unidentified cue, or whether energy itself is used as the decision variable.

4.3.2 Modeling exercises

4.3.2.1 Diotic models

Two modeling exercises are proposed for diotic stimuli based upon the results of Ch. 2. of this work, and the models used by Berg *et al.* (1992) and Green *et al.* (1992). Green and colleagues have studied the discrimination of narrowband spectra (including sub-critical band spectra) at a center frequency of 1 kHz. They were able to partially explain their results based on pitch cues (for bandwidths of about 1 critical band and narrower than 2 critical bands) and the fluctuation of the envelope spectrum for stimuli substantially narrower than a critical band (20 Hz at a 1 kHz center frequency). The stimuli featured in Ch.3 of this thesis could fall into either category. Thus it would be of interest to see predictions of both a pitch model, such as the envelope-weighted instantaneous frequency (described in detail in Berg *et al.*, 1992), and a model based on the modulation spectrum at the output of a single critical-band filter (described in detail in Green *et al.*, 1992; with more elaborate versions described in Dau *et al.*, 1997a; b).

4.3.2.2 Template-based models

The following discussion is intended to apply to any diotic or dichotic model that uses internal templates (e.g., Dau *et al.*, 1996a, b; Breebaart *et al.*, 2001a, b, c). For example, the Breebaart *et al.* (2001 a,b,c) model was tested in Ch.3 and found to be correlated to some listeners' data. However, the specific implementation used in Ch.3 did not take full advantage of the model's complexity (as parsimony was paramount in order to compare between models). Trial-by-trial responses were not simulated and a running template was not computed. Computation of a running template could be a useful modeling exercise that would not only exploit the full potential of the model, but would also investigate an often-heard criticism regarding experiments using reproducible noises. Critics point out that experimenters have no real way of knowing that listeners maintain the same detection strategy throughout an entire experiment. Admittedly, the author and other listeners have reported being influenced by particular noise waveforms, or even feeling temporarily confused for brief periods (i.e., tens of trials) during an experiment. Individual responses and waveform identification numbers were recorded on each trial for the experiments presented here, providing data suited for an interesting analysis of template-based models. Suppose that a template was constructed as the mean of several practice trials of randomly generated noise. Suppose also that this memory was a buffer of a limited number of waveforms in a first-in first-out configuration. It would be of interest to re-examine model predictions for the data in studies 1-4 as a function of the buffer length (or the number of internal representations of the stimuli used to compute an average template). This analysis is possible because responses to each waveform can be used to sort waveforms into *perceived* tone-plus-noise and noise-alone groups, regardless of the stimuli used for each trial.

One potential drawback of this modeling approach (and for that matter, a drawback of any of the models used in this study) would be that the potential use of short-term cues would not be captured by the template mechanism employed in the above models. Subjects reported for the dichotic detection tasks that relatively brief epochs of stimuli were often the basis for decisions. This fact compounds the modeling problem because the temporal locations of these epochs are

unknown, and may differ from waveform to waveform. Such a strategy would render useless the temporal weighting scheme used by the Breebaart model, which averages across waveforms. Further, an epoch-based strategy would likely require a rethinking (i.e., shortening) of the time constant used for smoothing the output of the binaural processor in the Breebaart model.

Recent evidence suggests that the relatively long estimates of binaural temporal windows, 60 to 200 ms (e.g., Grantham and Wightman, 1979; Kollmeier and Gilkey, 1990; Culling and Summerfield, 1998) may in fact be too long, and estimates on the order of 50 ms or shorter might be more suitable for modeling the current data (Kolarik and Culling, 2005). Indeed, researchers testing temporal aspects of binaural processing have reported time constants as short as 10 ms (e.g., Akeroyd and Bernstein, 2001). One other possibility is that listeners may employ more than one type of (potentially short-term) template. This strategy could be investigated by grouping waveforms by their respective hit and false alarm rates, and then investigating the templates that result from training the model using waveforms corresponding to high, moderate, and low hit rates.

Another suggestion for future modeling efforts is inspired by Hancock and Delgutte (2004). Results from Ch.3 of the current study suggest that a single binaural delay/attenuation model cannot explain detection for reproducible stimuli. The Hancock and Delgutte model was originally designed to predict interaural time difference (ITD) discrimination data and is based on recordings from the inferior colliculus of cat. The model employs a neuronal pooling strategy that optimally combines d' values across a population of model neurons tuned in best frequency and ITD according to distributions measured in cat. It is possible that responses of a population of channels tuned to a number of different ITD values are necessary to account for the current data.

Diotic and dichotic detection under restricted-cue conditions

A.1 INTRODUCTION

This Appendix describes a preliminary experiment that addresses the use of energetic vs. temporal cues as decision variables for tone-in-noise detection. The results of this experiment inspired the design of the experiment in described in Ch. 2, which more specifically tested the use of temporal-envelope vs. fine-structure based cues for tone-in-noise detection.

In this experiment, detection patterns (composed of hit rates and false-alarm rates) were estimated from subjects' responses to multiple presentations of several groups of N_0S_0 and N_0S_π reproducible noise waveforms. The overall energy variation across masking waveforms was known, and both random-phase noise and low-noise noise (noise with reduced envelope fluctuations; Pumplin, 1985) were used as masking stimuli. The variance of detection patterns was analyzed to test the hypothesis that energy was used as a detection cue for N_0S_0 stimuli. This analysis was repeated to assess the impact of altering the temporal properties of the stimulus waveforms using low-noise noise for both N_0S_0 and N_0S_π stimuli. Results for N_0S_0 stimuli suggest a detection strategy based on overall energy that was slightly influenced by temporal-envelope fluctuations in the signal. The exact role of envelope fluctuations in the detection task remains unclear for N_0S_0 stimuli. Results did not indicate a specific waveform feature (i.e. temporal envelope or energy) that was able to explain detection under all conditions, although a simple energy model was able to explain detection patterns estimated using N_0S_0 stimuli, predicting up to 84 percent of the variability in subjects' responses. Results for N_0S_π stimuli indicate that both fine-structure and temporal envelope are used in detection, suggesting that a more direct investigation of the roles of overall energy and waveform envelope and fine structure should be pursued.

A.1.1 Background

Despite many years of study, the exact process by which a listener detects a tone in a noise waveform remains unknown. Over the past half-century, a number of explanatory models have been used in attempts to predict the results of both diotic (e.g. Fletcher, 1940; Richards and Nekrich, 1993; and Carney *et al.*, 2002) and dichotic (e.g., Durlach, 1963; Hafter, 1971; and Colburn, 1977) tone-in-noise-detection experiments. Although these models are successful in predicting some detection thresholds, studies using reproducible-masking waveforms have shown that these models do not sufficiently capture the detection process on an individual-waveform basis (with the exception of energy-based model predictions for diotic stimuli; Gilkey and Robinson, 1986; Isabelle, 1995; Davidson *et al.*, 2006). These studies used a number of models to try to predict individual-subjects' hit and false-alarm rates for detecting tones in sets of reproducible masker waveforms.

Hit rates $[P(Y|T+N)]$ and false-alarm $[P(Y|N)]$ rates are *a posteriori* probabilities of reporting tone presence for multiple presentations of each tone-plus-noise (T+N) and each noise-alone (N) stimulus, respectively. If the $P(Y|T+N)$ and $P(Y|N)$ values are considered together as a set of waveform-dependent probabilities $[P(Y|W)]$, they form a detection pattern. Many dichotic models of tone-in-noise detection have proven unable to predict individual subjects' detection patterns (Isabelle, 1995). While some diotic energy-based models were able to partially capture the diotic detection patterns, they are known to fail in other detection tasks, such as those using a roving stimulus level, equal noise energies, low-noise noise, or Schroeder stimuli, suggesting

that temporal stimulus properties must be considered (e.g., Kidd *et al.*, 1989; Richards, 1992; Richards and Nekrich, 1993; Eddins and Barber, 1998; and Smith *et al.*, 1986).

Gilkey and Robinson (1986) and Davidson *et al.* (2006) showed that under diotic stimulus conditions, a multiple detector (MD) model composed of a linear combination of overall energy at the output of several critical bands surrounding the tone frequency was a good predictor of subjects' detection patterns. This finding suggests that overall stimulus energy may be a more salient cue for diotic tone-in-noise detection processes than temporal fluctuations in individual waveforms. Nevertheless, Richards (1992) found that listeners are able to perform the detection task even when stimulus energies are equalized across T+N and N stimuli. This study is a first attempt at determining the contributions of temporal-waveform structure to subjects' detection patterns estimated with diotic and dichotic reproducible stimuli. Here, overall stimulus energies (and magnitude spectra) were manipulated together with temporal stimulus properties in an attempt to clarify cues used for tone-in-noise detection.

This study extends previous work (e.g., Zheng *et al.*, 2002; Evilsizer *et al.*, 2002; and Davidson *et al.*, 2006) to begin a general examination of stimulus components that may be used to compute specific cues for tone-in-noise detection. One of the major goals of this work was to move away from the standard tone-in-reproducible-noise detection experiment that was designed to compare detection under different noise bandwidths, tone phases, etc., or to generate data suitable for modeling (e.g., Pfafflin and Matthews, 1966; Ahumada and Lovell, 1971; Isabelle and Colburn, 1991; Isabelle, 1995; and Evilsizer *et al.*, 2002). The design of the present experiment calls for the estimation of multiple detection patterns that can be compared (see Fig. 1-2) to test critical modeling assumptions as described below.

A.1.2 Experimental design

Here, rather than using an *a posteriori* modeling approach, models were used to guide *a priori* manipulations of the reproducible-noise waveforms. Specifically, signal-processing techniques were used to manipulate waveform energies, temporal structures, and temporal envelope properties. At this point it is worth specifically defining the usage of “temporal structure,” “temporal envelope,” and “temporal fine structure” in this document. The term “temporal structure” refers to the specific time-domain representation of the waveform as a whole, without regard for presentation level. The term “temporal envelope” refers to the slower fluctuations present in the time-domain waveform that can be removed either by half-wave rectification and low-pass filtering, or for our purposes, digitally using the Hilbert transform. “Temporal fine structure” refers to the carrier (e.g., zero crossings) present in the stimulus waveform, without regard for the envelope. “Temporal structure” includes both envelope and fine-structure information.

Four groups of stimuli (or cue conditions) were incorporated into the design of this study and were presented under both N_0S_0 and N_0S_π interaural configurations. A baseline group (random noise, random energy; RNRE) was composed of waveforms with random phase spectra and overall-energies that varied randomly across the set. A second group (low noise, random energy; LNRE) shared the same magnitude spectra as the first group, and also had varying overall energies, but had phase spectra that were selected using an adaptation of the low-noise noise (LNN) algorithm (Pumplin, 1985). (The LNN algorithm selects phases to minimize the 4th moment, or envelope fluctuation, of the noise waveform and is described in detail in Sec. A.2.1) The LNRE stimuli therefore had different temporal structures than the RNRE stimuli, and would effectively reduce any cues having a (monotonically increasing) relationship to the magnitude of temporal envelope fluctuations in the stimuli. A third group of waveforms (random noise, equal energy; RNEE) had the same temporal structures as the first group, but had no energy variations across waveforms within the T+N or N stimuli. The energies of the T+N and N waveforms from

the RNEE conditions were equalized to the average energies of the T+N and N stimuli from the RNRE condition, respectively. Finally, a fourth condition (LNEE; low noise equal energy) also had no variations in energy within T+N and N stimuli. Overall levels of the T+N and N LNEE stimuli were equalized to the average levels of T+N and N LNRE stimuli, respectively.

The influences of these manipulations were quantified by estimating subjects' detection patterns for each set of waveforms. If the stimulus manipulations altered cues used for the detection task, observable differences in subjects' detection patterns would result across the four stimulus conditions. By comparing the detection patterns across certain cue conditions, three naive hypotheses could be tested. Possible outcomes of these hypotheses are summarized in Table A-1.

The first hypothesis was that subjects rely on overall stimulus energies to perform the detection task. Although researchers have proven that listeners do not depend entirely upon overall energy to perform the detection task (e.g., Kidd *et al.* 1989; Richards, 1992; Richards, 1993), recent evidence suggests that listeners may employ this strategy when

Table A-1. Expected correlations for the 3 naive hypotheses involving the 3 comparisons (C) shown in Figs. 1 and 2. In C₁ the two sets of stimuli have different temporal structures, but correlated energies. In C₂ the two sets of stimuli have the same temporal structure but uncorrelated energies. In C₃ the two sets of stimuli have the same temporal structures, uncorrelated energies and less dynamic envelopes than in C₂.

	RNRE vs. LNRE	RNRE vs. RNEE	LNRE vs. LNEE
	C ₁	C ₂	C ₃
Energy	Moderate	Low	Low
Fine Structure	Low	High	High
Envelope	Low	High	Low

no effort is taken to make an overall-level cue unreliable (Davidson *et al.*, 2006). Under such a strategy, it was expected that the variance of detection patterns (i.e., the variance of hit and false-alarm rates across waveforms) estimated under the RNEE and LNEE conditions would be smaller than the variance of detection patterns estimated under the RNRE and LNRE conditions respectively, as there were no across-waveform differences (considering T+N and N stimuli separately) in overall energy in the equal energy conditions. Further, since the correlations of the energies of the RNRE and LNRE stimuli were known, only a moderate correlation between subjects' responses in RNRE and LNRE conditions was expected. As a corollary, it was also expected that detection patterns in the RNRE and RNEE conditions would *not* be correlated and that detection patterns in the LNRE and LNEE conditions would also *not* be correlated (because the overall stimulus energies did not vary across T+N or N waveforms under EE conditions).

The second hypothesis states that listeners use temporal cues (based on temporal fine structure or temporal-envelope fluctuations) to perform the detection task. Under this hypothesis, detection patterns estimated from the RNRE and LNRE conditions would *not* be correlated, as the stimulus temporal structure was not preserved between corresponding waveforms across the two sets of stimuli. As a corollary to this hypothesis, detection patterns estimated from RNRE and RNEE stimuli were expected to be correlated, as well as patterns estimated from the LNRE

and LNEE conditions. In each of the latter two comparisons, temporal structures were preserved between corresponding waveforms (but not overall energies).

The third hypothesis states that listeners use only temporal envelopes (e.g., a cue monotonically related to the magnitude of temporal envelope fluctuations) to perform the detection task. This hypothesis is based on the Richards (1992) observation that when a tone is added to a noise waveform, temporal fluctuations in the envelope of the noise waveform are effectively reduced. Such an observation requires that listeners code the average temporal fluctuation of the stimulus envelope. The temporal-envelope hypothesis calls for only weak or moderate correlations between detection patterns estimated from the LNRE and LNEE conditions, because such stimuli have intrinsically small envelope fluctuations with respect to those of the RNRE and RNEE conditions. Any use of a cue that is monotonically related the magnitude of envelope fluctuations should also be manifested as a significant change in the variance of detection patterns estimated under LN (low-noise) conditions with respect to RN (random-noise) conditions. The difference in detection-pattern variances between conditions is expected because the variance of the distributions of envelope fluctuation (4th moment, described below) are larger for RN stimuli than LN stimuli. In addition, if this temporal-envelope cue was employed, detection patterns estimated in the RNRE and RNEE conditions were expected to be correlated, as no attempt to reduce the magnitude of envelope fluctuations was made for these stimulus sets. Finally, under this hypothesis, no correlation should have existed between the RNRE and LNRE conditions, because the temporal structures of corresponding waveforms differed across these conditions.

A.2 METHODS

Experimental procedures were matched closely to those of Davidson *et al.* (2006), Evilsizer *et al.* (2002), and Gilkey *et al.* (1985). In this experiment, tone-in-noise detection was performed under both diotic (N_0S_0) and dichotic (N_0S_π) listening configurations using reproducible noises. Listening was completed in a single-walled sound attenuating booth (Industrial Acoustics Company, Bronx, NY) located in a small room within a larger, quiet, concrete-walled room. Five subjects ages 21-26 years completed the experiment and each had audiometrically-normal hearing. One subject (S5, the first author) had previous experience with tone-in-noise detection experiments.

A.2.1 Stimuli

Stimuli were created with MATLAB software (Mathworks, Natick, MA) and presented with a TDT System III (Tucker Davis Technologies, Gainesville, FL) RP2 D/A converter and TDH-39 headphones (Telephonics Corp., Farmington, NY). Within each interaural configuration, stimuli were generated according to one of 4 cue-manipulations. Each of the cue conditions contained 25 T+N and 25 noise-alone waveforms, such that 16 groups of 25 waveforms were considered (i.e., both T+N and N groups, in two interaural configurations, and in four cue conditions: RNRE, LNRE, RNEE, LNEE). Each waveform had a sub-critical bandwidth (according to Glasberg and Moore, 1990) of 50 Hz centered at 500 Hz and was 100 ms in duration.

Waveforms for the baseline condition (RNRE) were generated in the frequency domain using 5 frequency components with Rayleigh-distributed magnitudes and uniformly-distributed phases. Stimuli with temporal (phase) manipulations were generated using the same magnitudes as in the baseline condition, but had phases selected with an adaptation of the LNN algorithm as described in Pumplin (1985). This algorithm selected phase values that minimized the 4th moment of each waveform in the set of waveforms, effectively creating a set of stimuli that had smaller envelope fluctuations than RN stimuli. The choice of LNN phases allowed the temporal envelope, which is known to have an effect on tone-in-noise detection thresholds in both diotic

and dichotic conditions (Eddins and Barber, 1998), to be examined using individual masker waveforms. Specific details of the LNN algorithm are discussed below.

In order to minimize changes in envelope caused by adding the tone waveforms to the N waveforms, T+N waveforms were incorporated into the “error” function of the LNN algorithm (i.e., the function that is minimized). Within the LNN-algorithm “error” function, signals were added to the N waveforms at levels corresponding to the highest threshold across subjects (established during training). The highest threshold value was selected because average deviations in 4th moment increase with increasing signal-to-noise ratio (SNR). The “error” signal in the LNN algorithm was modified to simultaneously and equally weight the 4th moments of the N waveform, the (N_0S_0) T+N waveform, the π -phase (N_0S_π) T+N waveform, and the 0-phase (N_0S_0) T+N waveform. The resulting search produced a set of magnitudes, and two sets of phases (corresponding to low-noise noise and random-phase noise, such that the same sets of magnitudes and phases were used for all listeners). These magnitudes and phases were then used to generate the unscaled time-domain RN and LN N waveforms. The time-domain waveforms were then cyclically shifted such that the minimum of the envelope occurred at the onset and offset of the stimulus in order to reduce changes in stimulus properties caused by gating the stimuli. RN and LN waveforms were then normalized by the average RMS of all 25 RN and 25 LN waveforms respectively, and then multiplied by the RMS value of a 40-dB SPL spectrum level, 50-Hz bandwidth, 100-msec duration noise waveform (57 dB SPL). This normalization process ensured that random level and SNR variation existed across waveforms (denoted as RE, or random energy). Tone waveforms were then added at each individual subject’s threshold, and 10-msec \cos^2 ramps were applied to the T+N waveform, resulting in the final sets of LNRE and RNRE stimuli.

Stimuli with restricted energy cues (denoted as EE, or equal energy) were generated from the same unscaled time-domain-shifted waveforms used in the LN and RN conditions above. In order to minimize differences in energy across waveforms, each N waveform was scaled to an RMS level of exactly 57 dB SPL (40-dB SPL spectrum level), effectively eliminating variation in overall level and SNR across waveforms. Signal waveforms were added at the thresholds established during training (as above) and 10-ms \cos^2 ramps applied. Each of the RNEE and LNEE waveforms (with T+N and N groups treated separately) was then normalized to the mean level of the corresponding group of T+N or N RNRE or LNRE waveforms. In this manner, mean differences in level between T+N and N groups were preserved, while overall levels within the T+N and N groups were equalized.

The resulting RE waveform sets had standard deviations (across subjects and RN and LN noises) in level of about 2.3 dB SPL. The average difference in overall level of T+N and N waveforms for EE stimuli was about 3.2 dB SPL. The effectiveness of the LNN algorithm was tested using a Wilcoxon matched-pairs signed-ranks test (Sheskin, 2000). This non-parametric procedure tested the hypothesis that that the LN and RN 4th moment distributions came from different populations with different medians. Separate tests of T+N and N stimuli showed significant differences ($p < 0.001$) between the medians of the 4th moments of RN and LN stimuli for the various waveforms (e.g., diotic, dichotic right and left, etc.), indicating that the LNN algorithm was effective.

A.2.2 Training

Training procedures were similar to those detailed in Davidson *et al.* (2006) and will be briefly summarized here. An extensive training paradigm was used to allow subjects to form a stable decision strategy and criterion, such that stable performance occurred in the final testing procedure, which was a single-interval task using large numbers of trials at threshold. Threshold is defined here for each subject as the E_S/N_0 value in dB where $d' \approx 1$. Three separate training

tasks were completed, and each task was progressively more similar to the final testing procedure. The training procedures used 50-Hz bandwidth, 100-ms duration noise waveforms that were generated randomly on each trial (i.e., not the reproducible stimuli used in the testing procedure) and time shifted as described in Sec. A.2.1 to reduce onset transients. Randomly-generated noise was used to prevent any possible learning of the reproducible stimuli.

The following training and testing procedures were conducted under both the N_0S_0 and N_0S_π interaural configurations. In general, subjects received stimuli from only one interaural configuration per session (2-3 hours), and the use of N_0S_0 or N_0S_π stimuli alternated by session. In the rare cases where stimuli from both configurations were presented in the same session (such as to finish up a particular training or testing paradigm), the two conditions were never alternated within a session (or block). The initial listening configuration was randomized across subjects.

In the first training procedure, subjects completed 10-15 repetitions of a two-interval two-alternative forced-choice tracking procedure with trial-by-trial feedback to estimate a level for which $d'_{2AFC} = 0.77$ (Levitt, 1971). Each track was a fixed length of 100 trials. The step size was 4 dB for the first 2 reversals and 2 dB thereafter. Thresholds were estimated by averaging tone levels at reversals in the track excluding the first 4 or 5 reversals such that the number of averaged reversals was even.

The second training procedure was a single-interval, fixed-level task used to encourage stable performance at each subject's threshold. Approximately 10 blocks containing 100 trials with feedback were completed at +3, +1 and -1 dB relative to the threshold established in the two-interval task. Throughout the single-interval training procedures (and the testing procedure described in C), d' and bias (β , MacMillan and Creelman, 1991) were monitored. The d' values calculated from these blocks were used as a more accurate estimate of the tone level where d' was approximately equal to unity rounded to within 1-dB. Approximately 10 blocks were then run at that level. In the event a subject's threshold changed, the tone level was adjusted with 1-dB resolution until d' returned to near unity.

After a stable tone level was established, subjects completed approximately 10 100-trial blocks without feedback in order to determine whether d' values would remain near unity after feedback was removed. (Feedback was removed during "testing" to prevent any possible learning of the reproducible stimuli.) If necessary, tone-levels were adjusted in 1-dB steps to find the level resulting in $d' \approx 1$. The block length was then increased to 400 trials, and subjects completed 5 more blocks.

If β deviated by more than 15 percent from 1 (a value of 1 indicated an equal probability of guessing "tone" or "no tone"), subjects were given verbal feedback to "try and make an equal number of tone and no tone responses." The subjects were also notified that $\beta < 1$ indicates too many "tone" responses and $\beta > 1$ indicates too many "no tone" responses. The values of d' and β were computed using $P(Y|T+N)$ and $P(Y|N)$ across all stimulus waveforms from the four individual cue conditions. Because stimuli from the 4 cue conditions were interleaved and were all presented at the same signal-to-noise ratio, there was no control over variations in the within-cue-condition values of d' and β . Listeners were shown their respective d' and β values at the end of each block.

A.2.3 Testing

The testing procedure was identical to the final training procedure except that the reproducible noises described in Sec. A.2.1 were used as stimuli. Before each 400-trial block, 20 practice trials (that did not use reproducible stimuli) were presented with feedback. The testing paradigm called for 2 presentations of each T+N and each N stimulus from each of the 4 conditions. These presentations were randomly interleaved within each 400 trial-block.

However, a programming error resulted in unequal presentations of T+N and N stimuli in the LNEE condition only, such that the *a posteriori* probability of a tone ranged from 0.48 to 0.51 for waveforms 1-24 (the RNRE, RNEE and LNRE conditions were unaffected) and 0.33 for waveform 25. Accordingly, waveform 25 was eliminated from the analyses across all 4 conditions to maintain a balanced experiment. A total of 24 blocks were presented to each listener such that approximately 48 presentations of each T+N and each N waveform were presented at the final tone level. The narrowband noise waveforms used in training were random and did not have equal energies and were not considered low-noise noise. As a result, the tone level determined from the training procedure did not necessarily represent the level where $d' \approx 1$ for each subject when using the sets of reproducible noise waveforms. In these cases, the tone level was adjusted in 1-dB steps until $d' \approx 1$ for each subject. Each time the tone level changed, which occurred two or three times for each subject, the entire testing procedure was restarted. [Recall that thresholds estimated with LNN stimuli differ from those estimated with Gaussian noise (Hartman and Pumplin, 1988)]. The primary purpose for operating near unity d' and β was to ensure that reliable detection patterns were estimated, as operating well above threshold results in all $P(Y|T+N)$ having values near one and all $P(Y|N)$ having values near zero, and operating well below threshold results in chance performance. Any learning during this process was highly unlikely, as the long training procedure with feedback was designed to encourage subjects to establish a fixed decision strategy, and feedback was never used with reproducible noise waveforms.

A.3 Results and discussion

Several comparisons are made in the following sections: First, the reliability of the detection patterns is considered. Computing the reliability of each detection pattern establishes a reasonable upper limit for correlations amongst the various stimulus conditions in the experiment. Then, correlations between detection patterns from the various cue conditions are compared. Intersubject correlations and correlations between interaural configurations are then shown in order to reveal the potential use of similar strategies between listeners or between listening configurations. Note that for all but the between-subject analyses, an average subject was created by averaging the $P(Y|W)$ values across the four subjects.

A.3.1 Reliability of the data and detection performance

Tables A-2 through A-5 show detection-performance statistics as well as reliability statistics for all data collected in each of the 4 cue conditions under the N_0S_0 and N_0S_π interaural configurations. Overall measures of detection performance (d' values) are presented for each subject both across and within the 4 cue conditions. In general,

Table A-2. Detection performance and reliability statistics for the N_0S_0 interaural configuration. One signal-to-noise ratio (E_s/N_0) was used for each subject. Overall d' and β were computed using responses to waveforms in all conditions. Individual d' and β values are given for each of the 4 listening conditions (RNRE, random noise random energy; LNRE, low noise random energy; RNEE random noise equal energy; and LNEE, low noise equal energy). The coefficient of determination between responses from the first and the last half of the trials (r^2) and the proportion of predictable variance (V) are given for each condition, for hit and false-alarm rates considered together [P(Y|W)]. All r^2 values were significant ($p < 0.05$).

S	Overall			Condition	d'	β	P(Y W)	
	E_s/N_0	d'	β				r^2	$V_{P(Y W)}$
S1	5	1.03	1.18	RNRE	0.80	1.16	0.73	0.92
				LNRE	1.23	1.16	0.89	0.97
				RNEE	0.84	1.13	0.77	0.93
				LNEE	1.25	1.25	0.75	0.93
S2	4	1.11	1.07	RNRE	0.94	1.09	0.89	0.97
				LNRE	1.18	0.98	0.89	0.97
				RNEE	1.01	1.14	0.72	0.92
				LNEE	1.33	1.07	0.81	0.95
S3	6	0.77	0.98	RNRE	0.74	1.01	0.69	0.91
				LNRE	0.92	0.89	0.79	0.94
				RNEE	0.58	1.05	0.72	0.92
				LNEE	0.86	0.91	0.60	0.87
S4	4	0.94	0.99	RNRE	0.91	1.25	0.79	0.94
				LNRE	1.20	0.74	0.89	0.97
				RNEE	0.77	1.21	0.82	0.95
				LNEE	1.03	0.73	0.81	0.95
S5	5	1.14	0.96	RNRE	0.93	1.04	0.85	0.96
				LNRE	1.22	0.88	0.88	0.97
				RNEE	1.04	0.98	0.82	0.95
				LNEE	1.38	0.92	0.88	0.97
S_{avg}	4.8	0.99	1.03	RNRE	0.86	1.10	0.96	0.99
				LNRE	1.14	0.92	0.96	0.99
				RNEE	0.84	1.10	0.95	0.99
				LNEE	1.15	0.94	0.95	0.99

Table A-3. Reliability statistics for the N_0S_0 interaural configuration for hit and false-alarm rates are considered separately [$P(Y|T+N)$ and $P(Y|N)$]. The χ^2 statistic, the coefficient of determination between responses from the first and the last half of the trials (r^2), and the proportion of predictable variance (V) are given. All χ^2 values were significant ($p < 0.001$) and all r^2 values were significant ($p < 0.05$).

S	Condition	P(Y T+N)			P(Y N)		
		χ^2	r^2	$V_{P(Y T+N)}$	χ^2	r^2	$V_{P(Y N)}$
S1	RNRE	452	0.66	0.90	317	0.52	0.84
	LNRE	477	0.80	0.94	432	0.75	0.93
	RNEE	206	0.47	0.81	305	0.69	0.91
	LNEE	266	0.27	0.68	323	0.62	0.88
S2	RNRE	556	0.76	0.93	475	0.91	0.98
	LNRE	552	0.82	0.95	479	0.75	0.93
	RNEE	230	0.50	0.83	241	0.34	0.74
	LNEE	249	0.46	0.81	126	0.22	0.64
S3	RNRE	374	0.64	0.89	393	0.55	0.85
	LNRE	351	0.60	0.87	525	0.72	0.92
	RNEE	404	0.73	0.92	419	0.68	0.90
	LNEE	243	0.37	0.76	462	0.57	0.86
S4	RNRE	498	0.66	0.90	342	0.67	0.90
	LNRE	601	0.82	0.95	739	0.84	0.96
	RNEE	504	0.82	0.95	434	0.72	0.92
	LNEE	471	0.76	0.93	533	0.76	0.93
S5	RNRE	726	0.84	0.96	535	0.70	0.91
	LNRE	841	0.84	0.96	573	0.76	0.93
	RNEE	430	0.78	0.94	368	0.57	0.86
	LNEE	478	0.72	0.92	349	0.69	0.91
S _{avg}	RNRE	1813	0.94	0.98	1335	0.91	0.98
	LNRE	2310	0.90	0.97	2149	0.94	0.98
	RNEE	1163	0.91	0.98	1291	0.89	0.97
	LNEE	992	0.83	0.95	1256	0.90	0.97

Table A-4. Same as Table A-2, but for the N_0S_π interaural configuration.

S	Overall			Condition	d'	β	P(Y W)	
	E_S/N_0	d'	β				r^2	$V_{P(Y W)}$
S1	-4	0.87	1.13	RNRE	1.10	0.96	0.70	0.91
				LNRE	1.04	1.13	0.85	0.96
				RNEE	0.77	1.10	0.62	0.88
				LNEE	0.61	1.22	0.57	0.86
S2	-14	1.00	1.07	RNRE	1.31	0.85	0.86	0.96
				LNRE	1.10	1.11	0.78	0.94
				RNEE	1.00	0.98	0.71	0.91
				LNEE	0.67	1.22	0.66	0.90
S3	-15	1.01	1.18	RNRE	1.19	1.01	0.87	0.97
				LNRE	1.10	1.29	0.75	0.93
				RNEE	0.97	1.08	0.71	0.92
				LNEE	0.81	1.32	0.65	0.89
S4	-5	0.87	1.02	RNRE	0.93	1.22	0.88	0.97
				LNRE	1.06	0.79	0.87	0.96
				RNEE	0.86	1.22	0.87	0.97
				LNEE	0.71	0.91	0.92	0.98
S5	-16	1.00	0.97	RNRE	1.10	0.74	0.81	0.95
				LNRE	1.26	0.94	0.81	0.95
				RNEE	0.94	0.91	0.64	0.89
				LNEE	0.79	1.23	0.50	0.83
S_{avg}	-10.8	0.95	1.07	RNRE	1.11	0.96	0.95	0.99
				LNRE	1.10	1.04	0.93	0.98
				RNEE	0.90	1.06	0.92	0.98
				LNEE	0.70	1.17	0.90	0.97

Table A-5. Same as Table A-3, but for the N_0S_π interaural configuration. Some r^2 values were not significant ($r^2_{crit} = 0.16$ for $p < 0.05$) and one χ^2 value was not significant ($\chi^2_{crit} = 49.7$ for $p < 0.001$). These values are underlined.

S	Condition	P(Y T+N)			P(Y N)		
		χ^2	r^2	$V_{P(Y T+N)}$	χ^2	r^2	$V_{P(Y N)}$
S1	RNRE	159	0.16	0.57	206	0.52	0.84
	LNRE	182	0.63	0.89	233	0.60	0.87
	RNEE	215	0.43	0.79	185	0.33	0.73
	LNEE	178	0.40	0.77	140	0.35	0.74
S2	RNRE	349	0.76	0.93	150	0.35	0.74
	LNRE	560	0.79	0.94	98	<u>0.03</u>	0.30
	RNEE	258	0.42	0.79	115	0.21	0.63
	LNEE	297	0.68	0.90	74	<u>0.08</u>	0.44
S3	RNRE	348	0.87	0.97	69	<u>0.07</u>	0.42
	LNRE	267	0.51	0.83	61	<u>0.01</u>	0.18
	RNEE	209	0.59	0.87	77	<u>0.01</u>	0.18
	LNEE	155	0.36	0.75	<u>47</u>	<u>0.02</u>	0.25
S4	RNRE	582	0.77	0.93	697	0.89	0.97
	LNRE	743	0.79	0.94	828	0.83	0.95
	RNEE	610	0.83	0.95	692	0.87	0.97
	LNEE	937	0.92	0.98	1003	0.92	0.98
S5	RNRE	356	0.60	0.87	178	0.58	0.86
	LNRE	425	0.59	0.87	223	0.44	0.80
	RNEE	165	0.26	0.68	123	0.16	0.57
	LNEE	297	0.29	0.70	216	0.17	0.58
S _{avg}	RNRE	869	0.87	0.97	357	0.65	0.89
	LNRE	848	0.83	0.95	387	0.54	0.85
	RNEE	704	0.80	0.94	368	0.65	0.89
	LNEE	490	0.77	0.93	314	0.63	0.89

subjects with the highest thresholds had lower d' values (S3, Table A-2; S1 and S4, Table A-4), but for the majority of subjects and conditions the d' values were approximately unity. The majority of β values were also near unity, with some exceptions who (S1, Table A-2; S1 and S3, Table A-4) showed a slight bias ($\beta > 1$) to report “no tone” slightly more often. Detection patterns estimated within the different cue conditions were of primary interest, and accordingly, within-condition d' and β values were also of concern. Recall that the stimuli were randomly interleaved across conditions using a single SNR in order to foster a single decision strategy for each subject, and as a consequence, allowed experimenter control of only the overall d' and β . Thus, differences in d' and β from unity were not alarming, provided that the detection patterns were significantly reliable (see below).

In the N_0S_0 stimulus condition (Table A-2), d' values were generally lower for RN stimuli with respect to LN stimuli. These results were consistent with previous research that indicates, under diotic conditions, thresholds for LN stimuli are slightly lower than those for RN stimuli (Hartmann and Pumplin, 1988; Kohlrausch *et al.*, 1997; Eddins and Barber, 1998; and Eddins,

2001). S3 and S4 showed a slight bias toward tone responses for LN stimuli, which would be consistent with the use of envelope flattening as a detection cue.

Reliability statistics (Tables A-2 to A-5: r^2 , the correlation between estimates of the detection patterns computed from the first half of testing trials and from the last half of testing trials; and V , the proportion of predictable variance in the detection patterns) indicate that regardless of the within-condition deviations of d' and β from unity, N_0S_0 detection patterns were significant. The χ^2 values were significant ($\chi^2_{crit} = 49.7$ for $p < 0.001$) indicating that the variability in $P(Y|W)$ across waveform is more than expected by chance. The r^2 values were also significant [$r^2_{crit} = 0.16$ for $P(Y|T+N)$ or $P(Y|N)$, $r^2_{crit} = 0.18$ for $P(Y|W)$; $p < 0.05$], indicating that the probabilities estimated from the first half of listeners' responses were correlated to the probabilities estimated from the second half of the listeners responses.

The d' and β for the N_0S_π stimulus configuration were more variable across subjects than those for the N_0S_0 configuration. Table A-4 shows that in general, subjects had slightly lower d' values for the EE conditions and high β values for the LNEE condition with respect to the other three conditions. The low d' values observed in the EE conditions may be related to the level-equalization procedure. The right-ear and left-ear stimuli were scaled separately to the same overall level, resulting in a slight reduction of interaural-level differences for those stimuli. If listeners were relying on cues related to overall level differences between the ears, performance would have been poor. High β values for the LNEE condition indicated that listeners perceived both T+N and N stimuli as N stimuli. Such a perception could have been the result of temporally consistent envelopes (from the LN stimuli) combined with reduced interaural-level differences (from the normalization procedure) leading to consistent energies in each ear. The data from the N_0S_π condition were slightly less reliable on average. An exception was S3, whose data were very unreliable for N stimuli.

Tables A-2 through A-5 also show estimates of the proportions of predictable variance (V) computed separately for $P(Y|T+N)$, $P(Y|N)$, and $P(Y|W)$. These quantities are estimates of the upper bounds for predicting response variance across reproducible waveforms for particular subject, cue-condition, and interaural-configuration combinations, and are based on r^2 values using the formula described by Ahumada and Lovell (1971).

A.3.2 Comparisons between cue conditions

The primary purpose of this work was to investigate the relationships between detection patterns estimated under the various cue conditions. The design of the experiment called for three main comparisons between detection patterns (LNRE vs. RNRE, RNEE vs. RNRE, and LNEE vs. LNRE). These comparisons, along with comparisons of the variances of the detection patterns, indicated which of several possible cues (overall energy, temporal structure, temporal envelopes, and temporal fine structure) was primarily used for detection. The variance of a detection pattern is calculated by (separately) computing the variance of the z scores of the individual $P(Y|T+N)$ and $P(Y|N)$ values. Figures A-1 and A-2 show comparisons between detection patterns estimated under the different stimulus conditions for N_0S_0 and N_0S_π stimuli, respectively. The following description pertains to both figures. Each column corresponds to a particular comparison and each row to a particular subject. Each panel shows a scatter plot of the z -scores of $P(Y|T+N)$ (circles) and $P(Y|N)$ (squares) from two cue conditions. All z -scores were computed as the inverse of the normal cumulative distribution function (*cdf*) using $P(Y|W)$. Values of 0 or 1 were replaced with 1/48 or 47/48 respectively, such that the inverse *cdf* was defined. [Replacement with 1/48 occurred for 7 of the 1000 (2 interaural configurations x 4 conditions x 5 subjects by 25 noises) $P(Y|N)$ values that were equal to 0 and replacement with 47/48 occurred for 25 of the 1000 $P(Y|T+N)$ values that were equal to 1.] The scatter plots in the first column show comparisons between LNRE and RNRE $P(Y|W)$. In this comparison, the

temporal structures of corresponding waveform pairs differed, but the overall energies were positively correlated. ($r^2 \approx 0.62$; this correlation was reduced from 1 by the ramping procedure; when ramped, the overall energies of stimuli with different temporal structures but identical magnitude spectra, differ). The scatter plots in the second column show comparisons between RNEE and RNRE conditions. Corresponding waveform pairs in this comparison had identical temporal structures (and therefore identical “relative” magnitude and phase spectra), but differed in overall energies. In the third column, corresponding waveforms in the LNEE and LNRE conditions are compared. This comparison was similar to that shown in the second column, except stimuli were constructed using LN rather than RN phases. Thus, the waveforms in this comparison had less dynamic envelopes with respect to the RNEE-RNRE comparison. R^2 values are shown for each panel and were calculated using standard linear regression.

By comparing R^2 values across the three columns, inferences can be made about the cues used for detection. R^2 values near 0.62 in the first column and low R^2 values in the second and third columns would suggest that subjects relied on overall energies to perform the detection task. Low R^2 values in the first column and high R^2 values in the second and third columns would suggest that subjects used temporal structure to perform the detection task. Low R^2 values in the first column, high in the second, and low in the third would suggest that subjects used temporal envelopes to perform the detection task. These outcomes are summarized in Table A-1.

Because the upper limit of the expected R^2 values shown in the first column of Fig. A-1 is known to be smaller than 1, no tests of significant differences between correlations with column 1 are reported. Tests of correlated but non-overlapping correlations (Raghunathan *et al.*, 1996) were used to compare R^2 values from the second and third columns; significant values indicate that altering the envelopes of the stimuli influenced detection patterns (recall that energies were uncorrelated and magnitude spectra were the same within each comparison). Significant ($p < 0.05$) differences between R^2 values for each comparison are denoted with an asterisk. Results for the N_0S_0 and N_0S_π interaural configurations are presented separately below.

Differences between the variances of the detection patterns are presented in Fig. A-3. Differences in variance between the RE and EE conditions are shown in the upper row of each panel and differences in variance between the RN and LN conditions are shown in the lower row of each panel). Significant differences between the variances ($p < 0.05$; t test for homogeneity of variance for two dependent samples; Sheskin, 2000) of RE and EE detection patterns are indicated with asterisks in the upper row of each panel. Significant differences between the variances of LN and RN detection patterns are indicated with asterisks in the lower row of each panel⁵.

A.3.2.1. N_0S_0

Figure A-1 shows results for detection patterns estimated with N_0S_0 stimuli. All reported R^2 values were significant ($p < 0.05$). S1 and S2 had the highest R^2 values for the conditions in which temporal structures were preserved (columns 2 and 3), but showed no significant ($p < 0.05$) differences between R^2 values across the 3 comparisons. S3 through S5 (and the average subject) showed patterns of correlations similar to each other. For each of these listeners, R^2 values for comparisons in which temporal structures

⁵ Please note that the threshold for significant differences between variances depends upon more than the length of each bar shown in Fig. A-3, which shows only the differences in variances between the pairs of detection patterns tested. The test of *dependent samples* used for this procedure also depends on the correlation between each pair of detection patterns tested, as well as the overall variance of each pattern (not shown in Fig. A-3).

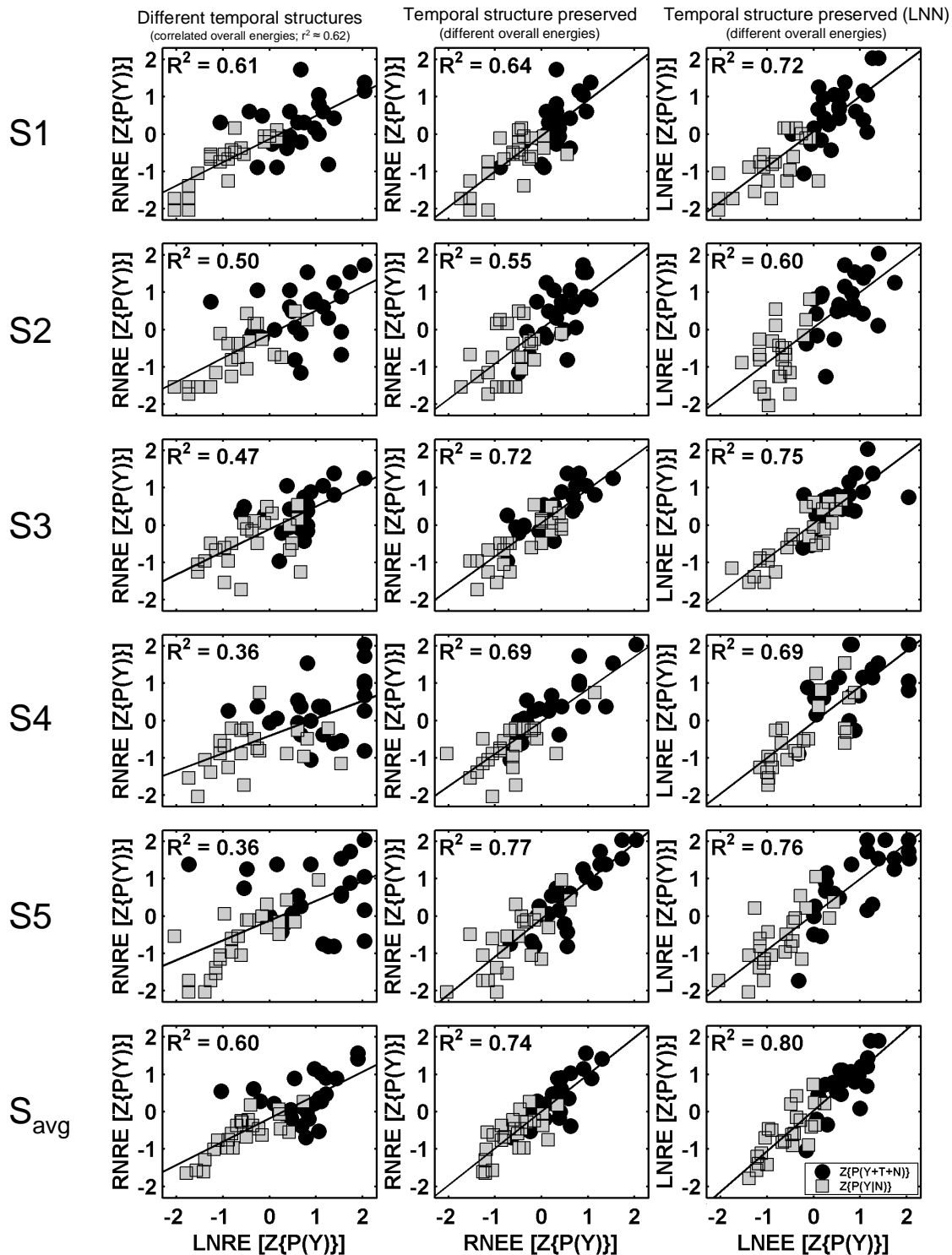


Figure A-1. Scatter plots showing comparisons between detection patterns estimated under the different cue conditions for the N_0S_0 interaural configuration. Each row represents a different subject, and each column a specific comparison. R^2 values are given for $z\{P(Y|W)\}$ and all R^2 values were significant ($p < 0.05$). Each point represents the z-score of a probability for an individual T+N (circles) or N (squares) waveform. If the detection patterns for each cue condition in each comparison were identical, all points would fall along the diagonal. Best-fit regression lines are shown for each comparison. Recall that waveform-overall energies are preserved in comparisons in the first column, waveform-temporal structures are preserved in comparisons in the second column and, and temporal structures using LNN waveforms are preserved in comparisons in the third column.

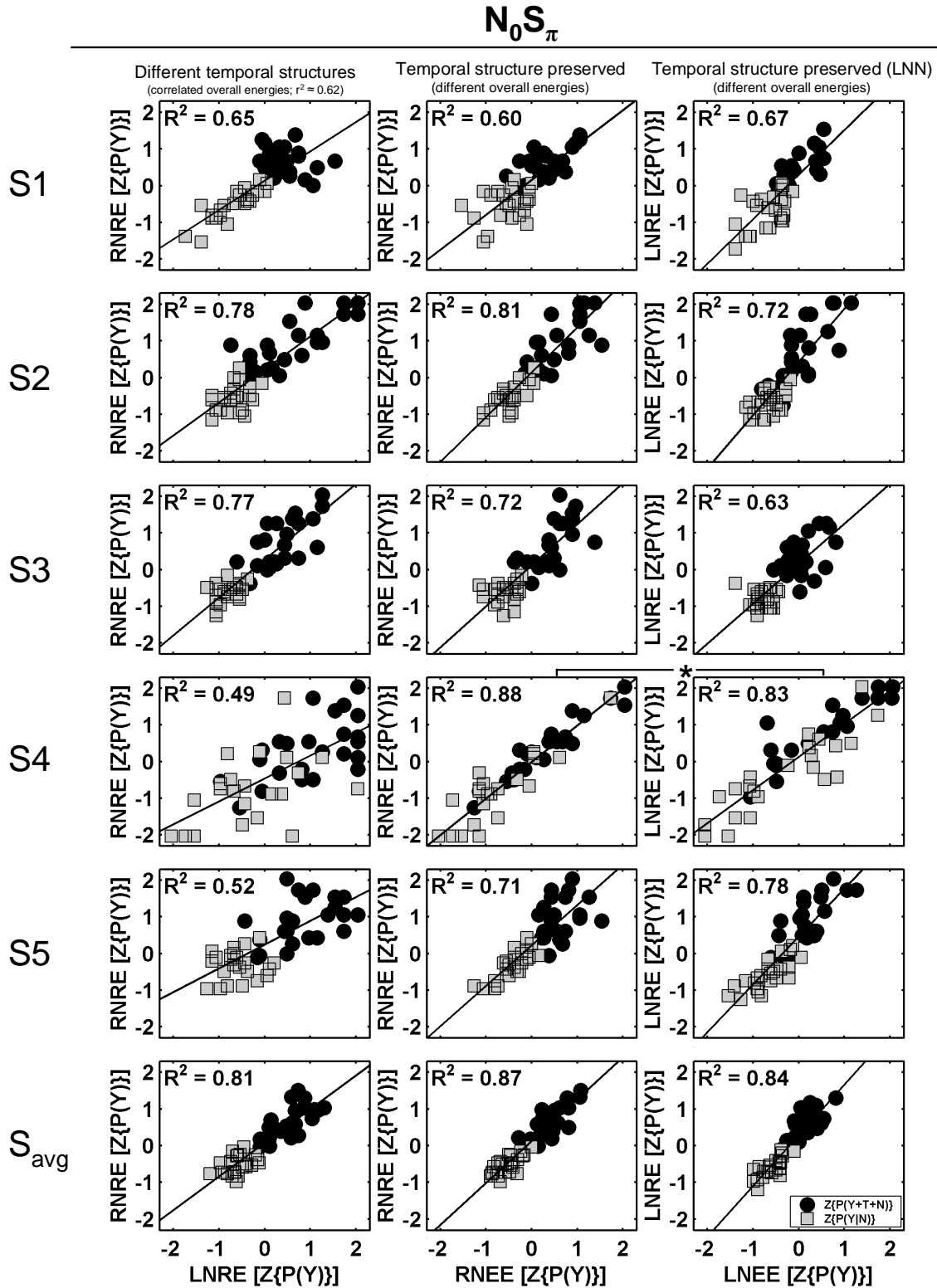


Figure A-2. Same as Fig. 1 except comparisons are shown for the $N_0 S_\pi$ interaural configuration. Stars indicate significant differences ($p < 0.05$) between correlations in the 2nd and 3rd columns.

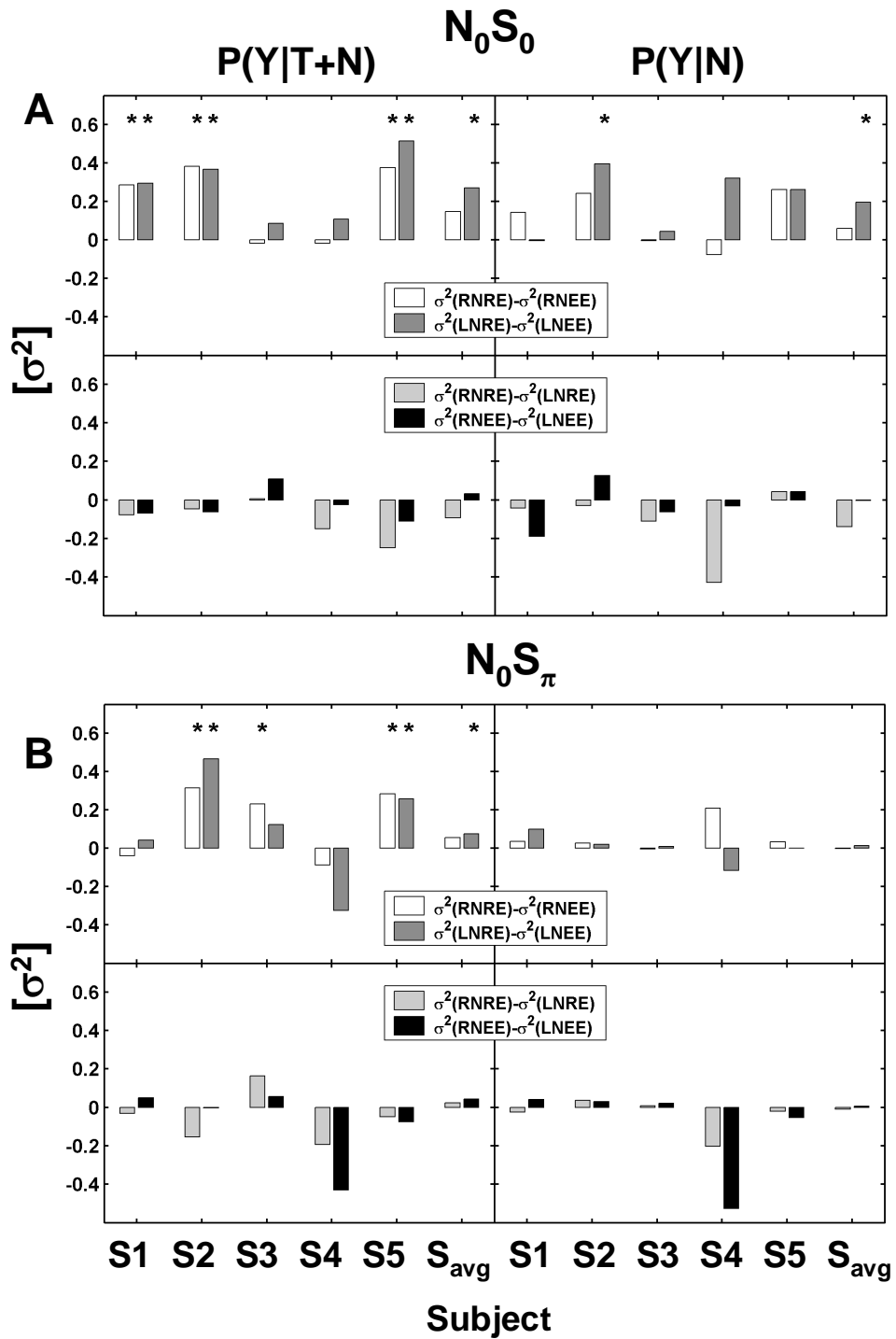


Figure A-3. Comparison of variances of detection patterns from the four cue conditions (within each subject). Each bar represents the difference of two variances. Significant differences ($p < 0.05$) between RE and EE conditions are shown in the top rows of panels A and B with an asterisk. Significant differences ($p < 0.05$) between RN and LN conditions are shown in the bottom rows of panel A and B with an asterisk.

were preserved (columns 2 and 3) were higher than R^2 values for comparisons in which temporal structures were not preserved (with correlated overall energies, column 1; tests of significant differences were not performed because the upper limit of the expected correlation in column 1 was 0.62). However, R^2 values for column 1 did approach the expected and maximum value for an energy cue for S1, S2 and S_{avg} .

Differences between the variances of the detection patterns in the RE and EE conditions are shown in Fig. A-3 A (upper row, for the N_0S_0 condition). Recall that under an energy model assumption, the variances of $P(Y|T+N)$ and $P(Y|N)$ from EE conditions are expected to be smaller than the variances of $P(Y|T+N)$ and $P(Y|N)$ from RE conditions. $P(Y|T+N)$ and $P(Y|N)$ are observed separately because overall energies were equalized separately for T+N and N stimuli, maintaining the average level difference between groups. Figure A-3 A (upper-left panel) show that of the 12 comparisons performed for $P(Y|T+N)$, 7 had significantly lower variances for EE conditions. Figure A-3 A, (upper-right panel) also shows the 12 comparisons of RE and EE variance performed for $P(Y|N)$. Here only 2 comparisons showed significantly lower variances in EE conditions. While these results do not argue strongly for the use of energy as a detection cue, note that in all but 3 of the 24 comparisons considered, the variances of RE detection patterns were larger than those of EE detection conditions. This finding suggests some effect of energy equalization, but is consistent with the notion that energy was certainly not the only cue used for detection. Nevertheless, the significant R^2 values in the first column of Figs. A-1 and A-2, and consistent changes in detection-pattern variance indicate that overall stimulus energy played a role as a cue for detection. This finding will be considered in the context of a basic energy model below.

Recall that if subjects were relying on cues related to the magnitude of envelope fluctuation to perform the detection task, the expected variability of detection patterns in the LN conditions would be significantly different than in the RN conditions. Under this hypothesis, R^2 values in the second column of Fig. A-1 (RNRE vs. RNEE) should also be significantly different from those in the third column of Fig. A-1 (LNRE vs. LNEE), because the temporal envelopes of the stimuli are significantly different between the two comparisons. No significant differences were found. Further, Fig. A-3 shows results from direct tests comparing the variances under RN and LN conditions separately for $P(Y|T+N)$ and $P(Y|N)$. None of the comparisons yielded significant differences in variance. Overall, these results indicate that listeners' decision variables were not strongly affected by a direct manipulation of temporal envelope fluctuations.

One of the main goals of this study was to examine the possible use of overall energy in tone-in-noise detection. Previous work with roving-level and equal-energy stimuli has demonstrated that subjects are able to use cues other than overall energy to perform a tone-in-noise detection task (Kidd *et al.*, 1989; Richards, 1992; Richards and Nekrich, 1993). Those studies examined only tone thresholds and asked what listeners are capable of doing without "natural" differences in overall energy. The question proposed here was quite different, and was inspired by the good energy-model fits achieved in Davidson *et al.* (2006) using reproducible stimuli. Here we ask if the "standard" tone-in-noise detection strategy is inherently different from that of an energy detector, or if listeners simply switch strategies in the face of roving-level or equal-energy stimuli. Put more simply, given the availability of overall energy cues, *how* are listeners performing the detection task? In this experiment, the average energy differences were preserved across T+N and N stimuli in the EE conditions while keeping the overall energies of all T+N waveforms equal and the overall energies of all N waveforms equal, allowing the use of an energy-based detection strategy in either the RE or EE conditions. The mere existence of a detection pattern with significant χ^2 values in the EE conditions is evidence against the use of a detection cue based entirely on overall energy. However, analyses of homogeneity of variance

indicated that detection patterns in the EE conditions tended to have smaller variances (although this level was above Bernoulli variability, as shown by the χ^2 tests in Sec. A.3.1), as would be expected if overall energy was used exclusively as a decision variable. Comparisons of P(Y|W) values between RNRE and LNRE conditions (which had correlated overall energies but different temporal structures) were also used to determine if listeners were using overall energy as a primary cue in the detection task. All subjects had significant R^2 values for this comparison and 3 of the 5 subjects had R^2 values that approached the expected correlation for an energy-based model. Such results were further explored using an energy model.

A simple energy-style model was implemented (the critical-band model from Davidson *et al.*, 2006) and used to predict the results of the present study. The model's decision variable was simply the energy at the output of a 4th-order gammatone filter centered at the tone frequency. Table A-6 shows the proportions of variance in the listeners' detection patterns explained the energy-model predictions. Note that R^2 values were significant R^2 values for RNRE and LNRE conditions for all subjects. The energy model was not used to predict EE stimuli, as energy variations in the output of the 75-Hz, 4th-order gammatone filter had a range of energy differences (within T+N or N

Table A-6. Proportions of variance in detection patterns estimated for N_0S_0 RE conditions by a model based on overall stimulus energy. All R^2 values (computed as the square of the correlation coefficient between energy model predictions and each subject's detection pattern) were significant ($p < 0.001$).

S	P(Y W)	
	RNRE	LNRE
S1	0.67	0.79
S2	0.73	0.84
S3	0.35	0.56
S4	0.34	0.67
S5	0.68	0.80
S_{avg}	0.67	0.82

waveforms) of less than 0.65 dB SPL (compared to 8.3 dB SPL for RE stimuli). Such small variations in energy would be undetectable due to internal noise.

Tests of correlated but non-overlapping correlations showed that the energy model predicted significantly ($p < 0.05$) more variance in LN detection patterns than in RN detection patterns for all subjects but S5. These results indicate that listeners behaved more like energy detectors for stimuli that had envelopes with reduced fluctuations. The results from Fig. A-2 showed a small effect of waveform temporal structure, but did not produce a striking pattern of R^2 values indicating envelope dominance as a detection cue (i.e., significantly different R^2 values in the second and third columns). Taken together, these results support the notion that overall energy prevails as a cue for detection when it is available, and that temporal variations in the stimuli also act as cues for detection when overall energy is made unreliable.

These findings suggested that when studying diotic tone-in-noise detection (with or without reproducible stimuli) it is important to prevent the use of overall energy as a detection cue (e.g., roving stimulus levels or equalizing stimulus energies). The use of such a simple

energy cue effectively dominates data that could be explored to learn about other cues important for detection of target signals in noise.

Corresponding waveforms in each of the cue conditions in the present study had the same spectral shapes, and detection patterns from each of the conditions were significantly correlated for most subjects. This indicates that spectral shape may have played a role as an additional detection cue. Green *et al.* (1992) reported on spectral shape discrimination of 20-Hz wide bands of Gaussian noise (with components spaced in 5-Hz increments) centered at 1000 Hz. They suggested that for stimuli falling within a single critical band the envelope power spectrum was a likely cue. Their model used a weighted sum of the normalized powers in 5 different modulation channels, and under-predicted listeners' thresholds by 3 to 5 dB when the signal was added to the center component of the spectrum. It is unclear whether such results are meaningful for the present study, given the 500-Hz signal frequency used here, due to the possibility that listeners may employ different strategies or operate using different cues at the two signal frequencies. Nevertheless, the model described by Green *et al.* (1992) is suggested as a candidate for future study in Ch. 4.

A.3.2.2 N_0S_π

Comparisons between detection patterns estimated with N_0S_π stimuli are shown in Fig. A-2. Recall that the reliability of the N_0S_π data was relatively poor compared to that of the N_0S_0 data (see Tables A-2 through A-5), consistent with the tightly clustered $z\{P(Y|W)\}$ values near the center of in each panel in Fig. A-2. This clustering would tend to obscure differences across the various cue conditions for some subjects.

If the magnitude of envelope fluctuations were related to the detection scheme used for N_0S_π stimuli, one would expect the variances of the detection patterns to be significantly different when estimated using LN rather than RN stimuli. Figure A-3 B, bottom shows that no significant differences in the variances of $P(Y|T+N)$ or $P(Y|N)$ between RN and LN conditions were observed, indicating that the magnitude of envelope fluctuations was not a primary cue for detection. The results of S4 showed significantly higher R^2 values for the second comparison than for the other two comparisons (but not because of reduced variance in the LN conditions), suggesting that this listener incorporated temporal-fine structure or the temporal structure of the entire waveform as part of the detection strategy. The R^2 values for the comparisons in the second and third columns of Fig. A-2 were not significantly different for the remaining subjects.

N_0S_π detection patterns were obscured by the unreliable nature of the data collected under those conditions. Unreliable detection patterns could have occurred for a number of reasons. One possible reason is the fact that level-equalization was performed binaurally, and may have led to the slightly lower d' values in EE conditions with respect to RE conditions (as shown in Table A-4) and also to the pattern of significant differences in variances of $P(Y|T+N)$ only between EE and RE conditions (Fig. A-3). Henning *et al.* (2005) performed a tone-in-noise detection experiment using 30-Hz wide, 110-ms duration noise centered at 500 Hz. They used a 40-dB dichotic rove that should have had a similar effect to equalizing overall energies across the two ears. They found that thresholds increased slightly (as observed here with the slightly lower d' values) for dichotic-rove conditions with respect to conditions where the levels were not roved or were roved diotically. As in the Henning *et al.* (2005) study, if listeners had adopted a strategy such as lateralization during training (which makes use of interaural level differences), the level equalization scheme could have led to less reliable results during testing. The present results do not point to the consistent use of particular N_0S_π waveform features by any of the subjects in this study.

A.3.3 Comparisons between subjects and between interaural configurations

Tables A-7 and A-8 show intersubject correlations for the four cue conditions in the experiment. (Correlations are presented in terms of r^2 , or the square of the correlation coefficient, which can be interpreted as the percentage of variance in one subject's detection pattern that is explained by the variance in another subject's detection pattern.)

Table A-7. Comparisons between subjects' [P(Y|W)] presented in terms of r^2 , the square of the correlation coefficient.

Interaural configuration	Intersubject Comparison	P(Y W)			
		RNRE	LNRE	RNEE	LNRE
N_0S_0	S1-S2	0.61*	0.86*	0.67*	0.69*
	S1-S3	0.57*	0.76*	0.58*	0.72*
	S1-S4	0.51*	0.76*	0.71*	0.69*
	S1-S5	0.72*	0.78*	0.72*	0.78*
	S2-S3	0.52*	0.72*	0.51*	0.58*
	S2-S4	0.50*	0.78*	0.59*	0.62*
	S2-S5	0.62*	0.80*	0.61*	0.66*
	S3-S4	0.52*	0.80*	0.52*	0.76*
	S3-S5	0.54*	0.65*	0.49*	0.57*
	S4-S5	0.54*	0.73*	0.55*	0.59*
N_0S_π	S1-S2	0.54*	0.37*	0.43*	0.15*
	S1-S3	0.57*	0.39*	0.57*	0.26*
	S1-S4	0.44*	0.53*	0.49*	0.47*
	S1-S5	0.53*	0.43*	0.57*	0.13*
	S2-S3	0.81*	0.80*	0.65*	0.64*
	S2-S4	0.35*	0.25*	0.35*	0.05
	S2-S5	0.68*	0.80*	0.58*	0.54*
	S3-S4	0.35*	0.28*	0.26*	0.15*
	S3-S5	0.71*	0.77*	0.66*	0.56*
	S4-S5	0.37*	0.22*	0.40*	0.00

* $p < 0.05$

Table A-8. Comparisons between subjects for [P(Y|T+N)] and [P(Y|N)] considered separately, presented in terms of r^2 , the square of the correlation coefficient.

Interaural configuration	Intersubject Comparison	P(Y T+N)			
		RNRE	LNRE	RNEE	LNRE
N ₀ S ₀	S1-S2	0.49*	0.72*	0.44*	0.04
	S1-S3	0.44*	0.62*	0.37*	0.43*
	S1-S4	0.30*	0.59*	0.53*	0.24*
	S1-S5	0.57*	0.65*	0.45*	0.50*
	S2-S3	0.49*	0.54*	0.39*	0.20*
	S2-S4	0.19*	0.74*	0.32*	0.33*
	S2-S5	0.51*	0.66*	0.28*	0.04
	S3-S4	0.27*	0.56*	0.30*	0.37*
	S3-S5	0.40*	0.66*	0.27*	0.28*
	S4-S5	0.35*	0.68*	0.31*	0.35*
N ₀ S _π	S1-S2	0.02	0.00	0.06	0.01
	S1-S3	0.04	0.07	0.15	0.04
	S1-S4	0.22*	0.27*	0.30*	0.30*
	S1-S5	0.03	0.00	0.35*	0.03
	S2-S3	0.53*	0.61*	0.23*	0.49*
	S2-S4	0.09	0.01	0.13	0.00
	S2-S5	0.30*	0.72*	0.10	0.27*
	S3-S4	0.09	0.01	0.05	0.01
	S3-S5	0.46*	0.46*	0.32*	0.33*
	S4-S5	0.17*	0.01	0.20*	0.10

Interaural configuration	Intersubject Comparison	P(Y N)			
		RNRE	LNRE	RNEE	LNRE
N ₀ S ₀	S1-S2	0.31*	0.70*	0.29*	0.34*
	S1-S3	0.31*	0.60*	0.56*	0.63*
	S1-S4	0.23*	0.58*	0.69*	0.69*
	S1-S5	0.61*	0.52*	0.50*	0.32*
	S2-S3	0.15	0.52*	0.38*	0.41*
	S2-S4	0.36*	0.52*	0.57*	0.42*
	S2-S5	0.33*	0.61*	0.20*	0.30*
	S3-S4	0.35*	0.71*	0.51*	0.72*
	S3-S5	0.30*	0.28*	0.40*	0.26*
	S4-S5	0.30*	0.47*	0.41*	0.22*
N ₀ S _π	S1-S2	0.04	0.01	0.00	0.04
	S1-S3	0.01	0.13	0.17*	0.03
	S1-S4	0.12	0.27*	0.27*	0.49*
	S1-S5	0.01	0.00	0.02	0.09
	S2-S3	0.13	0.12	0.00	0.00
	S2-S4	0.01	0.01	0.04	0.04
	S2-S5	0.08	0.17*	0.01	0.22*
	S3-S4	0.01	0.00	0.01	0.01
	S3-S5	0.02	0.16	0.00	0.00
	S4-S5	0.02	0.04	0.09	0.29*

* p < 0.05

Compared to previous work (Davidson *et al.* 2006), these correlations were slightly lower, which may have been a consequence of the reduced number of stimulus presentations (approximately 48 vs. 96) and increased within-subject variability. Correlations were lower for comparisons made in the N_0S_π condition than in the N_0S_0 condition. These reduced correlations between interaural configurations were partially a consequence of the lower reliability of the data from the N_0S_π configuration; but, in general, intersubject correlations were lower for dichotic than for diotic configurations (e.g., Evilsizer *et al.*, 2002; Isabelle, 1995). There was a relationship between intersubject correlation and threshold-tone levels, particularly for the N_0S_π configuration. Subjects with similar thresholds had more correlated detection patterns. For example, correlations between S1 and S4, S2 and S3, and S3 and S5 were significant for T+N stimuli in all four cue conditions in both interaural configurations. These subjects' thresholds differed at most by only 1 dB. Very low intersubject correlations occurred for pair-wise comparisons between S1 and S2, S1 and S3, and S1 and S5 for T+N stimuli; these subjects' thresholds differed by at least 10 dB. Similar patterns of correlations also occurred for responses to N stimuli, although not as frequently.

There was a relationship between cue condition and intersubject correlation for responses in the N_0S_0 configuration. Detection patterns were most similar between subjects in the LNRE condition. High between-subject correlations could be a consequence of high χ^2 values and relatively high d' values observed for the LNRE condition (model-data comparisons are also relatively higher for the LNRE conditions, as described in Ch.3), but likely also reflects the use of similar strategies among subjects for the LNRE condition.

Perhaps the most valuable information to come from these comparisons was that subjects with high N_0S_π thresholds (and thus with small MLDs, such as S1 and S4) tended to have correlated $P(Y|N)$ values, both between subjects and between interaural correlations, whereas subjects with low N_0S_π thresholds (and thus large MLDs, such as S2, S3, and S5) tended to have low intersubject correlations and low correlations between interaural configurations. Tables A-2 through A-5 reveal that S1 and S4 had slightly more reliable detection patterns (with higher predictable variances) which could account for part of this effect. A review of Evilsizer *et al.* (2002, Table A-4, S2) reveals high correlations between $P(Y|N)$ values for the subjects with the highest N_0S_π thresholds, indicating that this phenomenon was not unique to this study. Overall, this finding suggests that subjects with higher tone thresholds used strategies that were similar both across subjects and across interaural configurations. Subjects with lower thresholds used strategies for N_0S_π stimuli that were not correlated across subjects or with strategies used for N_0S_0 stimuli.

Tables A-9 and A-10 show correlations between detection patterns estimated using N_0S_0 stimuli and using N_0S_π stimuli. Values of r^2 are reported within each cue condition for each subject. The strongest between-configuration correlations occurred for subjects S1 and S4 across cue conditions. Recall that corresponding noise-alone waveforms were identical across the two interaural configurations. This pattern of correlations is consistent with the fact that S1 and S4 had the highest thresholds for N_0S_π stimuli and also the smallest differences between N_0S_0 and N_0S_π thresholds. The high correlation values suggest that these listeners were using a strategy for N_0S_π stimuli that was similar to the strategy used for N_0S_0 stimuli, accounting for the small differences

Table A-9. Comparisons between interaural configurations for $P(Y|W)$ presented in terms of r^2 , the square of the correlation coefficient.

S	P(Y W)			
	RNRE	LNRE	RNEE	LNRE
S1	0.51*	0.54*	0.64*	0.53*
S2	0.19*	0.14*	0.43*	0.29*
S3	0.28*	0.28*	0.20*	0.25*
S4	0.57*	0.49*	0.66*	0.68*
S5	0.17*	0.23*	0.29*	0.27*
S _{avg}	0.47*	0.45*	0.66*	0.67*

* $p < 0.05$

Table A-10. Comparisons between interaural configurations for $P(Y|T+N)$ and $P(Y|N)$ considered separately, in terms of r^2 , the square of the correlation coefficient.

S	P(Y T+N)				P(Y N)			
	RNRE	LNRE	RNEE	LNRE	RNRE	LNRE	RNEE	LNRE
S1	0.02	0.03	0.18*	0.04	0.66*	0.75*	0.56*	0.56*
S2	0.19*	0.31*	0.00	0.07	0.00	0.00	0.01	0.03
S3	0.00	0.01	0.00	0.17*	0.01	0.00	0.08	0.08
S4	0.17*	0.01	0.37*	0.40*	0.85*	0.79*	0.80*	0.89*
S5	0.01	0.10	0.00	0.02	0.02	0.00	0.09	0.01
S _{avg}	0.00	0.17*	0.16	0.01	0.50*	0.60*	0.54*	0.72*

* $p < 0.05$

between N_0S_0 and N_0S_π thresholds in these listeners. Correlations between interaural configurations for responses to T+N stimuli were significant in only a few cases and were, on average, smaller than correlations reported for responses to N stimuli.

Consider again that corresponding N_0S_0 and N_0S_π waveforms were identical when the tone was not present, but were very different when the tone was present. One might then expect $P(Y|N)$ values to be more similar between the two interaural configurations than $P(Y|T+N)$ values, provided that the same or similar detection strategies were used for each interaural configuration. However, such a pattern of interaural correlations would suggest of the use of different strategies for T+N and N trials. Given that listeners had no way of knowing *a priori* which trials contained the tone, the use of a strategy conditional on tone presence would have been impossible. The application of a diotic detection strategy to N_0S_π stimuli would not necessarily have produced $P(Y|T+N)$ values that were correlated across interaural configurations. However, even the partial use (e.g., on some trials) of a diotic strategy would certainly cause increased tone thresholds with respect to the use of a true binaural detection strategy.

A.4 Conclusions and future directions.

The primary conclusion of this work was that if overall energy was available as a cue for diotic tone-in-noise detection, listeners tended to use the energy-based cue, which explained up to 87 percent of the variance in $P(Y|W)$ for diotic stimuli. Accordingly, unless researchers are interested in examining the role of energy in tone-in-noise detection, future studies should

employ a scheme to make energy cues unreliable in studies of N_0S_0 detection in order to reveal less obvious and more interesting detection strategies.

There are several considerations worth emphasizing for the design of studies that compare detection patterns estimated using interleaved stimuli with arbitrary manipulations. First, one must be sure that the stimulus manipulations do not affect thresholds in the individual cue conditions to the point that the data may become unreliable. Second, the assessment of listeners' strategies based upon the changes in the variance of detection patterns [or alternatively, internal-to-external noise ratios (see Siegel and Colburn, 1989)] is not recommended because the necessary tests lack statistical power and there is great intersubject variability. Third, when comparing across cue conditions, a carefully selected "control" or "baseline" condition should be provided such that differences between this baseline condition and other conditions may be used when relative differences in correlations are small. Finally, a large number of trials is necessary ($\gg 50$) to establish highly reliable detection patterns. This number is larger than the numbers used in previous studies that did not interleave cue conditions within blocks. Future studies will make use of these recommendations to more directly observe the roles of temporal envelope and fine structure in diotic and dichotic tone-in-noise detection.

In Ch. 2, a more direct strategy was used to examine the roles of envelope and fine structure in diotic and dichotic signal detection. The experiment described in Ch. 2 incorporated much of what was learned from this study, including: double the number of trials for more reliable data, manipulation of stimulus components without assumptions about individual cues (other than energy; e.g., the 4th moment of the stimulus envelope), energy equalization for N_0S_0 stimuli (to effectively remove overall energy as a possible cue), and no energy equalization under N_0S_π conditions in order to prevent unnecessary interaural stimulus manipulations. Chapter 3 will address prediction of the detection patterns collected in the present study using traditional binaural decision variables, as well as various temporal detection schemes.

BIBLIOGRAPHY

- Ahumada, A., Marken, R., and Sandusky, A. (1975). "Time and frequency analyses of auditory signal detection," *J. Acoust. Soc. Am.* **57**, 385-390.
- Ahumada, A., and Lovell, J. (1971). "Stimulus features in signal detection," *J. Acoust. Soc. Am.* **49**, 1751-1756.
- Akeroyd, M. A. and Bernstein, L. R. (2001). "The variation across time of sensitivity to interaural disparities: Behavioral measurements and quantitative analyses," *J. Acoust. Soc. Am.* **110**, 2516-2526.
- Berg, B. G., Nguyen, Q. T., and Green, D. M. (1992). "Discrimination of narrow-band spectra. I. Spectral weights and pitch cues," *J. Acoust. Soc. Am.* **92**, 1911-1918.
- Boudreau, J.C. and Tsuchitani, C. (1968). "Binaural interaction in the cat superior olive S segment," *J. Neurophysiol.* **31**, 442-54.
- Blauert, J. (1981). "Lateralization of jittered tones," *J. Acoust. Soc. Am.* **70**, 694-698.
- Breebaart, J., van der Par, S., and Kohlrausch, A. (1999). "The contribution of static and dynamically varying ITDs and IIDs to binaural detection," *J. Acoust. Soc. Am.* **106**, 979-992.
- 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.
- Breebaart, J., van der Par, S., and Kohlrausch, A. (2001b). "Binaural processing model based on contralateral inhibition I. Dependence on spectral parameters," *J. Acoust. Soc. Am.* **110**, 1089-1104.
- Breebaart, J., van der Par, S., and Kohlrausch, A. (2001c). "Binaural processing model based on contralateral inhibition I. Dependence on temporal parameters," *J. Acoust. Soc. Am.* **110**, 1105-1117.
- Carney L. H., Heinz, M. G., Evilsizer, M. E., Gilkey, R. H., and Colburn, H.S. (2002). "Auditory phase opponency: A temporal model for masked detection at low frequencies," *Acta Acustica United with Acustica* **88**, 334-347.
- 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.
- Colburn, H. S., Isabelle, S. K., and Tollin, D. J. (1997). "Modeling binaural detection performance for individual masker waveforms," in R. Gilkey & T. Anderson (Eds.), *Binaural and Spatial Hearing in Real and Virtual Environments*, (New York, Erlbaum, pp. 533-556).
- Culling, J. F., and Summerfield, Q. (1998). "Measurements of the binaural temporal window using a detection task," *J. Acoust. Soc. Am.* **61**, 3540-3553.
- Dau, T., Kollmeier, D., and Kohlrausch, A. (1997a). "Modeling auditory processing of amplitude modulation: I. Detection and masking with narrowband carriers," *J. Acoust. Soc. Am.* **102**, 2892-2905.
- Dau, T., Kollmeier, B., and Kohlrausch, A. (1997b). "Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration in modulation detection," *J. Acoust. Soc. Am.* **102**, 2906-2919.
- 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.
- 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.
- Diercks, J. K., and Jeffress, L. A. (1962). "Interaural phase and the absolute threshold for tone," *J. Acoust. Soc. Am.* **34**, 981-984.
- Durlach, N. I. (1963). "Equalization and cancellation theory of binaural masking-level differences," *J. Acoust. Soc. Am.* **35**, 1206-1218.
- Durlach, N. (1972). "Binaural signal detection: Equalization and cancellation theory," in J.V. Tobias *Foundations of modern Auditory Theory*. Vol. 2 (New York, Academic Press, p.p. 2752-2760).
- Durlach, N. and Colburn, H. S. (1978). "Binaural Phenomena," in M. P. Friedman *Handbook of Perception*. Vol 4 (New York, Academic Press, p.p. 365-466).
- Eddins, D. A. (2001). "Monaural masking release in random-phase and low-noise noise," *J. Acoust. Soc. Am.* **109**, 1538-1549.
- 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.
- Edwards, A. L. (1979). *Multiple Regression and the Analysis of Covariance*. (New York, W. H. Freeman).

- Egan, J. P. (1965). "Masking-level differences as a function of interaural disparities in intensity of signal and noise," *J. Acoust. Soc. Am.* **38**, 1043-1049.
- Evilsizer M. E., Gilkey R. H., Mason C. R., Colburn H. S., and Carney L. H. (2002). "Binaural detection with narrowband and wideband reproducible noise maskers: I. Results for human," *J. Acoust. Soc. Am.* **111**, 336-345.
- Fitzpatrick, D. C., Kuwada, S., and Batra, R. (2000). "Neural Sensitivity to Interaural Time Differences: Beyond the Jeffress Model," *J. Neurosci.* **20**, 1605-1615.
- 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.
- Gilkey, R. H., Robinson, D. E., and Hanna, T. E. (1985). "Effects of masker waveform and signal-to-masker phase relation on diotic and dichotic masking by reproducible noise," *J. Acoust. Soc. Am.* **78**, 1207-1219.
- Glasberg, B. R., and Moore, B. J. C. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hearing Res.* **47**, 103-138.
- Goldberg, J. M. and Brown, P. B. (1969). "Responses of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization," *J. Neurophysiol.* **32**, 613-635.
- Goupell, M. J. (2005). "The use of interaural parameters during incoherence detection in reproducible noise," Ph.D. dissertation, Michigan State University.
- Goupell, M. J. and Hartmann, W. H. (2005). "Binaural models for the detection of interaural coherence," *J. Acoust. Soc. Am.* **117**, 2484 (Abs.).
- Goupell, M. J. and Hartmann, W. H. (2006). "Interaural fluctuations and the detection of interaural incoherence: Bandwidth effects," *J. Acoust. Soc. Am.* **119**, 3791-3986.
- Grantham, D. W. and Wightman, F. L. (1979). "Detectability of time-varying interaural correlation in narrow-band noise stimuli," *J. Acoust. Soc. Am.* **65**, 1509-1517.
- Green, D. M. (1983). "Profile analysis: A different view of auditory intensity discrimination," *J. Acoust. Soc. Am.* **38**, 133-142.
- Green, D. M., Berg, B. G., Huanping, D., Eddins, D. A., Onsan, Z., and Nguyen, Q. (1992). "Spectral shape discrimination of narrow-band sounds," *J. Acoust. Soc. Am.* **92**, 2586-2597.
- Haftner, E. R. (1971). "Quantative evaluation of a lateralization model of masking-level differences," *J. Acoust. Soc. Am.* **50**, 1116-1122.
- Hancock, K., and Delgutte, B. (2004). "A physiologically based model of interaural time difference discrimination," *J. Neurosci.* **24**, 7110-7117.
- Hartmann W. M. and Pumplin, J. (1988). "Noise power fluctuations and the masking of sine signals," *J. Acoust. Soc. Am.* **83**, 2277-2289.
- Henning, B. G., Richards, V. M., and Lentz, J. J. (2005). "The effect of diotic and dichotic level-randomization on the binaural masking-level difference," *J. Acoust. Soc. Am.* **118**, 3229-3240.
- Heinz, M. G., Zhang, X., Bruce, I. C., and Carney, L. H. (2001). "Auditory-nerve model for predicting performance limits of normal and impaired Listeners," *ARLO* **2** (3).
- Hind, J. E., Goldberg, J. M., Greenwood, D. D., and Rose, J. E. (1963). "Some discharge characteristics of single neurons in the inferior colliculus of the cat. II. Timing of the discharges and observations on binaural summation," *J. Neurophysiol.* **26**, 321-341.
- Hirsh, I. J. (1948). "The influence of interaural phase on interaural summation and inhibition," *J. Acoust. Soc. Am.* **20**, 761-766.
- Isabelle, S. K. (1995). "Binaural detection performance using reproducible stimuli," Ph.D. Dissertation, Boston University.
- Isabelle, S. K., and Colburn H. S. (1991). "Detection of tones in reproducible narrow-band Noise," *J. Acoust. Soc. Am.* **89**, 352-359.
- Jeffress, L. A. (1948). "A Place Theory for Sound Localization," *J. Comp. Physiol. Psychol.* **41**, 35-39.
- Jeffress, L. A., Blodgett, H. C., and Deatherage, B. H. (1952). "Masking of tones by white noise as a function of the interaural phase of both components. I. 500 cycles," *J. Acoust. Soc. Am.* **24**, 523-527.
- Joris, P. X. (1996). "Envelope Coding in the Lateral Superior Olive. II. Characteristic Delays and Comparison with Responses in the Medial Superior Olive," *J. Neurophys.* **76**, 2137-2156.
- Joris, P. X. (2003). "Interaural Time Sensitivity Dominated by Cochlea-Induced Envelope Patterns," *J. Neurosci.* **23**, 6345-6350.

- Joris, P. X., and Yin, T. C. T. (1992). "Responses to amplitude-modulated sounds in the auditory nerve of the cat," *J. Acoust. Soc. Am.* **91**, 215-232.
- Joris, P. X., and Yin, T. C. T. (1995). "Envelope Coding in the Lateral Superior Olive. I. Sensitivity to Interaural Time Differences," *J. Neurophys.* **73**, 1043-1062.
- Kay, R. H. (1982). "Hearing of modulation in sounds," *Physiol. Rev.* **62**, 894-975.
- Kiang, N. Y. S., Watanabe, T., Thomas, E. C., and Clark, L. F. (1965). *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve*. MIT, Cambridge, MA.
- Kidd, G., Mason, C. R., Brantley, M. A., and Owen, G. A. (1989). "Roving-level tone-in-noise detection," *J. Acoust. Soc. Am.* **86**, 1310-1317.
- Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, R., van de Par, S., Oxenham, A. J., and Puschel, D. (1997). "Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations," *Acta Acoustica* **83**, 659-669.
- Kolarik, A. J. and Culling, J. (2005). "Measuring the binaural temporal window," *J. Acoust. Soc. Am.* **117**, 2563. (Abs.)
- 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.
- Kuwada S., Stanford T. R., and Batra, R. (1987). "Interaural phase-sensitive units in the inferior colliculus of the unanesthetized rabbit: effects of changing frequency," *J Neurophysiol.* **57**, 1338-1360.
- Levitt. H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust Soc Am.* **49**, 467-477.
- Macmillan, N. A., and Creelman, C. D. (1991). *Detection Theory: A User's Guide*, (Cambridge U.P., New York).
- Marquardt, T., and McAlpine, D. (2001). "Simulation of binaural unmasking using just four binaural channels," *J. Assoc. Res. Otolaryn.* (Abs. 21716)
- McAlpine, D., Jiang, D., and Palmer, A. R., (2001). "A neural code for low-frequency sound localization in mammals," *Nature Neuroscience* **4**, 396-401.
- Metz, P. J., Bismarck, G. von, and Durlach, N. (1968). "Further results on binaural unmasking and the EC model. II. Noise bandwidth and interaural phase," *J. Acoust. Soc. Am.* **43**, 1085-1091.
- Moore, B. J. C. (1975). "Mechanisms of masking," *J. Acoust. Soc. Am.* **57**, 391-399.
- Moore, B. J. C. (2003). *An Introduction to the Psychology of Hearing*, (Academic Press, New York).
- Oppenheim, A. V., Schaffer, R. W., and Buck, J. R. (1999). *Discrete-time signal processing. 2nd ed.* (Prentice-Hall, New Jersey).
- Pfafflin, S. M., and Mathews, M.V. (1966). "Detection of auditory signals in reproducible noise," *J. Acoust. Soc. Am.* **39**, 340-345.
- Pumplin, J. (1985). "Low-noise noise," *J. Acoust. Soc. Am.* **78**, 100-104.
- Raghuathan, T. E., Rosenthal, R. and Rubin, D. B. (1996). "Comparing correlated but nonoverlapping correlations," *Psychological Methods* **1**, 178-183.
- Richards, V. M. (1992). "The detectability of a tone added to narrow bands of equal energy noise," *J. Acoust. Soc. Am.* **91**, 3424-3425.
- Richards, V. M., and Nekrich, R. D. (1993). "The Incorporation of level and level-invariant cues for the detection of a tone added to noise," *J. Acoust. Soc. Am.* **94**, 2560-2574.
- Shackelton, T. M. and Palmer, A. R. (2006). "Contributions of intrinsic neural and stimulus variance to binaural sensitivity," *J. Assoc. Res. Otolaryngology* **7**, 425-442.
- Shaw, W. A., Newman, E. B. and Hirsh, I. J. (1947). "The difference between monaural and binaural thresholds," *J. Exp. Psychol.* **37**, 229-242.
- Sheskin, D. J. (2000). *Parametric and Nonparametric Statistical Procedures*, (Chapman and Hall/CRC, New York).
- Siegel, R. A., and Colburn, H. S. (1989). "Binaural processing of noisy stimuli: Internal/external noise ratios under diotic and dichotic stimulus conditions," *J. Acoust. Soc. Am.* **86**, 2122-2128.
- Smith, B. K., Ulrich, S. K., Kohlrausch, A., and Schroeder, M. R. (1986). "Phase effects in masking related to dispersion in the inner ear," *J. Acoust. Soc. Am.* **60**, 1631-1637.
- Smith, Z. M., Delgutte, B., and Oxenham, A. J. (2002). "Chimaeric sounds reveal dichotomies in auditory perception," *Nature* **416**, 87-90.
- Tollin, D. J., and Yin, T. C. (2005). "Interaural phase and level difference sensitivity in low-frequency neurons in the lateral superior olive," *J. Neurosci.* **25**, 10648-10657.
- van de Par, S. and Kohlrausch, A. (1998). "Diotic and dichotic detection using multiplied-noise maskers," *J. Acoust. Soc. Am.* **103**, 2100-2110.
- Yin, T. C. T., and Chan, J. C. K. (1990). "Interaural time sensitivity in the Medial Superior Olive of Cat," *J. Neurophys.* **64**, 465-488.

- Zeng, F., Nie, K., Liu, S., Stickney, G. Del Rio, E., Kong, Y., and Chen, H. (2004). "On the dichotomy in auditory perception between temporal envelope and fine structure cues," *J. Acoust. Soc. Am.* **116**, 1351-1354.
- Zhang, X. (2004). "Cross-frequency coincidence detection in the processing of complex sounds," Ph.D. Dissertation, Boston University.
- 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 Rabbit," *J. Acoust. Soc. Am.* **111**, 346-356.
- Zilany, Muhammad S. A. and Bruce, Ian 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.
- Zurek, P. M. (1991). "Probability distributions of interaural phase and level differences in binaural detection stimuli," *J. Acoust. Soc. Am.* **90**, 1927-1932.

Vita

Name of author: Sean A. Davidson

Place of birth: Pittsfield, MA

Date of birth: July, 9 1979

Graduate and undergraduate schools attended:

Syracuse University, Syracuse, NY

Degrees awarded:

Master of Science in Bioengineering, 2003, Syracuse University

Bachelor of Science in Bioengineering, 2001, Syracuse University

Awards, honors, and professional experience:

Teaching Assistant Department of Bioengineering, 2001

Research Assistant, Department of Bioengineering, 2003

National Institutes of Health: Ruth L. Kirschstein National Research Service Award

Individual Fellowship, 2006