Binaural detection with narrowband and wideband reproducible noise maskers. III. Monaural and diotic detection and model results

Sean A. Davidson
Department of Biomedical and Chemical Engineering, Institute for Sensory Research, 621 Skytop Road, Syracuse University, Syracuse, New York 13244

Robert H. Gilkey
Department of Psychology, Wright State University, Dayton, Ohio, 45435 and Human Effectiveness Directorate, Air Force Research Laboratory, Wright-Patterson Air Force Base, Ohio 45433

H. Steven Colburn
Boston University Hearing Research Center, Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts 02215

Laurel H. Carney[a)]
Department of Biomedical and Chemical Engineering, Department of Electrical Engineering and Computer Science, and The Institute for Sensory Research, 621 Skytop Road, Syracuse University, Syracuse, New York 13244

(Received 5 December 2004; revised 13 January 2006; accepted 23 January 2006)

A single-interval, yes-no, tone-in-noise detection experiment was conducted to measure the proportion of "tone present" responses to each of 25 reproducible noise-alone and tone-plus-noise waveforms under narrowband (100 Hz), wideband (2900 Hz), monotic, and diotic stimulus conditions. Proportions of "tone present" responses (estimates of the probabilities of hits and false alarms) were correlated across masker bandwidths and across monotic and diotic conditions. Two categories of models were considered; one based on stimulus energy or neural counts, and another based on temporal structure of the stimulus envelope or neural patterns. Both categories gave significant correlation between decision variables and data. A model based on a weighted combination of energy in multiple critical bands performed best, predicting up to 90% of the variance in the reproducible-noise data. However, since energy-based models are unable to successfully explain detection under a roving-level paradigm without substantial modification, it is argued that other variations of detection models must be considered for future study. Temporal models are resistant to changes in threshold under roving-level conditions, but explained at most only 67% of the variance in the reproducible-noise data.
 [DOI: 10.1121/1.2177583]

PACS number(s): 43.66.Dc, 43.66.Ba [AK] Pages: 2258–2275

I. INTRODUCTION

The simple task of detecting a pure tone in the presence of a noise masker has been studied for more than half a century (e.g., Fletcher, 1940; Hawkins and Stevens, 1950; Jeffress, 1968; Patterson, 1976; Kidd *et al.*, 1989; Richards *et al.*, 1992). Yet, despite extensive study, a definitive explanation of the underlying mechanisms has not emerged. To a first approximation, detection performance can be predicted on the basis of differences in energy statistics between the noise-alone and tone-plus-noise stimuli in a narrow band of frequencies close to the tone frequency. This "critical-band" model can predict the bulk of the masking data and also forms the foundation for much psychoacoustic theory and research. However, the critical-band model is clearly wrong in detail. For example, single-channel energy-based models cannot explain results using the roving-level paradigm (e.g., Kidd *et al.*, 1989). Physiologically, it is known that cochlear tuning is level dependent (Rhode, 1971), and that interactions among frequency bands are inherent (e.g., suppression). Although the rate of auditory-nerve discharge varies monotonically with stimulus energy, the temporal pattern of discharge also has the potential to code differences between the tone-plus-noise and noise-alone stimuli. Finally, intersubject performance differences suggest that both cognitive and peripheral differences influence detection. That is, individual subjects may apply different detection strategies to the response of the auditory nerve in order to generate detection judgments.

This study is part of a series focused on psychophysical and physiological aspects of the coding of tones in noise (Evilsizer *et al.*, 2002; Zheng *et al.*, 2002). The tools of human psychophysics, animal behavior and, in ongoing studies, physiological recordings, are being applied to gain a better understanding of the cues that listeners use to detect tones in noise and the neural mechanisms that make the use of

[a)]Author to whom correspondence should be addressed.

these cues possible. The goal of this paper is to systematically measure and model tone-in-noise detection. Although our focus here was on "monaural" processing, monotic and diotic performance were directly compared and the findings were considered in the context of other relevant binaural results. We use reproducible noise data to evaluate the predictions of a variety of detection models, including the critical-band model. Modeling results under wideband and narrowband stimulus conditions were compared. These models were also compared to the data from eight subjects [four from the current study and four from the study of Evilsizer *et al.* (2002)].

A. Reproducible noise

We employed reproducible noise as a masking stimulus because it allowed for a more detailed comparison between subject responses and model predictions. In a classical tone-in-noise detection experiment, the masker waveform is generated, independently and *without replacement* on each trial, by a random or pseudorandom process, such that the same waveform is never presented twice. The performance of models and subjects is "averaged" across masker waveforms and then compared. Averaging the data in this way discards the information inherent in the trial-by-trial fluctuations in the statistics of the noise and in the subjects' responses to those fluctuations. In contrast, a tone-in-reproducible-noise detection experiment utilizes masker waveforms that are randomly selected on each trial *with replacement* from a small set (typically 10 to a few hundred) of noise waveforms. The performance of subjects and models can then be compared on a waveform-by-waveform basis. Note that the word "waveform" refers to any tone-plus-noise or noise-alone stimulus waveform. The phrase "masker waveform" refers only to the noise-alone waveform before the addition of a tone. Because subjects' responses show substantial and reliable variation across waveforms, the data from this type of reproducible-noise experiment provide a more demanding test for models. Indeed, models that make the same average (across waveforms) responses as the subjects can still fail to predict their responses to individual waveforms (e.g., Gilkey *et al.*, 1985).

Although different in experimental methodology, the reproducible-noise detection task in this study is analogous to more traditional tone-in-noise masking experiments in which the noise is created randomly and without replacement on each trial (e.g., Dolan, 1968). When the number of reproducible masker waveforms exceeds 10, subjects do not "learn" individual masker waveforms (Pfafflin, 1968). Here, 25 masker waveforms were used in each listening condition. Thus, as in studies where each masker waveform is created randomly and used without replacement, subjects in this study did not learn individual masker waveforms. In contrast, if only a single or a few masker waveforms are used (e.g., Langhans and Kohlrausch, 1992) subjects are most likely functioning differently because they have the ability to learn the individual waveforms.

B. Masker bandwidth

The critical-band model suggests that energy outside of the auditory filter centered at the tone frequency will not influence detection performance. However, a variety of evidence indicates that a more broadband process affects masked detection. Research on auditory physiology (e.g., Ruggero, 1973; Kiang and Moxon, 1974; Schalk and Sachs, 1980; and Costalupes *et al.*, 1984) has long suggested the presence of suppressive regions in the response of auditory-nerve fibers outside of the normal excitatory band. Similar effects have also been routinely observed in psychoacoustic data (e.g., Shannon, 1976). Other psychoacoustic findings suggest the presence of broader-bandwidth interactions (some extending more than 2 octaves above and below the tone frequency), both in cases when such interactions are advantageous [e.g., in profile analysis (Green, 1988) and comodulation masking release (Hall *et al.*, 1984)] and in cases when such interactions are disadvantageous (e.g., Neff and Callaghan, 1988). Reproducible-noise studies provide evidence for across-critical-band comparison in tone-in-noise detection (e.g., Ahumada and Lovell, 1971; Gilkey and Robinson, 1986) and for differences between wideband- and narrowband-masker conditions, at least under dichotic conditions (Evilsizer *et al.*, 2002).

C. N_0S_0 to N_mS_m comparison

Experimental results were directly compared under diotic (N_0S_0) and monaural (N_mS_m) conditions. Most models of binaural processing assume that detection patterns collected under N_mS_m and N_0S_0 conditions are "equivalent" (i.e., there is no masking-level difference between the N_mS_m and N_0S_0 configurations). However, despite the fact that comparable average performance is observed in these two conditions, it is possible that different strategies or different cell populations are used under each of the two conditions, and that detection statistics for individual waveforms may differ between the two conditions. Although reproducible noise has not previously been used to compare these conditions directly, such a comparison can reveal similarities and differences that may be obscured when the data are averaged across waveforms in a typical masking experiment. We directly compared the data for each subject under these two conditions.

D. Models of tone-in-noise detection

The major focus of this effort is to evaluate models of tone-in-noise detection. The set of models examined were broadly sampled from the range of ideas that have been applied to tone-in-noise detection and correspond to a diverse set of underlying physiological mechanisms. The specific models were selected to satisfy three criteria: they can be readily implemented and applied to reproducible-noise data, they have been successfully applied to tone-in-noise detection, and they have broad neurophysiological or psychoacoustic relevance. The critical-band model (Fletcher, 1940) is directly evaluated. A variation of the critical-band model, referred to as the multiple-detector model, is also evaluated. This model uses a linear combination of the energy at the

output of multiple critical-band filters as a decision variable (e.g., Ahumada and Lovell, 1971; Gilkey and Meyer, 1987). Energy-related models based on the average discharge rate of model auditory-nerve fibers (Heinz *et al.*, 2001b) were also considered.

Although energy is perhaps the most obvious decision statistic to compute from the narrowband output of an auditory filter, the addition of a tone to noise also changes the temporal properties of the filter output. Richards (1992), for example, has examined how the shape of the narrowband envelope differs between tone-plus-noise and noise-alone stimuli. We extend her analysis to the current conditions and refer to it as the envelope-slope model. Finally, we examine the predictions of a single-cell version of the phase-opponency model (Carney *et al.*, 2002), which uses coincidence detection of auditory-nerve discharges from fibers tuned to different frequencies to reveal similarities in discharge timing across these fibers associated with the presence of the tone. This model has been shown to describe significant features of the tone-in-noise masking data, including the minimal effects on subject performance of randomizing (or roving) overall stimulus level within a trial and across intervals for a 2-interval task (Carney *et al.*, 2002; Kidd, 1987; Kidd *et al.*, 1989). These effects cannot be captured by simple energy-based models, whose decision variables are based entirely on overall stimulus energy. The within-trial rove confounds the energy-based models, which select the interval with more energy as containing the tone, regardless of whether the tone was present, increasing thresholds by about 25% of the rove range (Green, 1984).

Overall, this study takes a broad view of data and preexisting models[1] for monotic and diotic tone-in-reproducible-noise masking data. The ability of each model to explain results for a reproducible-noise task as well as each model's ability to predict roving-level data are considered.

II. PSYCHOPHYSICAL EXPERIMENT

A. Methods

Experimental procedures were matched to those of Evilsizer *et al.* (2002) and Gilkey *et al.* (1985). N_0S_0 (diotic) and N_mS_m (monaural, left ear) interaural configurations were tested. All listening was completed in a double-walled sound attenuating booth (Acoustic Systems, Austin, TX). Four subjects participated in this study and ranged from 19 to 25 years in age. Each subject had audiometrically normal hearing. None of the subjects had prior experience with tone-in-noise detection experiments. Subjects occasionally commented on the various listening tasks, but were not solicited to do so. Subjects were debriefed as to the use of reproducible stimuli at the end of the final testing session.

1. Stimuli

Stimuli were generated and controlled by MATLAB software (Mathworks, Natick, MA). All stimuli were presented via a TDT System 3 (Tucker Davis Technologies, Gainesville, FL) RP2 programmable D/A converter and TDH-39 headphones (Telephonics Corp., Farmington, NY). Reproducible masker waveforms used by Evilsizer *et al.* (2002) were down-sampled from 50 kHz to 48.125 kHz with the MATLAB "resample" function to be compatible with a TDT System 3 sample rate. Each of the 25 wideband maskers was created from a broadband, Gaussian white noise by zeroing frequency components outside 100–3000 Hz. Narrowband maskers were created from the wideband maskers by zeroing frequency components outside 450–550 Hz. This process ensured that the spectral content of the narrowband maskers was identical to that of the wideband maskers within a range of 450–550 Hz. The masker spectrum level was 40 dB SPL. Tones and maskers were mixed in software and had 300 -ms durations including 10-ms cosine-squared on/off ramps. Tones were always added in the sine phase.

2. Training

The final testing procedure was a single-interval task involving large numbers of trials near threshold. In this study, we refer to threshold as the E_S/N_0 value (in dB) resulting in a $d'=1$. The modeling procedures assumed that the subjects were performing at the same level throughout these trials. Therefore, an extensive subject-training paradigm was employed to establish stable performance and decision criteria, and to minimize effects of learning. There were three separate training tasks. These tasks progressed in difficulty and similarity to the final testing procedure. The first was a two-interval, two-alternative forced-choice, tracking task with feedback. This was followed by a one-interval, fixed-level task with feedback and then by a one-interval, fixed-level task without feedback. For all tasks, subjects were given an unlimited amount of time to respond after the final observation interval. Random maskers were employed during all training tasks to prevent learning the individual reproducible noise waveforms.

In the first training procedure, two-down, one-up tracking with feedback was used to quickly estimate a tone level where $d'=0.77$ (the 70.7% correct point on the psychometric function as described by Levitt, 1971). Subjects were asked to use a computer mouse to click one of two large buttons corresponding to "Interval One" or "Interval Two." The buttons were presented following the second observation interval on an external computer screen visible through the window of the sound-attenuating booth. Immediately following each response, the word "correct" or "incorrect" was displayed on the computer screen for 700 ms. Each track had a duration of 100 trials. Each trial contained a pair of randomly generated maskers that were frozen across intervals. Intervals were separated by 500 ms of silence. The step size used in the adaptive track was 4 dB for the first two reversals and 2 dB thereafter. Threshold estimates were calculated by averaging the tone levels at reversals in the track, excluding the first 4 or 5 reversals such that the number of remaining reversals was even. Subjects completed 10–15 runs for each configuration and bandwidth before moving on to single-interval tasks.

In the second training procedure, a one-interval fixed-level task with feedback was used to familiarize subjects with a one-interval task and ensure thresholds were stable. Subjects were asked to click on one of two large buttons

corresponding to "Tone" or "No Tone." Feedback was presented by displaying either "Correct" or "Incorrect" on the monitor for 700 ms after the response button was clicked. Each block had 100 trials with tone levels +3, +1 or −1 dB with respect to the tone level determined by the two-interval, two-alternative forced-choice task. Two 100-trial blocks were completed for each of the three tone levels for each interaural configuration and bandwidth. If a subject's threshold changed, this sequence was repeated, after adjusting the tone level in 1-dB steps, until a tone level was determined that resulted in a stable d' approximately equal to unity for each bandwidth and stimulus configuration.

After a stable tone level for each condition was determined, subjects completed the third training procedure, a one-interval training task without feedback using random noise. This task was used to determine whether d' could be expected to remain near unity when trial-by-trial feedback was eliminated during the testing phase. In the rare case that the value of d' changed such that it was no longer near unity, the level was adjusted with 1-dB resolution until d' returned to near unity.

3. Testing

The testing procedure was identical to the final training procedure except that reproducible maskers were used during testing. When present, the tone was always at the level determined during training without feedback. Final analyses were performed on complete data sets at a single tone level for each subject, with a d' near unity for the duration of the testing period.

Each testing set consisted of four blocks of 100 trials. Before each set, 20 practice trials with feedback were presented using a tone level 2 dB above the testing level. These practice trials were included to help the subjects maintain an effective and consistent detection strategy during the course of the experiment, given that feedback was *never* presented while using reproducible maskers. Random maskers were used during the practice trials to prevent learning the reproducible stimuli. Using the level determined during the training procedure, 12 sets were run without feedback for each bandwidth and configuration. During testing, 25 reproducible masker waveforms were used. Each masker waveform was presented twice with a tone and twice without a tone in each block. Overall, 96 presentations of each tone-plus-noise and 96 presentations of each noise-alone stimulus were made (a total of 48 blocks) in each of the four testing conditions (narrowband and wideband; monotic and diotic).

Bias (β, MacMillan and Creelman, 1991) was computed across all waveforms in a particular listening condition. If β departed more than 15% from 1.0 (i.e., equally likely to respond "tone" and "no tone") on a given run, subjects were instructed to "try to make an equal number of 'Tone' and 'No Tone' responses." All trials were included in the study regardless of the resulting bias measure. No attempt was made to account for or correct data with bias deviating from unity, as this was not the primary focus of the study.

Hit rates, or the conditional probabilities of responding yes given a particular tone-plus-noise waveform [$P(Y|T+N)$], and false-alarm rates, or the conditional probabilities of responding yes given a particular noise-alone waveform [$P(Y|N)$], were computed in each bandwidth and stimulus configuration for individual tone-plus-noise and noise-alone waveforms, respectively. A third set of probabilities of responding yes given a particular stimulus waveform [$P(Y|W)$] was also created for each listening condition. That is, $P(Y|W)$ includes the 25 $P(Y|Y+N)$ and the 25 $P(Y|N)$ values. The set of 50 $P(Y|W)$ values is referred to as a detection pattern (see Fig. 1). Each value of $P(Y|W)$ represents the probability of responding "yes" or "tone present" to an individual waveform (either tone-plus-noise or noise-alone) and can be thought of as the likelihood of the tone being perceived in that particular waveform.

Comparisons between detection patterns were quantified in terms of the square of the correlation coefficient, or the coefficient of determination (r^2). Statistically significant ($p < 0.05$) r^2 values occurred above the critical value of $r^2 = 0.08$ for comparisons involving the set of 50 $P(Y|W)$ values, or above the critical value of $r^2 = 0.17$ for comparisons involving either the set of 25 $P(Y|T+N)$ or the set of 25 $P(Y|N)$ values. These critical values were established using a two-tailed t-test (Bruning and Kintz, 1968). The coefficient of determination allows for comparisons between experimental results and the regression analysis presented in the modeling section of this paper.

B. Results and discussion

Subject performance was characterized on two levels. Overall performance for the ensemble of waveforms was characterized by averaging across noise-alone waveforms and across tone-plus-noise waveforms, as well as across repeated presentations of the waveforms. These ensemble-level results provide traditional measures of performance (e.g., d' and β) that represent the subjects ability to perform the tasks, show consistent responding among subjects, and allow direct comparison to other studies. In addition to these traditional ensemble-level results, detection patterns were computed for each subject under each condition by averaging across repeated presentations of the individual waveforms, but not across waveforms. Note that an "average subject" was created by averaging the $P(Y|W)$ values across the four subjects in this study. Detection patterns of the individual and average subjects were then used for the following empirical comparisons: $P(Y|T+N)$ and $P(Y|N)$ values were compared between N_mS_m and N_0S_0 stimuli to test the hypothesis that detection patterns were the same between stimulus configurations, for both narrowband and wideband stimuli. $P(Y|T+N)$ and $P(Y|N)$ values were compared between narrowband and wideband stimuli to test the hypothesis that stimulus information outside a critical-band influences detection patterns in both N_mS_m and N_0S_0 conditions. Finally, detection patterns were compared across subjects to reveal the possible uses of different/similar strategies within each stimulus configuration and bandwidth.

1. Reliability of the data

The ensemble-level results are shown in the left side of Table I. Data are presented for each subject and the average

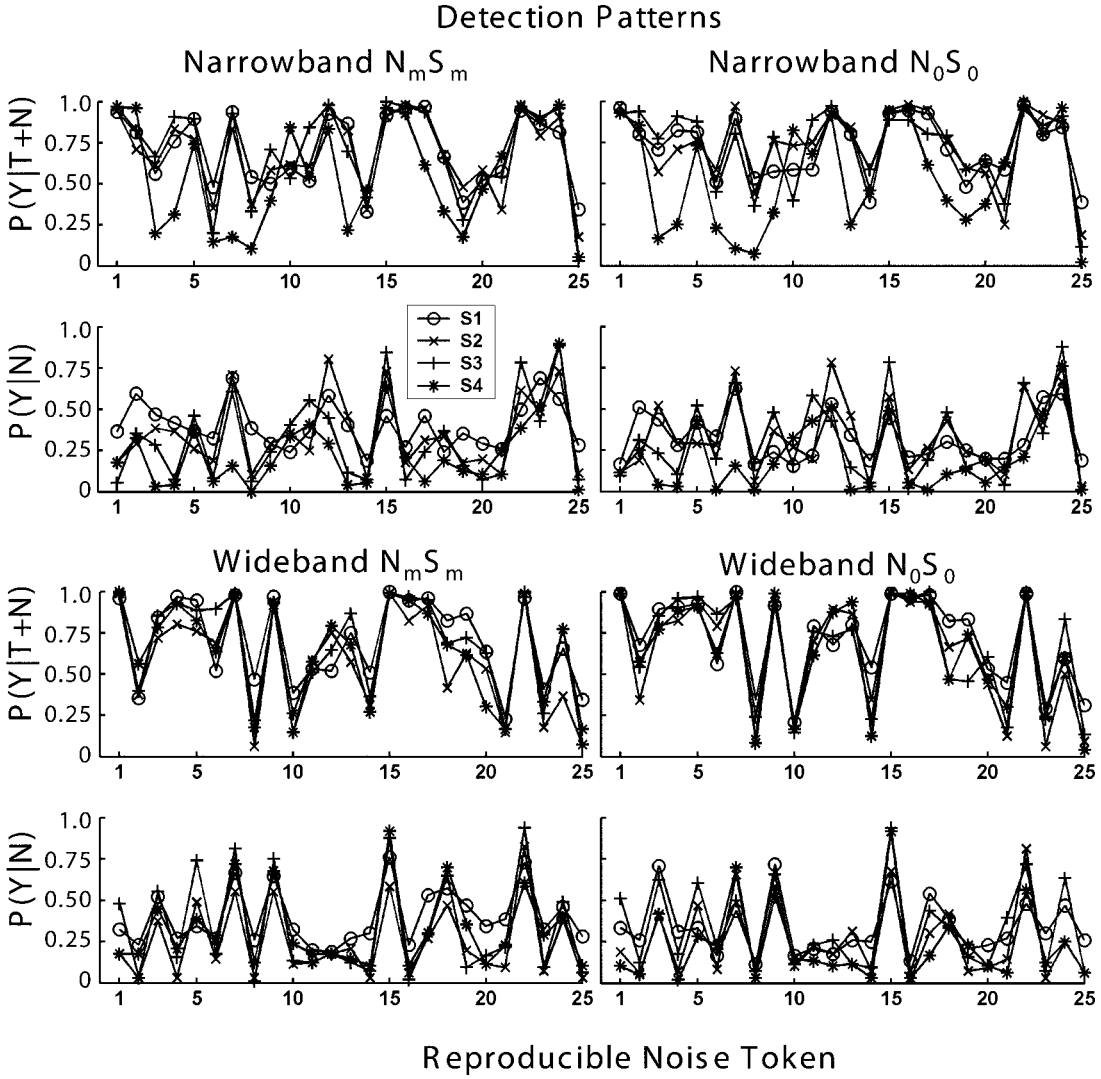

FIG. 1. Detection performance across reproducible maskers for the monaural and diotic stimulus conditions for narrowband maskers (top) and wideband maskers (bottom). Note that the horizontal axis is not a continuous variable; lines connect symbols corresponding to individual subjects to facilitate intersubject comparisons.

subject within each stimulus configuration and bandwidth (NB: narrowband, 450–550 Hz; WB: wideband 100–3000 Hz). Stimulus E_S/N_0 levels were within 3 dB for all subjects in all conditions tested. All d' values were within 20% of unity, thus subjects were tested at threshold. Bias results indicated that S4 had a tendency to report "tone absent" more often than "tone present" in all but the wideband N_mS_m stimulus configuration. *S*2 showed similar bias for the two wideband conditions. Other subjects showed little or no bias. Overall, subjects' performance levels were similar and near unity d' and bias.

The r^2 values between detection patterns constructed from the first and last 48 presentations of each stimulus and across the 25 individual tone-plus-noise and noise-alone waveforms in each bandwidth and stimulus condition were computed using $P(Y|T+N)$, $P(Y|N)$, and $P(Y|W)$ values (Table I). These first-half, last-half r^2 values indicate within-subject response consistency and serve as a reference for intersubject comparisons, as well as a reference for between-bandwidth and between-stimulus configuration comparisons. Subjects tended to respond more consistently on trials with the tone present than on trials with the tone absent. However, all coefficients were significant, and 52 of the 60 values were above 0.80, indicating that subjects were stable throughout the experiment.

2. Differences across waveforms

Results for each subject are presented in Table I and in Fig. 1 for each stimulus configuration. Figure 1 shows both monotic (left panels) and diotic (right panels) detection patterns. The hit-rates [$P(Y|T+N)$, upper panels] are based on trials containing the tone, and the false-alarm rates [$P(Y|N)$, lower panels] are based on trials without the tone.

Consistent with the results of other reproducible-masker studies, subjects' hit rates and false-alarm rates depended on the individual masker waveform (Evilsizer *et al.*, 2002; Isabelle and Colburn, 1991; Siegel and Colburn, 1989). The results of a χ^2 analysis with 24 degrees of freedom, using the procedure developed by Siegel and Colburn (1989), are shown in Table I. The χ^2 statistic was used to test the hypotheses that variation in $P(Y|T+N)$ and $P(Y|N)$ across

TABLE I. Psychophysical results across waveforms. E_S/N_0 is the ratio of tone energy to noise spectrum level (in dB). E_S/N_0, d', and β are shown for each combination of subject, interaural configuration, and noise bandwidth. NB and WB correspond to masker bandwidth: 100 Hz and 2900 Hz, respectively. The χ^2-values are given for performance across reproducible waveforms for both $P(Y|T+N)$ and $P(Y|N)$. The number of presentations of each masker for both $P(Y|T+N)$ and $P(Y|N)$ is given by n. The coefficient of determination, r^2, for the first half of trials vs the last half of trials is presented for both $P(Y|T+N)$ and $P(Y|N)$, individually, as well as across the entire ensemble of waveforms $[P(Y|W)]$. All χ^2 and r^2 values are significant to $p<0.001$.

						$P(Y\|T+N)$			$P(Y\|N)$			$P(Y\|W)$
Interaural configuration	BW	S	E_S/N_0	d'	β	χ^2	n	r^2	χ^2	n	r^2	r^2
N_mS_m	NB	S1	11.8	0.80	0.90	510.9	96	0.94	184.6	96	0.49	0.86
		S2	10.8	0.92	0.99	597.2	96	0.77	475.7	96	0.72	0.84
		S3	10.8	1.06	0.99	866.0	96	0.94	718.9	96	0.81	0.90
		S4	10.8	0.92	1.45	1003.4	96	0.92	602.2	96	0.86	0.91
		$S_{avg(4)}$	11.1	0.93	1.08	2379.6	384	0.96	1391.3	384	0.91	0.96
	WB	S1	11.8	0.80	0.91	699.5	96	0.92	278.5	96	0.72	0.88
		S2	10.8	0.96	1.40	926.3	96	0.94	599.3	96	0.83	0.92
		S3	10.8	0.90	1.00	1017.1	96	0.96	952.5	96	0.90	0.95
		S4	9.8	0.86	1.06	856.1	96	0.84	543.0	96	0.81	0.88
		$S_{avg(4)}$	10.8	0.88	1.09	3240.9	384	0.98	2057.7	384	0.96	0.98
N_0S_0	NB	S1	12.8	1.08	0.94	400.5	96	0.79	233.3	96	0.64	0.87
		S2	10.8	1.09	0.94	593.5	96	0.79	498.6	96	0.72	0.86
		S3	10.8	1.12	0.96	608.4	96	0.86	666.4	96	0.84	0.92
		S4	10.8	1.01	1.58	984.7	96	0.94	610.0	96	0.92	0.95
		$S_{avg(4)}$	11.3	1.07	1.11	1907.3	384	0.94	1427.3	384	0.87	0.95
	WB	S1	11.8	1.03	0.94	728.9	96	0.88	300.0	96	0.81	0.92
		S2	10.8	1.08	1.25	1138.0	96	0.92	647.8	96	0.88	0.94
		S3	9.8	0.91	1.03	1066.6	96	0.94	671.9	96	0.88	0.94
		S4	10.8	1.17	1.37	1079.8	96	0.92	686.6	96	0.86	0.93
		$S_{avg(4)}$	10.8	1.04	1.15	3776.1	384	0.97	1903.0	384	0.98	0.98

maskers was not due to chance (i.e., that the detection patterns were waveform dependent). All χ^2 statistics greatly exceeded the $p<0.001$ significance level ($\chi^2_{crit}=51.18$), and thus the null hypotheses of only random variations in detection patterns were rejected. Note that a larger χ^2 statistic indicates a more reliable detection pattern.

3. Comparisons between stimulus configurations

This experiment tested the hypothesis that psychophysical detection patterns are the same for the N_mS_m and N_0S_0 stimulus configurations. Detection patterns were significantly ($p<0.001$) correlated between the N_mS_m and N_0S_0 stimulus configurations for both narrowband and wideband maskers (see Table II). In Fig. 1, the similarity between N_mS_m and N_0S_0 detection patterns for individual subjects is also visible.

In this study, tone levels producing unity d' for individual subjects were within about 1 dB for the N_mS_m and N_0S_0 conditions. These results were consistent with the failure to find $N_mS_m-N_0S_0$ MLDs in several other experiments (Sever and Small, 1979; Egan *et al.*, 1969; Egan, 1965; Hirsh and Burgeat, 1958; Blodgett *et al.*, 1958) that used free-running noise and similar masker levels as in this experiment. However, at lower masker levels, some have observed an $N_mS_m-N_0S_0$ MLD.[2] In a study that used a *single* frozen noise masker, Langhans and Kohlrausch (1992) compared monaural and diotic thresholds and found an $N_mS_m-N_0S_0$ MLD. It is not surprising that the results in these two studies are different, since subjects could use the learned properties of the single noise waveform in their study. As previously explained, our study was designed to provide results that are comparable to performance in traditional experiments with random noise.

TABLE II. Correlations between detection patterns for N_mS_m and N_0S_0 stimulus configurations. Narrowband and wideband r^2 values are presented for each subject. These r^2 values were calculated from responses across the 25 tone-plus-noise waveforms $[r^2_{P(Y|T+N)}]$, 25 noise-alone waveforms $[r^2_{P(Y|N)}]$, as well as across the ensemble of all 25 tone-plus-noise and 25 noise-alone waveforms $[r^2_{P(Y|W)}]$.[a]

Bandwidth	Subject	$r^2_{P(T+N)}$	$r^2_{P(Y\|N)}$	$r^2_{P(Y\|W)}$
NB	S1	0.94	0.70	0.90
	S2	0.86	0.88	0.92
	S3	0.81	0.86	0.90
	S4	0.96	0.86	0.95
	$S_{avg(4)}$	0.89	0.90	0.91
WB	S1	0.79	0.64	0.83
	S2	0.92	0.92	0.95
	S3	0.90	0.82	0.89
	S4	0.88	0.90	0.92
	$S_{avg(4)}$	0.95	0.91	0.95

[a]All values $p<0.001$.

TABLE III. Correlations between detection patterns for narrow (100 Hz) and wide (2900 Hz) stimulus bandwidths. Individual-subject r^2 values are presented within either the N_mS_m or N_0S_0 stimulus configuration and were calculated from responses across the 25 tone-plus-noise waveforms [$r^2_{P(Y|T+N)}$], 25 noise-alone waveforms [$r^2_{P(Y|N)}$], as well as across the ensemble all 25 tone-plus-noise and 25 noise-alone waveforms [$r^2_{P(Y|W)}$].

Stimulus configuration	Subject	$r^2_{P(Y\|T+N)}$	$r^2_{P(Y\|T)}$	$r^2_{P(Y\|W)}$
N_mS_m	S1	0.21[a]	0.03	0.37[b]
	S2	0.41[b]	0.31[b]	0.54[b]
	S3	0.29[a]	0.38[b]	0.49[b]
	S4	0.07	0.11	0.23[b]
	$S_{avg(4)}$	0.28[a]	0.29[a]	0.48[b]
N_0S_0	S1	0.36[b]	0.09	0.56[b]
	S2	0.41[b]	0.41[b]	0.59[b]
	S3	0.50[b]	0.46[b]	0.61[b]
	S4	0.06	0.07	0.24[b]
	$S_{avg(4)}$	044[b]	0.42[b]	0.64[b]

[a] $p<0.05$.
[b] $p<0.01$.

4. Comparisons between bandwidths

Between-bandwidth comparisons were performed to determine whether stimulus information outside the critical-band centered at the tone frequency influenced detection patterns. Recall that narrowband maskers were created by removing all components outside the band centered at 500 Hz; thus, narrowband and wideband stimuli should be identical within the 500-Hz critical band. Between-bandwidth r^2 values are presented in Table III and the underlying correlations are also visible in the detection patterns in Fig. 1. Although between-bandwidth r^2 values were significant for most subjects in both stimulus configurations (recall that for $n=25$, $r^2>0.17$ is significant, $p<0.05$), values were significantly lower than those observed between the first 48 and last 48 presentations of each masker (first-half, last-half r^2, Table I). A test for significant difference between two nonindependent correlations (Bruning and Kintz, 1968) was performed between the first-half, last-half correlations and cross-bandwidth correlations to determine if the two sets of correlations differed significantly. Results showed significant differences ($p<0.01$) for 37 of the 64 tests (4 subjects $\times 2$ bandwidths $\times 2$ stimulus configurations $\times 2$ tone cases, with and without tone $\times 2$ halves). Of the 27 that failed at the $p<0.01$ level, 15 involved statistically insignificant ($p<0.05$) cross-bandwidth correlations of 0.34 or less. Given that the between-bandwidth variation in detection patterns was statistically larger than the within-bandwidth variation in detection patterns (between the first half and the last half of trials recorded), these results indicate that detection performance was influenced by stimulus information outside one critical-band for both monotic and diotic conditions. It is notable that the noise added outside the critical band did not affect threshold (i.e., E/N_0 where d' was near unity), but it did affect detection patterns.

Narrowband maskers had the same magnitude and phase as wideband maskers for components in the 100-Hz region centered on the stimulus frequency. Consequently, stimulus information in the wideband masker falling inside one critical-band was approximately (ignoring frequency components present in the auditory filter skirts) the same as that of the narrowband masker, while information falling in adjacent critical-bands was unique to the wideband masker.

The effect of the noise added outside the critical band presents a challenge for models based on energy in a single critical band. These models cannot correctly predict changes in detection patterns unless energy differences in the filter skirts are responsible for the differences between the two conditions. Another hypothesis is that listeners use different detection strategies for different noise bandwidths, causing reduced correlations (Evilsizer *et al.*, 2002). In this case, a model based on a single critical band may still be used, but such models would require different decision variables for the different bandwidths.

5. Comparisons among subjects

Comparisons were performed between detection patterns, $P(Y|W)$, for each subject-subject pair to reveal intersubject consistency and the possible uses of different/similar strategies within each stimulus configuration and bandwidth. The means and ranges of intersubject r^2-values are provided in each experimental condition in Table IV.

All possible intersubject correlations were computed and were found to be significant ($p<0.05$) for all monaural stimuli. Within the monaural stimulus condition, higher intersubject correlations were observed for wideband stimuli relative to narrowband stimuli. S4 was the least correlated to other subjects in the monaural stimulus condition (see Fig. 1) and also had the lowest correlation across stimulus bandwidths.

Intersubject correlations were lower for narrowband conditions relative to wideband conditions in both the N_mS_m

TABLE IV. Mean, minimum, and maximum intersubject r^2 values [$P(Y|W)$, calculated from responses across the ensemble of all 25 tone-plus-noise and 25 noise-alone waveforms] for the subjects in the current study and the subjects from Evilsizer *et al.* (2002).

Interaural configuration	Bandwidth	mean[$r^2_{[P(Y\|W)]}$]	min[$r^2_{[P(Y\|W)]}$]	max[$r^2_{[P(Y\|W)]}$]
N_mS_m	NB	0.67	0.48	0.84
	WB	0.85	0.81	0.89
N_0S_0	NB	0.62	0.31	0.81
	WB	0.72	0.51	0.91

TABLE V. Proportions of predictable variance estimated by four different methods (see text for details) for N_mS_m and N_0S_0 stimulus configurations and NB (100 Hz) and WB (2900 Hz) conditions.

Interaural configuration	BW	Statistical method			
		V_p	$\sigma_p^2/\sigma_{\text{Tot}}^2$	$\sigma_{\text{ext}}^2/\sigma_{\text{Tot}}^2$	$\sigma_{\text{Waveform}}^2/\sigma_{\text{Tot}}^2$
N_mS_m	NB	0.990	0.994	0.996	0.939
	WB	0.995	0.995	0.998	0.975
N_0S_0	NB	0.988	0.997	0.997	0.941
	WB	0.996	0.998	0.998	0.977

and N_0S_0 listening configurations in this study and also in both the N_0S_0 and N_0S_π listening configurations in Evilsizer *et al.* (2002). Subjects in the current study reported more difficulty with the narrowband task, which may have led to increased variability in detection strategies and reduced intersubject correlations for narrowband conditions. Diotic intersubject correlations were computed for the four subjects in this work (S1–S4), as well as the four subjects in the Evilsizer *et al.* (2002) study (S5–S8). No noticeable decrease in intersubject correlation occurred when comparing across studies (i.e., when comparing between S1–S4 and S5–S8), although S4 and S7 did show the lowest intersubject correlations [0.67 and 0.52 for 100-Hz N_0S_0 $P(Y|T+N)$ and $P(Y|N)$, respectively]. As in the monotic case, diotic trials using wideband maskers had higher intersubject correlations than those using narrowband maskers. However, low intersubject correlations were rare, and all but 6 of the 112 diotic intersubject correlations were significant at the $p<0.05$ level. Overall, these results indicate that detection patterns were consistent across subjects, except for S4 and S7.

6. Internal noise

In previous studies of detection using reproducible noise, investigators have taken advantage of the fact that the amount of variability in the responses across a set of waveforms is available (i.e., the detection pattern) to estimate the relative amount of internal variability. The ratio of internal to external noise would be large if the external noise sample had a small effect on performance and small if performance was highly variable across noise waveforms. This measure is also useful for comparison to other psychophysical results. To estimate the internal-to-external noise ratio, the internal variance (σ_{int}^2) is assumed to be equal to 1 for all masker waveforms, and the external variance (σ_{ext}^2) is estimated from the sample variance of the z scores across the set of waveforms; the ratio ($\sigma_{\text{int}}^2/\sigma_{\text{ext}}^2$) is then the reciprocal of the square root of the external variance (Siegel and Colburn, 1989). Internal-to-external noise ratios (in standard deviations) ranged from approximately 1.0 to 4.0 across the four subjects in our study[3] (similar to values reported in Spiegel and Green, 1981; Siegel and Colburn, 1989; and Isabelle and Colburn, 1991). Monotic and diotic results were similar for each subject, indicating that the amount of internal noise did not substantially change between the monotic and diotic tasks.

7. Predictable variance

In Sec. III, various models are used to predict the changes in $P(Y|W)$ observed across the combined set of 25 noise-alone and 25 tone-plus-noise waveforms. As discussed in the previous sections, the data are quite reliable, suggesting that the models have the potential to predict a substantial portion of the variability in $P(Y|W)$. Table V presents four estimates of the portion of predictable variance for the average subject under each of the four stimulus conditions.

The first column shows estimates of the proportion of predictable variance based on the first-half, last-half correlations for the results described in Table I. These estimates were obtained using the equation presented by Ahumada and Lovell (1971),

$$V_p = \frac{r_{12}}{r_{12} + 0.5(1 - r_{12})}, \tag{1}$$

where V_p is the proportion of predictable variance, and r_{12} is the first-half, last-half correlation. [Note because the first-half, last-half correlations from the Evilsizer *et al.* (2002) experiment were not readily available; these estimates are based on only the four subjects in this experiment, even under the N_0S_0 condition.]

The second column of Table V presents estimates of the proportion of predictable variance calculated from the assumption that the only source of error is that associated with estimating a proportion based on Bernoulli trials, such that

$$\frac{\sigma_p^2}{\sigma_{\text{Tot}}^2} = 1 - \left[\frac{E\left[\dfrac{P(Y|W)[1 - P(Y|W)]}{n_{\text{trials}}} \right]}{\sigma_{\text{Tot}}^2} \right], \tag{2}$$

where σ_p^2 is the predictable variance, σ_{Tot}^2 is the total variance in the $P(Y|W)$ values for the average subject, $P(Y|W)$ is the probability of a "tone present" response for each tone-plus-noise and each noise-alone waveform, n_{trials} is the number of trials used to estimate each $P(Y|W)$, and the expected value is across the combined set of 50 waveforms (25 tone-plus-noise and 25 noise-alone waveforms).

The third column presents estimates of the proportion of predictable variance based on the internal-to-external noise ratio ($\sigma_{\text{int}}^2/\sigma_{\text{ext}}^2$) discussed in Sec. II B 6. Assuming that the internal variance estimate is reduced by the number of trials,

$$\frac{\sigma^2_{\text{ext}}}{\sigma^2_{\text{Tot}}} = \frac{\sigma^2_{\text{ext}}}{\left[\sigma^2_{\text{ext}} + \frac{\sigma^2_{\text{int}}}{n_{\text{trials}}}\right]} = \frac{1}{\left[1 + \frac{\left(\frac{\sigma^2_{\text{int}}}{\sigma^2_{\text{ext}}}\right)}{n_{\text{trials}}}\right]}, \quad (3)$$

where the internal-to-external noise ratio ($\sigma^2_{\text{int}}/\sigma^2_{\text{ext}}$) is the reciprocal of the variance in the z scores of $P(Y|W)$ across the waveforms (i.e., predictable variance), σ^2_{Tot} is the estimated total variance across the z scores of $P(Y|W)$ for each waveform for the average subject, and n_{trials} is the number of trials used to estimate each z score of $P(Y|W)$.

The fourth column presents estimates of the proportion of predictable variance based on partitioning the variance from separate Waveform × Subject Analyses of Variance computed for each condition. A randomized-blocks design was used, such that each cell had a single value, $P(Y|W)$ for that waveform and that subject. Adapting equations from Kirk (1995, pp. 258 and 267) for the Random-Effects Model, we have

$$\text{MS}_{\text{Waveform}} = n_{\text{subject}} \cdot \sigma^2_{\text{Tot}} = \sigma^2_{\text{error}} + \sigma^2_{(\text{Waveform}\times\text{Subject})} + n_{\text{subject}} \cdot \sigma^2_{\text{Waveform}}$$

and

$$\text{MS}_{\text{Residual}} = \sigma^2_{\text{error}} + \sigma^2_{(\text{Waveform}\times\text{Subject})}.$$

Therefore,

$$\frac{\sigma^2_{\text{Waveform}}}{\sigma^2_{\text{Tot}}} = \frac{\text{MS}_{\text{Waveform}} - \text{MS}_{\text{Residual}}}{\text{MS}_{\text{Waveform}}}, \quad (4)$$

where $\sigma^2_{\text{Waveform}}$ is the variance associated with the effect of waveform (i.e., predictable variance), $\text{MS}_{\text{Waveform}}$ is the mean square for the effect of Waveform, $\text{MS}_{\text{Residual}}$ is the mean square for the residual error based on the Subject by Waveform interaction, n_{subject} is the number of subjects, and σ^2_{Tot} is the estimated total variance in the $P(Y|W)$ values for the average subject.

A similar analysis can be applied to estimate the proportion of predictable variance in the responses of individual subjects as opposed to the proportion of predictable variance in the average subject, as presented here. The results of such an analysis are reported in the Appendix.

All four estimates suggest that over 93% of the variance is potentially predictable in all four conditions, with the estimates of the proportion of predictable variance ranging from 0.939 to 0.998 across methods and conditions.

III. MODELS FOR MASKED DETECTION

A. General approach

Two general categories of models were considered in this study. The general structure of each model is illustrated in Fig. 2: six were based on energy or average firing rates: [CB: critical band; MD: multiple detector; and four AN-count-based models (L: low spontaneous-rate; H: high spontaneous rate; L_{SS}: steady-state, low spontaneous-rate; and H_{SS}: steady-state, high spontaneouse-rate)], and two were

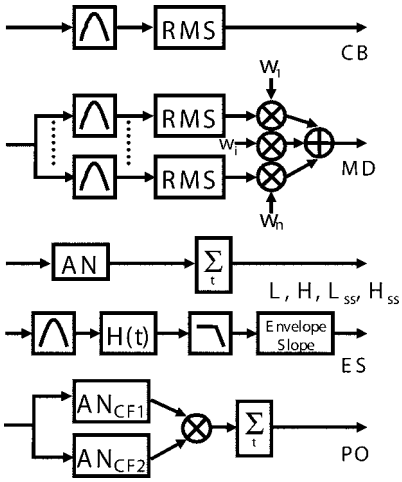

FIG. 2. Summary of model structures. The model inputs were pressure waveforms as a function of time. Each model's output was a decision variable corresponding to a given input waveform. Models, listed top to bottom are: CB, critical band; MD, multiple detector; low (L) and high (H) spontaneous-rate and steady-state versions of low (L_{SS}) and high (H_{SS}) spontaneous-rate auditory-nerve fibers; envelope slope (ES); and phase opponency (PO). $H(t)$ denotes the absolute value of the complex analytic function.

based on temporal variation: [an envelope-based model (ES: envelope slope), and the phase-opponency model (PO)]. The first category combines energy-based models with models based on neural counts of AN fibers, which are a first approximation of a physiologically-based energy estimate. The temporal models considered here were based on either the envelopes of narrowband-filtered stimuli (ES) or on fine-structure, as assessed using a physiological coincidence-detection-based model for detection (PO). The specific function and structure of each model will be described in detail in the following sections. All model simulations used a 50 -kHz sampling rate.

These models used either stimulus-based or physiologically-based decision variables that were computed from specific noise-alone and tone-plus-noise waveforms. Model outputs were compared to the individual data of each subject and also to the average subject [S_{avg}; including subjects from the Evilsizer *et al.* (2002) study for the N_0S_0 condition]. Although the models tested here were monaural, modeling results were compared to both monotic and diotic psychophysical data because subjects' detection patterns were highly correlated across the N_mS_m and N_0S_0 stimulus configurations. Model predictions were quantified using the following logistical regression procedure (Gilkey and Robinson, 1986). Model decision variables were fit to hit and false-alarm rates for individual subjects as well as to those for the average subject. An ogive function with two parameters was used for each case,

$$P'(Y|W) = \frac{1}{1 + e^{[[-(\log_{10}(x)-\mu)]/\theta]}}, \quad (5)$$

where $P'(Y|W)$ is the probability of a model "yes" response, x is the model decision variable, μ is the threshold parameter, and θ is the slope parameter (Gilkey and Robinson, 1986; MacMillan and Creelman, 1991). A Nelder-Mead sim-

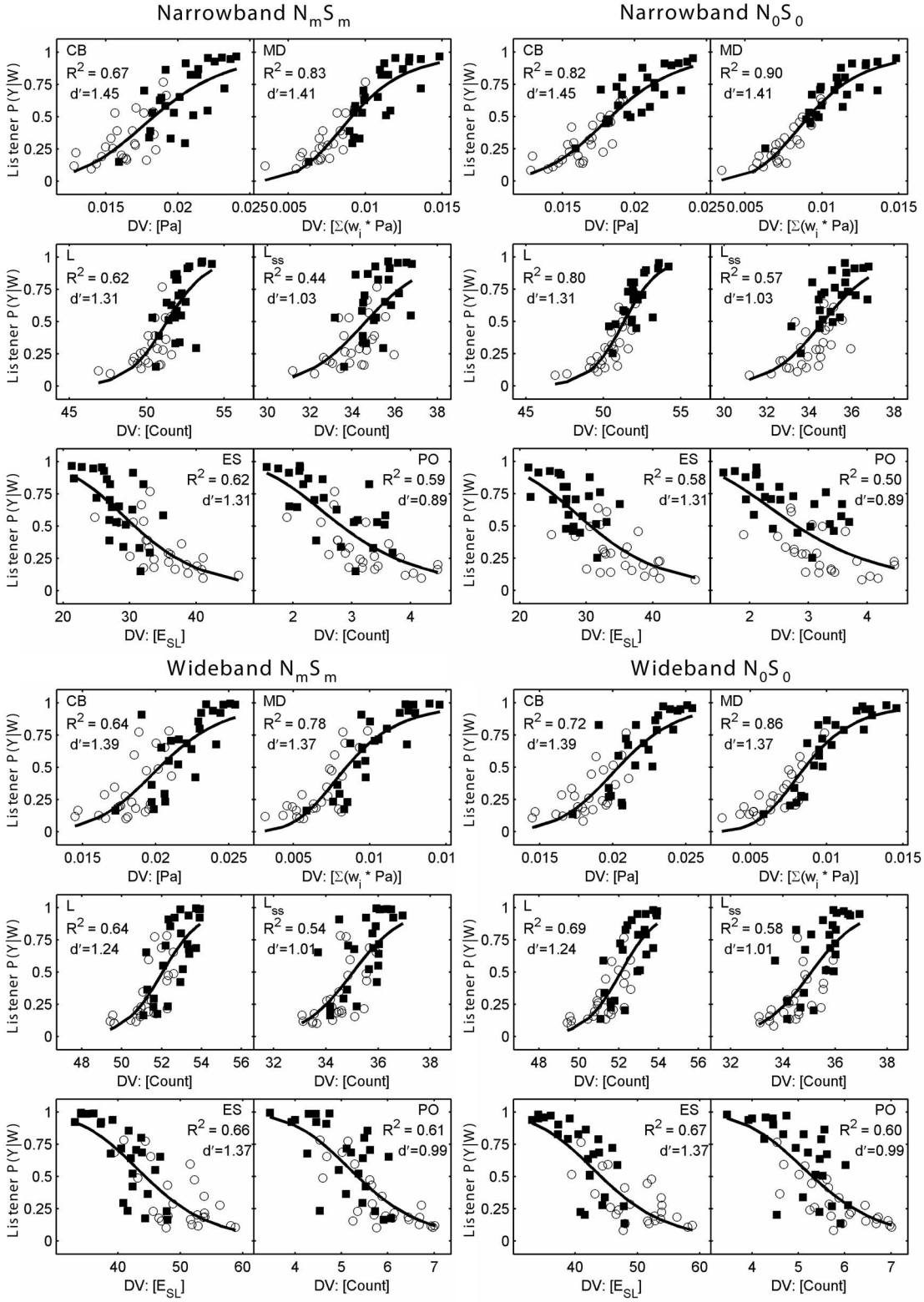

FIG. 3. Model predictions. Logistical regression was used to quantify the ability of the models in Fig. 2 to predict the average subject's $P(Y|W)$ values. Each panel shows the resulting ogive, and a scatter plot of model decision variables vs the average subject's $P(Y|W)$ values, subdivided into $P(Y|T+N)$ (squares, the probability of responding yes to a tone-plus-noise waveform) and $P(Y|N)$ (circles, the probability of responding yes to a noise-alone waveform). R^2 values and model d's are also shown. A perfect model would be indicated by all points in the scatter plot falling along the ogive and $R^2=1$. Model abbreviations are as in Fig. 2. Note that only the low-spontaneous-rate AN models are shown because of the relatively poor performance of high-spontaneous-rate AN models (see text).

plex direct search (MATLAB fminsearch) was used to minimize the sum of squared deviations between the set of subject $P(Y|W)$ values and the set of model $P'(Y|W)$ values. The search produced threshold and slope parameters corresponding to the best model fit for each individual subject and for the average subject. Model fits to the average subject are shown in Fig. 3. Note that although the plots in Fig. 3 explicitly show tone-plus-noise and noise-alone data, models were fit to the combined tone-plus-noise and noise alone data, that is, all 50 $P(Y|X)$ values. The resulting threshold and slope parameters were used in Eq. (5) to transform the model decision variable corresponding to each tone-plus-noise waveform into a model $P(Y|T+N)$, and the model decision variable corresponding to each noise-alone waveform into a model $P(Y|N)$, yielding a set of model hit and false-alarm rates. The following R^2 statistic was used to quantify the ability of each model to predict the data. The proportion of variance accounted for by the fit is defined as

$$R^2 = \frac{\sum_{i=1}^{50} [P'(Y|W_i) - P(Y|W_i)]^2}{\sum_{i=1}^{50} [P(Y|W_i) - \overline{P(Y|W)}]^2}, \quad (6)$$

where $P(Y|W)$ is the subject's probability of a "yes" response and $P'(Y|W)$ is the model's probability of a "yes" response. The numerator of Eq. (6) is the sum of squared errors between individual model and subject probabilities across waveforms, where i indicates the waveform index. The denominator is the sum of squared deviations between subject probabilities corresponding to individual waveforms and the mean subject probability across waveforms. The symbol R^2 is used to indicate the nonlinear fit between the model decision variable and the subject $P(Y|W)$ values, but is otherwise analogous to the r^2 values between sets of $P(Y|W)$ values reported in Sec. II. The use of model parameters in addition to the slope and threshold based on Eq. (5) is specifically reported in each model description below.

Each model was run at an E_S/N_0 of 10.8 dB, the median signal-to-noise ratio of all listeners combined across all conditions. Model d' values were computed as a relative sensitivity measure from the raw decision variables (i.e., before any fitting was performed). Thus, a single d' was computed for each model in each bandwidth as the difference in the means of the model decision variable between tone-plus-noise and noise-alone waveforms, divided by the square root of the average variance of model decision variables for tone-plus-noise and noise-alone waveforms. Individual model d' values were expected to be identical under the N_mS_m and N_0S_0 listening conditions because the models were monaural and d' values were computed from decision variables based on the same stimulus waveforms (before any fitting procedures were completed).

None of the models in this study included internal noise. Accordingly, if we assume the subjects' internal-to-external noise ratios are equal to 1 (internal-to-external noise ratios for the average subject ranged from 1.0 to 1.3), and that the model sensitivity equals the average listeners' sensitivity, then we expect model d' values to be larger than those of the subjects by a factor of $\sqrt{2}$ (i.e., since the model contained no internal noise, and the models and subjects experienced the same external noise, the overall model noise was half that of the listener, and thus the model d' was larger by a factor of $\sqrt{2}$).

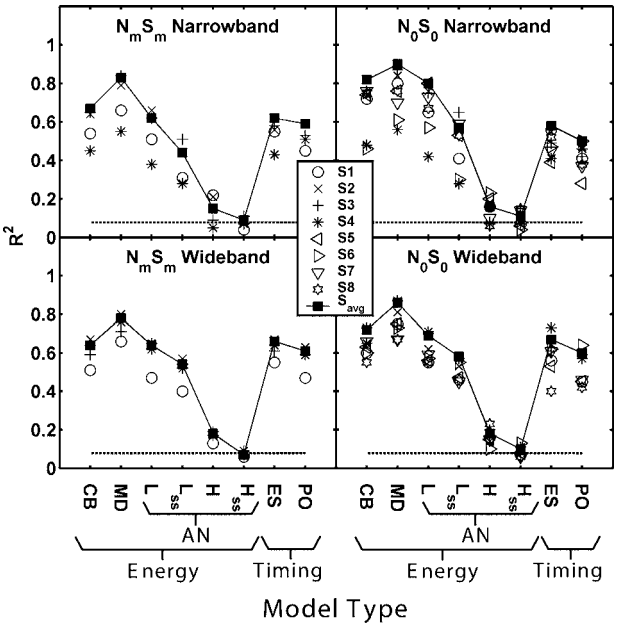

FIG. 4. Summary of model-subject R^2 values for combined tone-plus-noise and noise-alone trials for the eight models tested in this work. Individual subject data are shown by different symbols. The average-subject data are shown by the solid line with closed squares. Note that the horizontal axis is not a continuous variable; lines connect symbols corresponding to the average subject to facilitate comparisons between models. Model abbreviations are as in Fig. 2. Recall that CB, MD, and AN models use energy-based decision variables. ES and PO use temporally-based decision variables. R^2 values exceeding the $p<0.05$ significance levels are above the thick-dashed line.

Modeling results are presented in Figs. 3 and 4. Each panel in Fig. 3 shows the ogive fit of an individual model to the average subject. Each panel also shows a scatter plot of the average subject's $P(Y|T+N)$ (squares) and $P(Y|N)$ (circles) values versus the model decision variable for each waveform. Model abbreviations are as in Fig. 2. Values for R^2 and d' are reported in the panels of Fig. 3 for each model in all four listening conditions. Note that the high-spontaneous-rate auditory-nerve models were omitted from Fig. 3 because of their relatively poor performance. In Fig. 4, model types are listed on the abscissa and the ordinate shows the proportion of variance in subject detection patterns predicted by each model (R^2). Scatter plots for individual-subject model predictions are available at http://web.syr.edu/~lacarney/auditory.htm. In the following sections, each model is explicitly described and evaluated. A final section addresses effects of roving the stimulus level.

B. Critical-band model

1. Methods

Several energy-based models were implemented and compared to monotic and diotic psychophysical data collected from this study and from Evilsizer *et al.* (2002). The first energy-based model, the simple critical-band (CB) model, was composed of a 4th-order linear gammatone filter

centered at the tone frequency (500 Hz). The basic structure of the CB model is shown in the CB panel of Fig. 2. The equivalent-rectangular bandwidth was fixed at 75 Hz to correspond to the auditory-filter estimates of Glasberg and Moore (1990). The rms energy of the filter output was used as the decision variable, similar to the "Energy Mnemonic" described in Gilkey and Robinson (1986). This model had no free parameters in addition to the slope and threshold of the ogive fit [Eq. (5)].

2. Results and Discussion

This model predicts 64%–82% of the variance in the average subject's detection patterns as shown in Figs. 3 and 4 (CB). Model sensitivity was slightly higher for narrowband stimuli ($d'=1.45$) than for wideband stimuli ($d'=1.39$), however both of the resulting d' values were consistent with what would be expected for a model with no internal noise (i.e., $d' \approx \sqrt{2}$, as described above). The finding that stimulus energy was related to subject responses is consistent with previous studies (e.g., Pfafflin and Mathews, 1966; Ahumada *et al.*, 1975; Gilkey and Robinson, 1986). Critical-band model predictions were significant for all subjects in all conditions; however, a substantial portion of the variance in the average subject's detection pattern was unaccounted for under wideband conditions. The lower R^2 values under wideband conditions (with respect to narrowband conditions) can be understood by considering that the model, in contrast to the experimental data (see Table III), predicts a high correlation between narrow and wideband results ($r^2=0.98$).

C. Multiple-detector model

As noted previously, the results of Ahumada and Lovell (1971), Gilkey and Robinson (1986), Isabelle and Colburn (1991), and Evilsizer *et al.* (2002) suggest that information outside the critical-band influences detection patterns for wideband, tone-in-noise detection tasks. In the current section, a more sophisticated energy-based model is explored. Referred to here as the multiple-detector model (MD), this model is capable of incorporating energy outside the critical-band centered at the tone frequency.

1. Methods

The basic model structure is shown in Fig. 2. A linear combination of the RMS output of seven 4th-order, 75-Hz equivalent-rectangular bandwidth (ERB) linear gammatone filters, spaced at 75-Hz intervals from 275 to 725 Hz, was used as the decision variable (adapted from Gilkey and Robinson, 1986). The coefficients for the linear combination were determined using a two-stage fit. The Matlab fminsearch function was used to iteratively adjust six of the seven weights in the linear combination (the weight for the 500 -Hz channel was fixed at 1.0); within each iteration the fminsearch function was called again to find the best fitting ogive between the model decision variable (i.e., the linear combination with the "current" weights) and the subject's $P(Y|W)$ values.

2. Results and discussion

The best-fitting weights had a characteristic shape for wideband stimuli; the strongest weight, which was positive, was observed for the band centered at the tone frequency, and smaller magnitude negative weights were observed for frequency bands above and below the tone frequency.

Because our study also included narrowband (100-Hz) stimuli, it was of interest to consider the importance of filters located more than a critical band away from the tone frequency and centered outside the region of greatest stimulus energy. It was found that dramatically different weights could be obtained for filters more than one critical band above and below the tone frequency with little impact on the variance explained by the model for narrowband stimuli. In order to investigate this phenomenon, the results from a seven-filter MD model were compared to those from a three-filter MD model, each fit to the average subject using narrowband stimuli. The three-filter MD model used only three 4th-order, 75-Hz equivalent-rectangular bandwidth (ERB) linear gammatone filters centered at 425, 500, and 575 Hz, and was otherwise identical to the 7-filter model. Partial correlation coefficients were computed to measure the variability explained by the 7-filter model, while already accounting for the variability explained by a model using only the 3 filters closest to the tone frequency. These coefficients were found to be insignificant ($p>0.1$) for both N_mS_m and N_0S_0 stimulus conditions, indicating that, as expected, energy at filters more than a critical band from the tone frequency did not significantly improve MD predictions for narrowband stimuli. Weights for wideband stimuli were found to be consistent across listeners for all seven filters. The MD model was therefore restricted to the three filters closest to the tone frequency for narrowband stimuli, while all seven filters were included for the wideband stimuli. This model used a total of two free parameters for narrowband stimuli, and six free parameters for wideband stimuli (the central filter weight was always fixed at 1.0), in addition to the slope and threshold of the ogive fit. Figure 5 shows weight patterns across the outputs of the filters in the multiple-detector model for the average subject for all four stimulus conditions.

None of the other models explored in this study included additional model parameters (beyond the slope and threshold of the ogive function) that were *fit to subject data*. In order to more closely compare the multiple-detector model to other models in this work, weight patterns derived for the average subject in the N_0S_0 conditions were used for all MD model predictions. Using the average subject's weights for all subjects served to eliminate the process of fitting additional model parameters to individual subject data, while still fitting the model to the subject data in an average sense. Filter weights were very consistent across listeners.

Results for the MD model using the average subject's weights are shown in Figs. 3 and 4. This model accounted for more variance in the average subjects' detection patterns than any of the other models discussed here. The MD model was able to predict 78%–90% of the variance in the average subject's detection patterns. Figures 3 and 4 show, interestingly, that this model made better predictions for narrowband

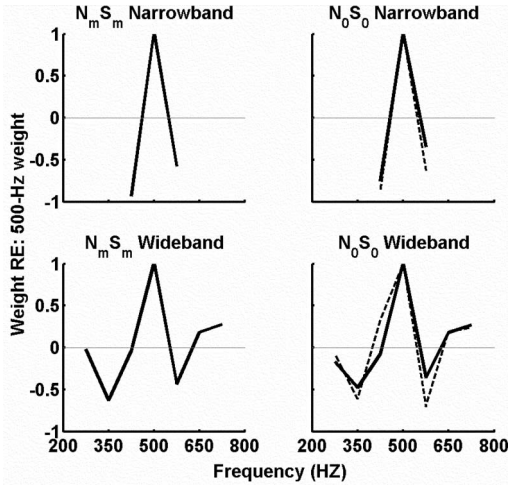

FIG. 5. Spectral weights from the multiple-detector model. The average subject is shown by the thick line. Subject data were taken from this work and that of Evilsizer *et al.* (2002). The thick, dashed line denotes the weighting pattern resulting from a multiple-detector model fit to phase-opponency model hit and false-alarm rates for the narrowband and wideband N_0S_0 stimulus configurations.

stimuli than wideband stimuli. Also note that N_0S_0 predictions were better than N_mS_m predictions, which was expected, because weight patterns were derived in the N_0S_0 condition and the data were averaged across more subjects in the N_0S_0 condition than in the N_mS_m condition. This model was slightly less sensitive under the wideband condition ($d'=1.37$) than the narrowband condition ($d'=1.41$).

Incorporating energy present in channels outside a single critical bandwidth improved model-subject correlations with respect to the critical-band model, which used the energy at the output of a *single* 75-Hz bandwidth filter centered at 500 Hz. This improvement occurred for both wideband and narrowband stimuli. Although little stimulus energy existed outside the critical-band for narrowband stimuli, the reduced weights in the filters adjacent to the filter centered at 500 Hz must have been responsible for improving model predictions with respect to the critical-band model, and thus this information should not be ignored. Implications of the weight pattern are discussed below.

Unlike other models discussed in this study, which had only two free parameters (the slope and intercept of the psychometric function), the MD model incorporated either four or eight free parameters (the three narrowband or seven wideband filter weights with central weights fixed at one, and the slope and mean of the psychometric function). Therefore, it is possible that the increases in R^2 for the MD model with respect to the CB model could have resulted simply from the additional free parameters. To examine this issue, a MD model was fit to the data an additional 100 times, using randomly-created independent sets of 25 noise-alone and 25 tone-plus-noise waveforms within each of the 100 fits. Three filters were used for narrowband stimuli and seven filters were used for wideband stimuli. Central to the test was the fact that the energy at the output of the filter centered at the tone frequency was always computed from the *original sets* of noise waveforms, while energies at the adjacent filters were computed from the *independent sets* of random noise waveforms. The independent sets of random-noise waveforms were not related to those used in the experiment; therefore, any improvement in the fits of the MD model over that of the CB model for the random sets of waveforms must be explained by the fact that the MD model incorporated more free parameters. Using this analysis, R^2-values from the MD model were compared to the R^2 values from the 100 fits under narrowband and wideband, N_0S_0 and N_0S_π listening conditions.

It was found that for 397 out of 400 random-noise fits (for the average subject across all conditions) the R^2 values were below the R^2 value obtained when the MD model was fit using the original set of waveforms. This result suggests that the increases in R^2 for the MD model relative to the CB model was significant, and that the subjects had indeed used energy in the frequency regions outside the critical band as a basis for their decisions about the presence of the signal.

One might attempt to interpret the shape of the MD model's weighting function (Fig. 5) in terms of physiological mechanisms. For example, the combination of central positive and surrounding negative weights is compatible with a neural lateral-inhibitionlike mechanism that may be present in cells with characteristic frequencies (CFs) tuned to frequencies surrounding the 500-Hz stimulus. Consistent with this interpretation, Kopp-Scheinpflug *et al.* (2002) report the presence of broadly tuned inhibitory inputs to spherical bushy cells (SBCs) in the AVCN. They propose that these inputs could be responsible for increased frequency selectivity seen in SBCs. However, it is possible to obtain a weighting function including both negative and positive weights with a purely excitatory model. The dashed, black lines in Fig. 5 show the results of a multiple-detector model fit to a purely-excitatory model (described in detail in Sec. III F) in the N_0S_0 listening condition. Note that the overall shape is similar to that produced by the multiple-detector model fits to psychophysical data. The multiple-detector modeling results could predict about 49% of the variance in narrowband and about 65% of the variance in wideband detection patterns generated by the excitatory model. Thus, the shape of the weighting function cannot be used explicitly to identify excitatory vs inhibitory mechanisms. The negative weights for the filters adjacent to the filter centered at the tone frequency may be explained by phase differences (associated with filter shape) between the central filter and adjacent filters.

D. AN-count-based models

1. Methods

A physiologically-based energy model was tested using the Heinz *et al.* (2001b) AN model and a simple neural-count-based detection strategy. One motivation for testing this model was that the relatively broad tuning of AN fibers at high sound levels may explain the bandwidth effects in the psychophysical results. High-spontaneous-rate (HSR; 50 spikes/s) and low-spontaneous-rate (LSR; 1 spike/s) model fibers were tested. For LSR simulations, the parameters of the inner-hair-cell-to-auditory-nerve synapse were set to reproduce the dynamic range of low-frequency LSR

rate-level functions in response to tones at the fiber's characteristic frequency (Heinz, personal communication). The decision variable was the mean discharge rate (count) of the model fiber, either including or excluding the onset-response (first 5 ms), such that four models were considered. Models denoted as steady-state (SS) excluded the first and last 50 ms of the AN response (onset and offset response). Mean counts were computed from the AN model's rate function; no spikes were generated, and thus the internal noise generally associated with the Poisson properties of AN fibers was not included here.

2. Results and discussion

Results for the low-spontaneous-rate AN-count-based model are shown in Figs. 3 and 4 [L (low-spontaneous rate); L_{SS} (low-spontaneous-rate steady-state)], while high-spontaneous-rate AN count based models are shown in Fig. 4 only [H (high spontaneous rate); H_{SS} (high-spontaneous rate steady-state]. Predictions for HSR fibers, which saturate at low SPLs (i.e., have small dynamic ranges), accounted for only 7%–18% of the variance in the average subject's detection patterns and thus were not included in Fig. 3. HSR-fiber models produced d' values ranging from 0.47 to 0.56, well below the expected value of 1.41, due to effects of rate saturation.

For the narrowband results, the LSR model had $d' = 1.31$, and for wideband results, the LSR model had $d' = 1.24$. The LSR_{SS} model had lower sensitivity for both narrowband ($d' = 1.03$) and wideband ($d' = 1.01$) stimuli, indicating that information in the onset response of the fibers is useful for the detection process. Model LSR fibers predicted between 44% and 80% of the variance in the average subject's narrowband detection patterns and between 54%–69% of the variance in the average subject's wideband detection patterns.

In general, the predictions of LSR AN count-based models were similar to the CB model discussed in Sec. III B and the LSR_{SS} model made slightly poorer predictions. Model predictions based on HSR fibers were barely significantly better than chance at the $p < 0.05$ level and predictions based on HSR_{SS} fibers were insignificant in some conditions (see Fig. 4). In particular, the wider-than-critical-band tuning of AN fibers at high sound levels did not improve upon the performance of the critical-band model for describing detection patterns in response to wideband stimuli. The count-based detection strategy is essentially a peripheral transformation of the single-critical-band detector discussed earlier. However, unlike the critical-band model, the AN-count-based models are subject to saturation and rate-dependent variance (Colburn *et al.*, 2003). This characteristic poses problems for HSR fibers coding high-level stimuli such as those used here, or for coding stimuli with broad dynamic range (e.g., when the level is roved). No improvement was observed using model AN fibers over the CB or MD models. Thus, it was of interest to determine if temporal models for tone-in-noise detection could improve upon these predictions.

E. Envelope-slope model

A number of findings in the literature suggest that subjects can use cues other than energy for detection in noise, whether or not energy cues are available. For example, Richards and Nekrich (1993) found that listeners are able to perform a two-alternative forced-choice tone-in-noise detection task when no reliable level differences between noise-alone and tone-plus-noise stimuli are present. Richards and Nekrich (1993) also measured effects of energy on detection performance with separate level-discrimination tasks for noise-alone and tone-plus-noise stimuli. Performance in these conditions did not fully explain results in which level cues are present, suggesting the existence of cue(s) that do not rely on energy. One such cue is derived from temporal envelope fluctuations in the stimulus waveform and is known as the envelope-slope statistic (see Fig. 2, ES).

1. Methods

We investigated envelope cues using the following modified version of the Zhang (2004) envelope-slope statistic, which was adapted from the statistic derived by Richards (1992),

$$E_{SL} = \frac{\sum_t |x[t - \Delta t] - x[t]|}{\sum_t x[t]} \tag{7}$$

where $x[t]$ is the Hilbert envelope of the output of a 4th-order linear gammatone filter centered at the stimulus frequency (75-Hz ERB), and Δt is the time step over which differences were computed (the sampling frequency). The envelope of the gammatone filter was computed using a Hilbert transform and then low-pass filtered (10th-order maximally-flat IIR filter with a cutoff of 250 Hz) to ensure that any fine structure was removed and that the model decision variable was based solely on envelope fluctuations. The statistic was normalized to remove overall level and duration effects. When the tone was added to the noise waveform, the envelope-slope statistic of the tone-plus-noise complex decreased relative to the envelope-slope statistic of the noise-alone waveform, indicating tone presence. (A tone has an envelope slope equal to 0.) This model used no additional parameters other than the slope and threshold of the ogive fit.

2. Results and discussion

The ES model was slightly more sensitive for wideband stimuli ($d' = 1.37$) than for narrowband stimuli ($d' = 1.31$). Figures 3 and 4 (ES) reveal that this decision variable[4] was significantly correlated with the psychophysical data across reproducible maskers for all conditions. Envelope-slope model predictions approached the critical-band model predictions for wideband stimuli. The envelope-slope model accounted for approximately 65% of the variance in the average subject's wideband detection patterns and approximately 60% of the variance in the average subject's narrowband detection patterns. All predictions were significant at the $p < 0.05$ level.

F. Phase-opponency model

1. Methods

We also tested a single-cell version of the phase-opponency model (Carney *et al.*, 2002), which is a physiologically-based model that takes advantage of the temporal fine structure in AN responses. A block diagram of the single-cell phase-opponency model is shown in Fig. 2 (PO). Two model auditory-nerve fibers were used (Heinz *et al.*, 2001b). The center frequencies of the two fibers were selected (459 and 542 Hz) such that they were symmetrically spaced around the 500-Hz target frequency and had phase responses that differed by 180° at the target frequency at the level used. These AN model outputs were cross correlated. When the tone was present, both AN fibers tended to synchronize with the 500-Hz tone; and the 180° phase shift between the fibers made it unlikely that discharges would occur simultaneously for the two fibers. That is, coincidences were less likely to occur when the target was present. The coincidence-detector output counts were approximated with Eq. (8), which was based on more general expressions for neural coincidence detector output [e.g., Eq. (2) in Colburn (1977), Eq. (6) in Colburn (1996); and Eq. (B1) in Heinz *et al.* (2001a)], specifically,

$$C_M = n_{\text{fib}}^2 T_{\text{CW}} \int_{T_{\text{DUR}}} r_i(t,F_i) r_j(t,F_j) dt, \tag{8}$$

where C_M is the monaural coincidence-detector count, r_i and r_j are rate functions of AN fibers with differing CFs, n_{fib} is the number of AN fiber inputs for each CF, T_{DUR} is the interval over which the coincidence detector response is computed, and T_{CW} is the time window for coincidences. This equation assumes that the width of the time window T_{CW} is narrow compared with the rate of fluctuation in the rate functions in Eq. (8). As noted above, the current model did not generate discharge times; thus the internal noise associated with AN responses was not included. When the two onset probabilities were multiplied in Eq. (8), they exceeded realistic levels and did not produce decision variables correlated to subject data. Therefore, only the sustained portion of the response was used (i.e., the first and last 50 ms of the coincidence detector response were excluded). The $n_{\text{fib}}^2 T_{\text{CW}}$ term at the left of Eq. (8) accounted for the number of fibers at each CF incident on the coincidence-detector cell [Eq. (4.2) in Zhang, 2004] and for the coincidence-window duration [Eq. (6) in Colburn (1996)]. Ten identical fibers were incident on each model coincidence-detecting cell at each CF, and the coincidence-window duration was 20 μs (Carney *et al.*, 2002). Three parameters [AN center frequencies (2) and the number of fibers incident on each cell], were used in this model, in addition to the slope and threshold of the ogive fit.

2. Results and discussion

The PO model was less sensitive than the other models described here for narrowband ($d'=0.89$) and wideband ($d'=0.99$) conditions, as well as less sensitive than expected for a model with no internal noise. As shown in Figs. 3 and 4 (PO) this model was able to predict 50%–59% of the variance in the average subject's narrowband detection patterns, and about 60% of the variance in the average subject's wideband detection patterns.

The PO panels in Fig. 4 show that the phase-opponency model predicted less variance in the psychophysical data than the multiple-detector, envelope-slope and critical-band models. A population version of the PO model (Carney *et al.*, 2002) was also tested, but showed no improvement over the single-cell version for these conditions.

G. Effects of roving stimulus level in a 2-interval detection task

Model predictions were also considered under conditions in which stimulus levels were roved within trials in a 2-interval, 2-alternative forced-choice detection task using random maskers (i.e., not reproducible-noise maskers). Although no psychophysical data were obtained in this study for the roving-level condition, there is little doubt about the nature of these results and that they have important implications for models of detection.

Although both the CB and MD models were able to account for a large amount of variance in the reproducible-noise psychophysical data, these models (in their current implementations) did not correctly predict published thresholds in a roving-level task. For example, Kidd *et al.* (1989) reported psychophysical roving-level data that do not match threshold predictions for a critical-band model. The MD model requires that the negatively-weighted filters be excited in order to be robust to the roving-level paradigm. That is, in the case of a wideband stimulus, the random amount of energy contributed by the noise to the filter at the tone frequency can be effectively removed by adjacent filters that encompass only noise and have approximately equal but negative weights. However, if the stimulus bandwidth is narrow enough that negatively-weighted filters are effectively not stimulated, the MD model shows an increase in threshold that is similar to the critical-band model. Since the current form of the MD model incorporated only the weights derived in this work, it would not be expected to describe this roving-level experiment. The MD model could be improved to make better predictions using roving-level stimuli. That is, the detection mechanism could be modified to adjust either the number of filters and/or their bandwidths for each stimulus bandwidth, or adjust the weights of a fixed set of filters for each stimulus bandwidth, such that the scaled output of the negatively-weighted filters effectively cancels the noise energy at the filter centered at the tone frequency.

Counts of AN fibers are monotonically related to energy; therefore, like the critical band model, the AN models cannot successfully predict thresholds for a roving-level task. If the AN model were expanded into a population model (i.e., that incorporated some sort of weighted sum of AN outputs, analogous to the MD model), its ability to explain results for the roving-level task would involve the same constraints as the MD model (i.e., inaccurate predictions if the bandwidth of the noise was insufficient to excite negatively-weighted filters, and the requirement to readjust weights for each stimulus bandwidth), in addition to rate-saturation problems.

The ES model was unaffected by roving the stimulus level until very narrow noise bandwidths (10 Hz), for which there was an increase in threshold that was greater than that observed in psychophysical data (Kidd *et al.*, 1989).

Carney *et al.* (2002) showed that a phase-opponent detection mechanism can successfully explain results (within 3 dB) for the roving-level detection task for maskers with bandwidths of at least 300 Hz. Further, Zhang and Carney (2004) and Zhang (2004) have shown that a more general cross-frequency coincidence-detection model can explain thresholds in a narrowband condition with roving-level maskers. The general coincidence-detection model is capable of predicting detection thresholds within 3 dB of human listeners in conditions with a 32-dB rove and a 100-Hz stimulus bandwidth, and does so using only HSR fibers. Recall that HSR fibers cannot explain thresholds based on average discharge rate due to saturation, and that HSR fibers comprise the majority of AN fibers.

Thus, it appears that the multiple-detector, phase-opponency, and envelope-slope models are capable of for predicting results from roving-level tasks with wideband noise, however each of these models breaks down when the bandwidth of the stimulus becomes narrow [bandwidth $<$300 Hz for the single cell PO model, $<$75 Hz (1 critical band) for the MD model, and $<$10 Hz for the ES model].

IV. SUMMARY AND FUTURE DIRECTIONS

This study measured detection in monotic and diotic masking conditions with reproducible noise waveforms. Measurements showed that using either monotic or diotic conditions resulted in similar distribution of performance over the individual waveforms. This result adds to the evidence that these conditions involve the same decision processing.

Predictions of energy-based, auditory-nerve, envelope, and phase-opponent cross-frequency coincidence-detection models were evaluated by comparing model decision variables and subject detection patterns. Results showed that the multiple-detector model could explain the most variance in subject detection patterns followed by the critical-band model. The critical-band model, however, cannot explain performance in a roving-level paradigm. The multiple-detector model will require a mechanism to adapt its filters and/or filter weights to specific stimulus conditions in order to predict thresholds under conditions where the stimulus level is roved. The phase-opponency and envelope-based models, although explaining less variance in the subject's detection patterns than energy-based models, are robust to changes in stimulus level. In contrast to the temporal models, HSR AN-count-based physiological implementations of energy-based models saturate, preventing accurate prediction of detection patterns. LSR AN-count-based models accurately predicted subject detection patterns, but were unable to explain performance in a roving-level task. In general, these results suggest the need to consider alternative temporal models for tone-in-noise detection, as well as to pursue the multiple-detector model's potential to describe roving-level results. Ongoing studies designed to manipulate specific detection cues will help direct future temporal-modeling efforts. The energy-based multiple-detector model will also be explored further.

ACKNOWLEDGMENTS

This work was supported by NIDCD R01-01641 (L.H.C.), and R01 DC00100 (H.S.C.). Part of this work was supported by a Research Challenge grant from the Ohio Board of Regents (R.H.G.). We thank Susan Early for her editorial comments and Valerie Shalin and Julio Mateo for assistance with statistical analyses.

APPENDIX: PREDICTABLE VARIANCE FOR INDIVIDUAL-SUBJECT FITS

Extending the analysis presented in Sec. II B 7, it is possible to derive the expected proportion of predictable variance available when fitting models individually to each subject's data, as compared to fitting models to data that have been averaged across subjects, as is the emphasis in this paper. To obtain an estimate of the error variance that is separate from the waveform by subject interaction, the data were rearranged as follows: N_0S_0 and N_mS_m data were combined (to increase the number of trials in the analysis; this decision is justified because the differences between the two conditions are small), and the 192 trials obtained for each combination of waveform and bandwidth were randomly grouped into 12 blocks of 16 trials. Separate Waveform $\times$ Subject $\times$ Block random-effects ANOVAs were performed for each bandwidth. Working from the expected values of the Mean Squares, as described by Kirk (1995, p. 462), and combining terms appropriately, we obtained separate estimates for the variances associated with the effect of waveform (σ_W^2), subject (σ_S^2), block (σ_B^2), the interaction between waveform and subject (σ_{WS}^2), the interaction between waveform and block (σ_{WB}^2), the interaction between subject and block (σ_{SB}^2), and the interaction among waveform, subject, and block combined with error ($\sigma_{\mathrm{WSB}}^2+\sigma_\varepsilon^2$). This procedure lead to some negative variance estimates (which, for our data, were small in magnitude) that were set to zero, as is common practice (e.g., Maxwell and Delaney, 2004, p. 487). For the case in which the models were fit to data averaged across subjects, as in the current study, the proportion of predictable variance is given by

$$\frac{\sigma_P^2}{\sigma_T^2}=\frac{\sigma_W^2}{\sigma_W^2+\frac{\sigma_{\mathrm{WS}}^2}{N_S}+\frac{\sigma_{\mathrm{WB}}^2}{N_B}+\frac{\sigma_{\mathrm{WSB}}^2+\sigma_\varepsilon^2}{N_SN_B}}, \tag{A1}$$

where σ_P^2 is the predictable variance in the average detection pattern,σ_T^2 is the total variance in the average detection pattern, N_S is the number of subjects, and N_B is the number of 16-trial blocks combined to estimate each value in the detection pattern. For the case in which the models were fit separately to the data of individual subjects, predictable variance is given by

TABLE VI. Estimated proportion of predictable variance in $P(Y|W)$ values averaged across the N_mS_m and N_0S_0 stimulus configurations under both narrow and wide stimulus bandwidths. Results are presented for Eq. (A1) which estimates the proportion of predictable variance for the average subject, and equation Eq. (A2), which estimates the proportion of predictable variance for the case in which models are fit to the data of individual subjects. Results are also presented for Eq. (4) using the subject data that was combined across the N_mS_m and N_0S_0 stimulus configurations (see text).

Bandwidth	Eq. (A1)	Eq. (A2)	Eq. (4) Combined
NB	0.944	0.989	0.945
WB	0.982	0.991	0.982

$$\frac{\sigma_P^2}{\sigma_T^2} = \frac{\sigma_W^2 + \sigma_S^2 + \sigma_{\mathrm{WS}}^2}{\sigma_W^2 + \sigma_S^2 + \sigma_{\mathrm{WS}}^2 + \frac{\sigma_{\mathrm{WB}}^2}{N_B} + \frac{\sigma_{\mathrm{SB}}^2}{N_B} + \frac{\sigma_{\mathrm{WSB}}^2 + \sigma_\varepsilon^2}{N_B}}, \quad \text{(A2)}$$

where σ_P^2 is the predictable variance in the individual detection patterns, and σ_T^2 is the total variance in the individual detection patterns.

The estimated proportions of predictable variance for Eqs. (A1) and (A2) are shown in Table VI (columns 1 and 2) for the two bandwidth conditions, with $N_S=4$ and $N_B=12$. Note that both Eq. (A1) and Eq. (4) estimate the proportion of predictable variance for the case when models are fit to data averaged across subjects. The values shown for Eq. (A1) (Table VI, column 1) are slightly larger than those shown for Eq. (4) (Table V, column 4). This difference occurred because the analyses in Table VI were based on 192 trials per waveform (i.e., the data were combined across the N_mS_m and N_0S_0 conditions), whereas the analyses in Table V were based on 96 trials per waveform. To allow a direct comparison between Eq. (A1) and Eq. (4), Eq. (4) was applied to the results of data that had been combined across binaural conditions, such that 192 trials per waveform occurred; these results are shown in Table VI (column 3) and are in agreement with the results of Eq. (A1).

Comparing results from Eq. (A1) and Eq. (A2) (shown in the first two columns of Table VI), we see, as expected, that fitting the models to individual subjects increased the proportion of predictable variance. This was particularly true for the narrow bandwidth case, because there was greater across-subject variability for that stimulus condition.

[1]One model not considered here is that of Dau *et al.* (1996). We were unable to meaningfully apply their model's structure to our single-interval reproducible-noise task.

[2]Although no MLD was found in this study, others have found an $N_mS_m - N_0S_0$ MLD, presumably only at low masker levels. Dolan (1968) observed increasing $N_mS_m - N_0S_0$ MLD for 150-Hz and 300-Hz tones as the spectrum level of a 1000-Hz low-pass filtered noise masker was lowered from 65 to 20 dB SPL. MLDs were less than 1 dB for the 65 dB SPL spectrum level noise and up to 4 dB for the 20 dB SPL spectrum-level noise. Diercks and Jeffress (1962) found a small $N_mS_m - N_0S_0$ MLD (2.8 dB) for a 250-Hz tone, but the masker levels for which this result was achieved were unpublished. Shaw *et al.* (1947) also found an $N_mS_m - N_0S_0$ MLD of 5.8 dB; however masker levels were also unpublished in that work. It is assumed that the latter two studies used relatively low noise spectrum levels in accordance with the results of Dolan (1968). Thus, if this experiment were repeated at lower masker levels that yielded MLDs, it is possible that detection patterns would differ between the two conditions.

[3]We also computed internal-to-external noise ratios for the Evilsizer *et al.* (2002) study. In that study, N_0S_π noise ratios were larger than diotic noise ratios, with the exception of S7, who had the largest estimated noise ratios (ranging from 2.6 to 3.6 in narrowband N_0S_0 and N_0S_π conditions, respectively) and had the lowest first-half, last-half correlations [ranging from 0.52 to 0.76; see Table I of Evilsizer *et al.* (2002)]. This finding indicates that this subject may have changed strategies, thereby increasing internal noise estimates.

[4]An envelope-slope based model was also considered with the basilar-membrane and inner-hair-cell portions of the Heinz *et al.* (2001b) AN model. Results are not presented here because predictions were not significantly different from the envelope-slope model incorporating a gammatone filter. A model based on the standard deviation of the envelope (Richards, 1992) was also considered, but was not included here because predicted detection patterns were not correlated to the psychophysical data.

Ahumada, A., and Lovell, J., (**1971**). "Stimulus features in signal detection," J. Acoust. Soc. Am. **49**, 1751–1756.

Ahumada, A., Marken, R., and Sandusky, A., (**1975**). "Time and frequency analyses of auditory signal detection," J. Acoust. Soc. Am. **57**, 385–390.

Blodgett, H. C., Jeffress, L. A., and Taylor, R. B., (**1958**). "Relation of masked threshold to signal duration for various interaural phase combinations," Am. J. Psychol. **71**, 283–290.

Bruning, J. L., and Kintz, B. L., (**1968**). *Computational Handbook of Statistics* (Scott, Foresman and Company, Glenview, IL).

Carney, L. H., Heinz, M. G., Evilsizer, M. E., Gilkey, R. H., and Colburn, H. S., (**2002**). "Auditory phase-opponency: A temporal model for masked detection at low frequencies," Acta. Acust. Acust. **88**, 334–347.

Colburn, H. S., (**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., (**1996**). "Computational models of binaural processing," in *Auditory Computation*, edited by H. L. Hawkins, T. A. McMullen, A. N. Popper, and R. R. Fay (Springer, New York), pp. 332–400.

Colburn, H. S., Carney, L. H., and Heniz, M. G., (**2003**). "Quantifying the information in auditory-nerve responses for level discrimination," J. Assoc. Res. Otolaryngol. **4**, 294–311.

Costalupes, J. A., Young, E. D., and Gibson, D. J., (**1984**). "Effects of continuous noise backgrounds on rate response of auditory-nerve fibers in cat," J. Neurophysiol. **51**, 1326–1344.

Dau, T., Püschel, D., and Kohlrausch, A., (**1996**). "A quantative model of the 'effective' signal processing in the auditory system," J. Acoust. Soc. Am. **99**, 3615–3622.

Diercks, J. K., and Jeffress, L. A., (**1962**). "Interaural phase and the absolute threshold for tone," J. Acoust. Soc. Am. **34**, 981–984.

Dolan, T. R., (**1968**). "Effect of masker spectrum level and masking-level differences at low signal frequencies," J. Acoust. Soc. Am. **44**, 1507–1512.

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.

Egan, J. P., Lindner, W. A., and McFadden, D., (**1969**). "Masking-level differences and the form of the psychometric function," Percept. Psychophys. **6**, 209–215.

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.

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

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

Gilkey, R. H., and Meyer, T. A., (**1987**). "Modeling subject responses in a reproducible-noise masking task," J. Acoust. Soc. Am. **82**, 92–93.

Gilkey, R. H., and Robinson, D. E., (**1986**). "Models of auditory masking: A molecular phsychophysical approach," J. Acoust. Soc. Am. **79**, 1499–1510.

Glasberg, B. R., and Moore, B. J. C., (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Green, D. M., (**1984**). "Profile analysis: A different view of auditory intensity discrimination," Am. Psychol. **38**, 133–142.

Green, D. M., (**1988**). *Profile Analysis* (Oxford University Press, New York).

Hall, J. W., Haggard, M. P., and Fernandes, M. A., (**1984**). "Detection in

noise by spectro-temporal pattern analysis," J. Acoust. Soc. Am. **76**, 50–56.
Hawkins, J. E., and Stevens, S. S., (**1950**). "The masking of pure tones and of speech by white noise," J. Acoust. Soc. Am. **22**, 6–13.
Heinz, M. G., Colburn, H. S., and Carney, L. H., (**2001a**). "Rate and timing cues associated with the cochlear amplifier: Level discrimination based on monaural cross-frequency coincidence detection," J. Acoust. Soc. Am. **110**, 2065–2084.
Heinz, M. G., Zhang, X., Bruce, I. C., and Carney, L. H., (**2001b**). "Auditory-nerve Model for Predicting Performance Limits of Normal and Impaired Listeners," ARLO **2**, 91–96.
Hirsh, I. J., and Burgeat, M., (**1958**). "Binaural effects in remote masking," J. Acoust. Soc. Am. **30**, 827–832.
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., (**1968**). "Mathematical and electrical models of auditory detection," J. Acoust. Soc. Am. **44**, 187–203.
Kiang, N. Y., and Moxon, E. C., (**1974**). "Tails of tuning curves of auditory-nerve fibers," J. Acoust. Soc. Am. **55**, 620–630.
Kidd, G., (**1987**). "Auditory discrimination of complex sounds: The effects of amplitude perturbation on spectral shape," in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NY), pp. 16–25.
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.
Kirk, R. E., (**1995**). *Experimental Design: Procedures for the Behavioral Sciences*, 3rd ed. (Brooks/Cole, Pacific Grove, CA).
Kopp-Scheinpflug, C., Dehmel, S., Dörrscheidt, G. J., and Rübsamen, R., (**2002**). "Interaction of excitiation and inhibition in anteroventral cochlear nucleus neruons that receive large endbulb synaptic endings," J. Neurosci. **22**, 11004–11018.
Langhans, A., and Kohlrausch, A., (**1992**). "Differences in auditory performance between monaural and diotic conditions. I. Masked thresholds in frozen noise," J. Acoust. Soc. Am. **91**, 3456–3470.
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 University Press, New York).
Maxwell, S. E., and Delaney, H. D., (**2004**). *Designing Experiments and Analyzing Data: A Model Comparison Perspective* (Lawrence Erlbaum Associates, Mahwah, NJ), pp. 469–517.
Neff, D. L., and Callaghan, B. P., (**1988**). "Effective properties of mulit-component simultaneous maskers under conditions of uncertainty," J. Acoust. Soc. Am. **83**, 1833–1888.
Patterson, R. D., (**1976**). "Auditory filter shapes derived with noise stimuli," J. Acoust. Soc. Am. **59**, 640–654.
Pfafflin, S. M., (**1968**). "Detection of Auditory Signal in Restricted Sets of Reproducible Noise," J. Acoust. Soc. Am. **43**, 487–490.
Pfafflin, S. M., and Mathews, M. V., (**1966**). "Detection of auditory signals in reproducible noise," J. Acoust. Soc. Am. **39**, 340–345.
Rhode, W. S., (**1971**). "Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique," J. Acoust. Soc. Am. **49**, 1218–1231.
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.
Ruggero, M. A., (**1973**). "Responses to noise of auditory-nerve fibers in the quirrel monkey," J. Neurophysiol. **36**, 569–587.
Schalk, T. B., and Sachs, M. B., (**1980**). "Nonlinearities in auditory-nerve fiber responses to bandlimited noise," J. Acoust. Soc. Am. **67**, 903–913.
Sever, J. C., and Small, A., M., (**1979**). "Binaural critical-bands," J. Acoust. Soc. Am. **66**, 1343–1350.
Shannon, R. V., (**1976**). "Two-tone unmasking and suppression in a forward-masking situation," J. Acoust. Soc. Am. **59**, 1460–1470.
Shaw, W. A., Newman, E. B., and Hirsh, I. J., (**1947**). "The difference between monaural and binaural thresholds," J. Exp. Psychol. **37**, 229–242.
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.
Spiegel, M. F., and Green, D. M., (**1981**). "Two procedures for estimating internal noise," J. Acoust. Soc. Am. **70**, 69–73.
Zhang, X., (**2004**). "Cross-frequency coincidence detection in the processing of complex sounds," Ph.D. dissertation, Boston University.
Zhang, X., and Carney, L. H., (**2004**). "Models for monaural cross-frequency coincidence-detectors in tone-in-noise detection tasks," Abstract 842, Association for Research in Otolaryngology.
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.