

Research Article

Tone-in-Noise Detection Using Envelope Cues: Comparison of Signal-Processing-Based and Physiological Models

JUNWEN MAO¹ AND LAUREL H. CARNEY²

¹*Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY, USA*

²*Departments of Biomedical Engineering and Neurobiology and Anatomy, University of Rochester, Rochester, NY, USA*

Received: 12 March 2014; Accepted: 16 September 2014; Online publication: 30 September 2014

ABSTRACT

Tone-in-noise detection tasks with reproducible noise maskers have been used to identify cues that listeners use to detect signals in noisy environments. Previous studies have shown that energy, envelope, and fine-structure cues are significantly correlated to listeners' performance for detection of a 500-Hz tone in noise. In this study, envelope cues were examined for both diotic and dichotic tone-in-noise detection using both stimulus-based signal processing and physiological models. For stimulus-based envelope cues, a modified envelope slope model was used for the diotic condition and the binaural slope of the interaural envelope difference model for the dichotic condition. Stimulus-based models do not include key nonlinear transformations in the auditory periphery such as compression, rate and dynamic range adaptation, and rate saturation, all of which affect the encoding of the stimulus envelope. For physiological envelope cues, stimuli were passed through models for the auditory nerve (AN), cochlear nucleus, and inferior colliculus (IC). The AN and cochlear nucleus models included appropriate modulation gain, another transformation of the stimulus envelope that is not typically included in stimulus-based models. A model IC cell was simulated with a linear band-pass modulation filter. The average discharge rate and response fluctuations of the model IC cell were compared to human performance. Previous studies have predicted a significant amount of the variance across reproducible noise maskers in listeners' detection using stimulus-

based envelope cues. In this study, a physiological model that includes neural mechanisms that affect encoding of the stimulus envelope predicts a similar amount of the variance in listeners' performance across noise maskers.

Keywords: tone-in-noise detection, envelope, physiological model

INTRODUCTION

Hearing in noise is a challenge for listeners both with and without hearing loss, yet the healthy auditory system accomplishes this task with remarkable ability. Previous studies using stimulus-based signal-processing strategies have identified envelope-based cues that play an important role in detection. The goal of this study was to determine whether nonlinear physiological models, which both distort and enhance envelope cues, can perform as well as the simple stimulus-based models in predicting listeners' detailed performance.

Early studies of tone-in-noise detection (e.g., Blodgett et al. 1958, 1962; Dolan and Robinson 1967) used random noises for each trial to obtain listeners' detection thresholds. Two listening conditions have been tested most extensively (Fig. 1): the diotic condition, in which identical noise-alone or tone-plus-noise stimuli are presented to the two ears, and the dichotic condition, in which out-of-phase tones and in-phase noises are presented to the two ears for the tone-plus-noise trials.

Reproducible noises, a set of pre-generated random noises, have been used to explore listeners' detailed detection performance across different noise

Correspondence to: Laurel H. Carney · Departments of Biomedical Engineering and Neurobiology and Anatomy · University of Rochester · Rochester, NY, USA. Telephone: 585-276-3948; email: laurel_carney@urmc.rochester.edu

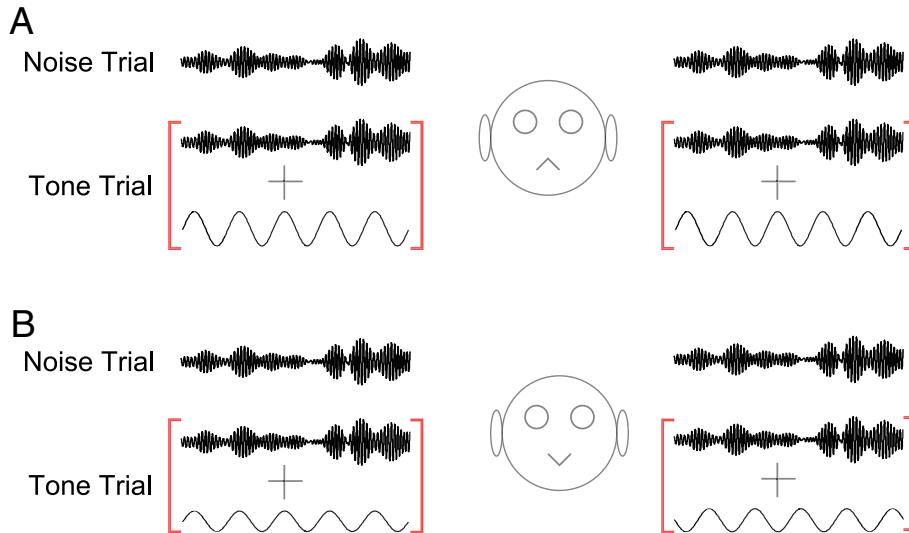


FIG. 1. A schematic diagram illustrates two basic listening conditions: **A** diotic condition, $N_O S_O$, and **B** dichotic condition, $N_O S_{\pi}$. Note that the tone is inverted in the lower right panel (after Moore 2003).

waveforms (Pfafflin and Mathews 1966; Gilkey et al. 1985; Siegel and Colburn 1983; Isabelle and Colburn 1991; Evilsizer et al. 2002; Davidson et al. 2006; Mao et al. 2013). Listeners' responses vary across repetitions of any given stimulus, due to unknown factors which are referred to as internal noise that are assumed to be independent across trials. Each reproducible noise waveform is tested multiple times to yield average response patterns that reduce the effects of internal noise. Listeners' performance can be described in terms of the set of hit (correct identification of "tone present") and false alarm (FA) rates for each reproducible noise, referred to as a detection pattern (Davidson et al. 2006) (Fig. 2). Individuals' detection patterns are highly consistent over time (Gilkey et al. 1985; Isabelle and Colburn 1991; Evilsizer et al. 2002; Davidson et al. 2006) and, in many conditions, across listeners, suggesting the use of similar strategies and cues. Several models have been proposed to predict listeners' performance using specific cues. In each model, a decision variable (DV) is computed based on a feature of the stimulus, and DVs from the set of reproducible noises are compared to listeners' detection patterns.

For diotic detection, energy within a critical band (Fletcher 1940) predicts a significant amount of the variance in listeners' performance across different maskers (Davidson et al. 2009; Mao et al. 2013). However, the energy cue cannot explain performance in roving-level conditions, in which the overall sound level varies in each trial (Kidd et al. 1989). An envelope slope (ES) cue (Richards 1992; Zhang 2004) is robust for the roving-level condition and predicts a significant amount of the variance in listeners' detection patterns (Davidson et al. 2009;

Mao et al. 2013). Models based on temporal fine structure such as the phase opponency model (Carney et al. 2002) also predict a smaller but significant portion of the variance in detection patterns. A recent study shows that predictions based on an optimal combination of energy and temporal cues approach the predictable variance in detection patterns (the common variance among different listeners' performance) for the diotic condition (Mao et al. 2013).

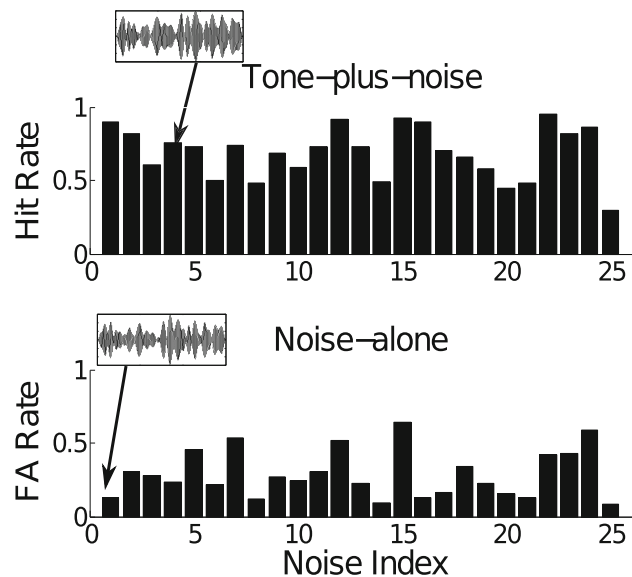


FIG. 2. Detection pattern (hit and FA rates) of the average listener for diotic narrowband waveforms. The horizontal axis shows noise index; the insets show examples of tone-plus-noise (top) and noise-alone (bottom) waveforms. Note that listeners' responses vary across reproducible waveforms; responses were highly consistent within and across individual listeners for this stimulus condition.

For dichotic detection, interaural level difference (ILD) and/or interaural time difference (ITD) cues have been used to predict listeners' performance (Davidson et al. 2009; Isabelle 1995); however, these predictions were substantially lower than the predictable variance. A binaural envelope slope cue, the slope of the interaural envelope difference (SIED), yields significantly better predictions than ILD and/or ITD cues (Mao and Carney 2014). Thus, among stimulus-based models, those using envelope cues are most successful in predicting listeners' performance for both diotic and dichotic conditions.

Envelope cues are important for detection in noise, but these cues are affected by every stage of neural processing. Nonlinearities in the auditory periphery that affect envelopes, including compression, rate saturation, and rate and dynamic range adaptation, are included in recent computational models for the auditory nerve (AN) (Zilany and Bruce 2006, 2007; Zilany et al. 2009, 2014; Zilany and Carney 2010). Dau et al. (1996) and Breebaart et al. (2001) simulated auditory signal processing for masked detection using models that included linear band-pass filters, rectification, and adaptation. The current study combined recent nonlinear AN models with models for envelope processing in the cochlear nucleus and inferior colliculus (IC) (Nelson and Carney 2004), where envelope tuning is first observed along the ascending auditory pathway (reviewed in Joris et al. 2004). Using the detailed detection performance available from reproducible noise studies, it was possible to investigate the physiological processing of envelope cues for masked detection and to test the hypothesis that physiological models can perform as well as the simpler signal-processing-based models in predicting listeners' performance. For physiological models, responses were assessed based on the average discharge rate and on temporal fluctuations in the responses of monaural and binaural model IC cells. Given the success of the stimulus-based ES and SIED models for predicting listeners' performance, these models served as benchmarks for the physiological models to investigate the effectiveness of cues extracted using models with nonlinear neural mechanisms that influence envelope coding.

METHODS

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

Datasets

Listeners' detection performance for diotic and dichotic tone-in-noise detection was obtained from previous experiments (Davidson et al. 2006; Evilsizer et al. 2002; Mao and Carney 2014). Detection in the presence of reproducible noise maskers was tested on each listener in these studies. Each single-interval trial had a 50 % chance of being either a noise-alone or tone-plus-noise trial, and the masker waveform was randomly chosen from the set of reproducible waveforms. Listeners responded "tone present" or "tone not present", and their performance was described in terms of hit rate (proportion of correct response of tone present for tone-plus-noise waveforms) and FA rate (proportion of responding tone present for noise-alone waveforms) for each reproducible masker waveform. The set of hit and FA rates across the ensemble of maskers is referred to as a detection pattern (Davidson et al. 2006). Figure 2 shows the detection pattern of the average listener (i.e., averaged performance across individual listeners) for diotic narrowband waveforms.

Data from diotic and dichotic listening conditions were used in this study. Both narrowband (452–552 Hz) and wideband (100–3,000 Hz) waveforms were used for diotic and dichotic conditions. The narrowband maskers were created by extracting the 100-Hz band of the spectrum from 452 to 552 Hz from the wideband maskers. The spectrum level of the noise waveform was 40 dB sound pressure level (SPL) (overall noise level was 60 dB SPL for narrowband waveforms and approximately 75 dB SPL for wideband waveforms). For the predictions here, the 500-Hz tone level was set to each listener's threshold. Listeners' detection thresholds for the dichotic condition were approximately 10 dB lower than those in the diotic condition. This decrease in threshold, the well-known binaural masking level difference (Moore 2003), was due to the binaural differences introduced in the dichotic condition. In this study, data from a total of eight listeners for the diotic condition (S1–S4 from Evilsizer et al. 2002 and S5–S8 from Davidson et al. 2006) and six listeners for the dichotic condition (S1–S4 from Evilsizer et al. 2002 and S9 and S10 from Mao and Carney 2014) were analyzed.

Diolic Models for Tone-in-Noise Detection

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

Stimulus-Based Model. The original ES model (Richards 1992; Zhang 2004; Davidson et al. 2009) quantifies changes in envelope fluctuations. Because the addi-

tion of a tone to a narrowband noise waveform results in a decrease of the envelope fluctuations, a low value for the DV indicates that the testing waveform is more likely to be a tone-plus-noise stimulus. By inspecting the frequency components of envelopes from tone-plus-noise and noise-alone stimuli, it was determined that the largest differences in envelope energy were within 50–150 Hz (Mao et al. 2013). Thus, a sixth-order band-pass filter with a center frequency of 100 Hz ($Q=1$) was used to extract the envelope frequency range of interest. Figure 3A shows the schematic diagram of the modified ES model. The Hilbert transform is used to compute the analytic signal from the output of a fourth-order gammatone filter (center frequency of 500 Hz). The envelope is obtained from the analytic signal, and the DV of the model was calculated as the integral of the half-wave rectified slope of the envelope. The difference between this modified model and the original ES model (Richards 1992; Zhang 2004; Davidson et al. 2009) is that a tenth-order low-pass filter (cutoff frequency at 250 Hz, aiming to exclude the high frequency fine-structure components) is replaced by

the band-pass filter to extract envelope cues from the most informative frequency range. A previous study showed that predictions based on the modified ES model (Mao et al. 2013) were more consistent with listeners' performance than those using the original ES model.

Physiological Model. In the physiological model, the stimulus was passed through a series of phenomenological models along the ascending auditory pathway (Fig. 3B). First, a human version of the AN model (Zilany et al. 2009, 2014; Ibrahim and Bruce 2010) was used to obtain the AN synapse output. The input to the AN model was first processed by a middle ear filter, followed by a set of band-pass filter paths that provided inputs to the inner hair cell (IHC). The IHC response provided the input to the synapse model, which provided the final model AN response. This AN model has been shown to simulate responses to a range of different stimuli accurately, including pure tones, forward masked stimuli, and amplitude-modulated (AM) stimuli (Zilany et al. 2009). Next, model AN responses were used as inputs to a cochlear nucleus (CN) model (Nelson and Carney 2004). Inhibitory and excitatory AN responses tuned to the same frequency were processed through low-pass filters representing convolution with postsynaptic potential waveforms and were then weighted by excitatory and inhibitory strengths and combined to provide the response of the CN model. For the purposes of the current study, the function of the CN model was to provide modest temporal tuning (i.e., increased synchrony to the envelope) as observed for several cell types in the anteroventral cochlear nucleus (Frisina et al. 1990; Gai and Carney 2006, 2008); no attempt was made to match the model to a particular cell type or CF (Gai and Carney 2006). Increased synchrony was accomplished using an inhibitory strength that was 0.6 times the excitatory strength. Higher values of the inhibitory input strength would yield stronger synchrony to the envelope (e.g., appropriate for onset cells), but little or no sustained response to noise stimuli, which was important for studying responses to the reproducible maskers. Thus, an intermediate value of inhibitory strength that resulted in improved synchrony but still yielded sustained responses was used for all simulations here. This value was the same as that used for the CN stage of the same-frequency inhibitory-excitatory (SFIE) model in Nelson and Carney (2004).

Krishna and Semple (2000) and Nelson and Carney (2007) showed that approximately half of the cells in the IC have band-pass tuning to AM. The IC cell is simulated with a modulation filter to represent this tuning in the model. Specifically, the IC responses are modeled by a band-pass modulation filter, with a peak, or best modulation frequency (BMF) that

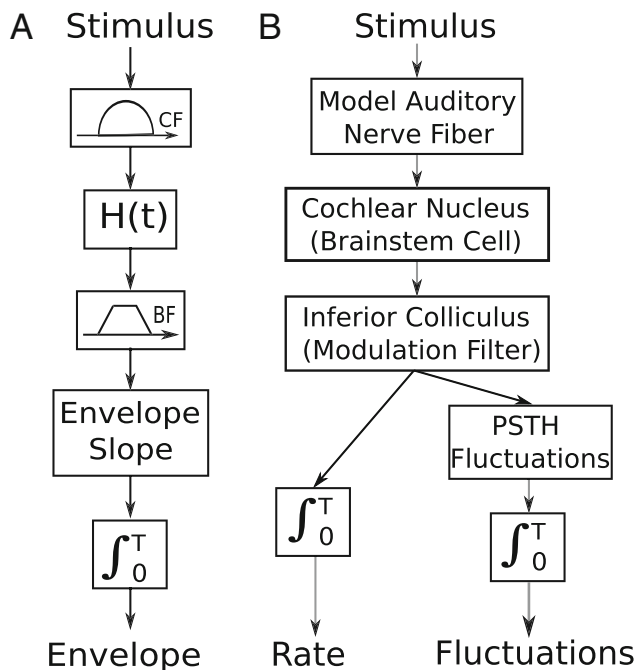


FIG. 3. Schematic diagrams of the monaural envelope models. **A** Stimulus-based modified ES model. The envelope was obtained from the analytic waveform computed from the Hilbert transform of a fourth-order gammatone-filtered waveform; a sixth-order band-pass filter with center frequency of 100 Hz was used to extract the envelope frequency range of interest. Cue values were computed as the integral of the half-wave rectified slope of the envelope fluctuations. **B** Physiological envelope model. The stimulus was passed through the AN and CN phenomenological models and the IC modulation filter to obtain the synapse output. The neural envelope cue was computed based on the rate and fluctuation of the model synapse output.

receives its input from the synapse output of the CN model. The SFIE-type IC model used by Nelson and Carney (2004) acts as a modulation filter, and a number of center frequencies can be achieved by carefully choosing time constants for the excitatory and inhibitory inputs. The Q value of the SFIE-type model is approximately 1.5. Preliminary results showed that better predictions were obtained with a broader filter. In this study, a sixth-order band-pass filter was used that allowed more flexible adjustment of the center frequency, as compared to the original SFIE model, and a Q value of 1 was used (see Appendix). The order of the filter used here was determined by the phase range obtained from physiological recordings of IC cells in awake rabbits (unpublished observations). Two basic neural mechanisms were used to obtain envelope cues from the physiological model: average discharge rate, which was computed as the averaged synapse output, and a measure related to the fluctuations of the model's peri-stimulus time histograms, which was obtained from the integral of the half-wave rectified derivative of the model response.

Dichotic Models for Tone-in-Noise Detection

Similar to the diotic condition, both stimulus-based and physiological envelope models were used in the dichotic condition. For noise-alone trials in the dichotic condition, listeners were presented with identical noise stimuli to the two ears; thus, there were no binaural stimulus differences in the noise-alone trials. Stimulus-based models that depend on binaural differences cannot make any predictions for dichotic noise-alone trials. As a result, only hit rates were predicted for the stimulus-based models in the dichotic condition. In contrast, for the physiological model, binaural differences were available in models that included cell inputs with mismatched center frequencies, and thus, FA rates could be predicted in that case.

Stimulus-Based Model. For the dichotic condition, binaural differences occur because of the addition of out-of-phase tones to in-phase noises at the two ears. The SIED model (Mao and Carney 2014) focuses on the binaural envelope difference cues. Figure 4A shows a schematic diagram of the SIED model, in which envelopes from the contralateral and ipsilateral sides are extracted from the analytic signal computed from a fourth-order gammatone-filtered stimulus. The binaural envelope difference was calculated based on the difference between the computed monaural envelopes. Specifically, the SIED cue value was calculated as the time integral of the half-wave rectified slope of the envelope differences. It has been shown that the SIED cue represents a nonlinear combination of interaural time and level differences (Mao and Carney 2014).

Physiological Model. Computation of the dichotic physiological envelope cue is shown in Figure 4B. Similar to the diotic physiological model (Fig. 3B), binaural stimuli are first passed through a series of phenomenological cell models along each monaural pathway. Model synapse outputs are obtained from AN and CN models for both contralateral and ipsilateral sides. The excitatory response from the contralateral CN model is combined with a delayed (2 ms) inhibitory response from the ipsilateral CN model via an inhibitory interneuron. The combination of the CN outputs represents the binaural difference, which includes both interaural time and level differences. The combined excitatory and inhibitory inputs are sent to the IC modulation filter. For the IC model, a band-pass modulation filter centered at 50 Hz was used; this modulation frequency contains the largest envelope differences related to tone presence in the dichotic condition (Mao and Carney 2014). The dichotic physiological model envelope cues based on either rate or response fluctuations were obtained from the model IC synapse output.

Evaluation of Model Predictions

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

RESULTS

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

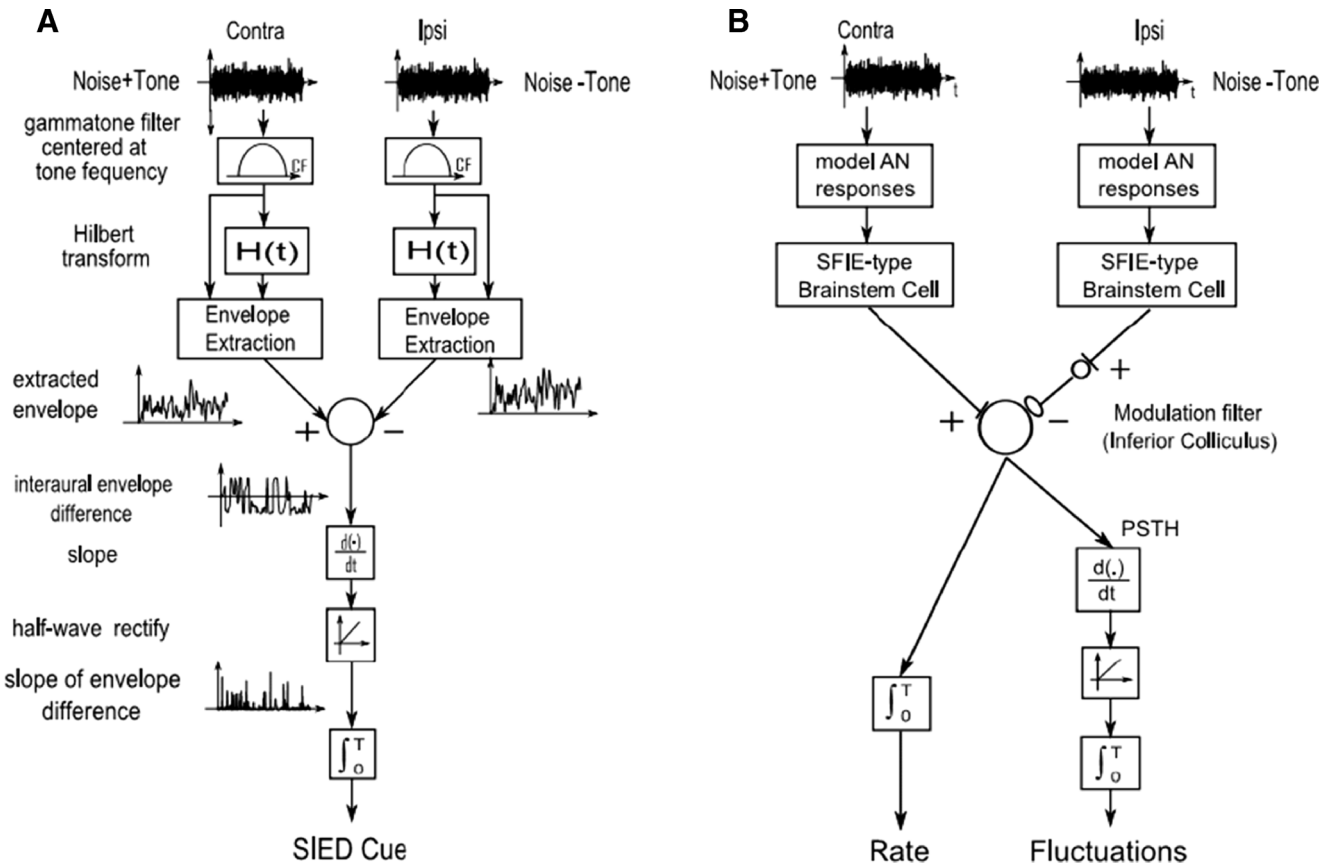


FIG. 4. A schematic diagram of the dichotic envelope models. **A** Dichotic stimulus-based SIED cue. Envelope was extracted using analytic signal computed from the Hilbert transform of the fourth-order gammatone-filtered waveforms. The SIED cue was computed as the time integral of the half-wave rectified slope of the envelope difference at the two ears. **B** Dichotic physiological envelope cue. Binaural stimuli were passed through the AN and CN models;

excitatory contralateral and delayed inhibitory ipsilateral CN outputs were combined to compute the binaural envelope difference. Responses from the CN outputs were sent to the IC cell. The IC cell was simulated with a band-pass modulation filter, and the envelope cue from the IC cell was computed based on the average rate and fluctuation of the model's response.

Detection patterns were highly correlated across different pairs of listeners in the diotic narrowband and wideband and dichotic wideband conditions (Mao et al. 2013; Mao and Carney 2014), indicating that listeners used a similar strategy to detect tones in noise in each of these conditions. In this study, model predictions are only shown for the average listeners in these three conditions. For the dichotic narrowband condition, in which listeners' patterns were not significantly correlated (Mao and Carney 2014), model predictions are shown for individual listeners. Model predictions using stimulus-based envelope cues were reported in previous studies (Mao et al. 2013; Mao and Carney 2014). Here, quantitative comparisons of stimulus-based and physiological envelope cues are shown in Tables 1 and 2.

Diotic Physiological Cues

Figure 5(A–C) shows model predictions of average listeners' narrowband detection patterns using stimulus-based envelope cues, average rates, and

fluctuations computed from the model IC cell responses. Predictions based on the same cues for the average listeners' wideband detection patterns are shown in Figure 5(D–F). In each panel, the x -axis shows the model center frequencies and the y -axis shows the proportion of variance in the detection pattern that is explained by the model.

In both narrowband and wideband conditions, the trends of model predictions across different frequency channels were similar, with the highest correlation to listeners' detection patterns obtained at or near the 500-Hz target tone frequency. In addition, maximal predictions from the stimulus-based envelope cue and the physiological rate and fluctuations cues were similar in these two conditions (Fig. 5(A–C) for narrowband; Fig. 5(D–F) for wideband). In the narrowband condition, the frequency range that yielded the highest correlation to listeners' hit rates was approximately 530 Hz for both stimulus-based and physiological cues. The reason that the maximal correlation to listeners' performance occurred for the 530-Hz channel was likely due to the phase

TABLE 1

| Stimulus-based and physiological model predictions of diotic and dichotic tone-in-noise detection patterns for the average listener | | | | | | |
|---|----------------------|--------------------|----------------------------|--------------------|--------------------|----------------------------|
| | Narrowband condition | | | Wideband condition | | |
| | Stimulus-based ES | Physiological rate | Physiological fluctuations | Stimulus-based ES | Physiological rate | Physiological fluctuations |
| Diotic condition | | | | | | |
| Hit | <i>0.53</i> | <i>0.50</i> | <i>0.54</i> | <i>0.50</i> | <i>0.53</i> | <i>0.48</i> |
| FA | <i>0.46</i> | <i>0.32</i> | <i>0.33</i> | <i>0.38</i> | <i>0.52</i> | <i>0.46</i> |
| Dichotic condition (hit) | See Table 2 | | | <i>0.53</i> | <i>0.46</i> | <i>0.52</i> |

The values listed indicate the proportion of variance explained by each model; the values in italics indicate predictions that are statistically significant ($r^2 \geq 0.16$; $p < 0.05$). For these values, the center frequency for the model cell was 500 Hz and the best modulation frequency for the IC band-pass modulation filter was 100 Hz in the diotic condition and 50 Hz for the dichotic condition

properties of the gammatone filter, as no significant difference in envelope energy was observed across these channels. In the wideband condition, maximal correlations to listeners' detection patterns were obtained from model cells tuned near the frequency of the 500-Hz tone.

Dichotic Physiological Cues

For the stimulus-based SIED cue, FA rates cannot be predicted because there are no interaural differences in the identical noise-alone stimuli that were presented to the listener. In the physiological models, assuming that model IC cells receive inputs from AN synapse outputs with mismatched center frequencies (Joris et al. 2006), predictions of both hit and FA rates can be computed. In this section, each figure shows dichotic model predictions with different combinations of mismatched center frequencies. As described previously, the average listener was analyzed in the wideband condition because listeners' detection patterns were highly correlated with each other, and individual listeners were analyzed for the narrowband condition because different listeners apparently use different strategies for detection of tones narrowband stimuli. In particular, different binaural combinations of frequency channels provide better predictions for different individual listeners in the narrowband dichotic condition (Mao and Carney 2014).

In Figure 6, dichotic model predictions of hit and FA rates for the average listener in the wideband condition (A, B) and three individual listeners in the narrowband condition [S1 (C, D), S3 (E, F), and S4 (G, H)] are shown. These individual listeners were chosen because their results were representative of the three other listeners. In each panel, the axes show the center frequencies of the model cells that received stimuli presented to the left and right ears. Predictions for the binaural models that receive matched 500-Hz frequency channels from both ears

are shown at the lower left of each panel; the other predictions are for binaural combinations of different (mismatched) frequency channels from the two ears.

For the wideband condition, the trends of model predictions using mismatched AN inputs are similar: the highest correlation of hit rate was observed for model cells that received left and right inputs with AN tuning near the tone frequency of 500 Hz; for FA rates, predictions with ipsilateral inputs around 500 Hz were high (Fig. 6(A, B)). For the narrowband condition, the trends in the predictions varied across listeners: some were best predicted by frequency channels around 500 Hz, and others were best predicted by frequency channels away from the tone frequency for both ears (Fig. 6(C–G)). For instance, S3's detection patterns were best predicted by rate information from frequency channels near 500 Hz, and S4's hit rates were best predicted by frequency channels that were approximately one critical bandwidth apart. The diversity in these results implies that different listeners use different strategies for narrowband tone-in-noise detection, which also may explain

TABLE 2

| Stimulus-based and physiological model predictions for individual listeners' hit rates in the dichotic narrowband condition | | | | | |
|---|-------------------------|--------------------|-------------|----------------------------|-------------|
| Model predictions | Stimulus-based SIED cue | Physiological rate | | Physiological fluctuations | |
| | Hit | Hit | FA | Hit | FA |
| S1 | <i>0.18</i> | <i>0.31</i> | <i>0.23</i> | <i>0.30</i> | <i>0.22</i> |
| S2 | <i>0.31</i> | <i>0.15</i> | <i>0.36</i> | <i>0.22</i> | <i>0.36</i> |
| S3 | <i>0.00</i> | <i>0.26*</i> | <i>0.28</i> | <i>0.31*</i> | <i>0.24</i> |
| S4 | <i>0.16</i> | <i>0.52*</i> | <i>0.56</i> | <i>0.50*</i> | <i>0.29</i> |
| S5 | <i>0.10</i> | <i>0.28</i> | <i>0.41</i> | <i>0.38</i> | <i>0.30</i> |
| S6 | <i>0.13</i> | <i>0.14</i> | <i>0.22</i> | <i>0.12</i> | <i>0.15</i> |

The values listed indicate the maximum value of the percent variance accounted for within the range of CFs in Figures 6 and 7. Values in italics indicate statistical significance ($r^2 \geq 0.16$; $p < 0.05$)

* $p < 0.05$, indicating predictions that were significantly improved compared to the stimulus-based SIED cue

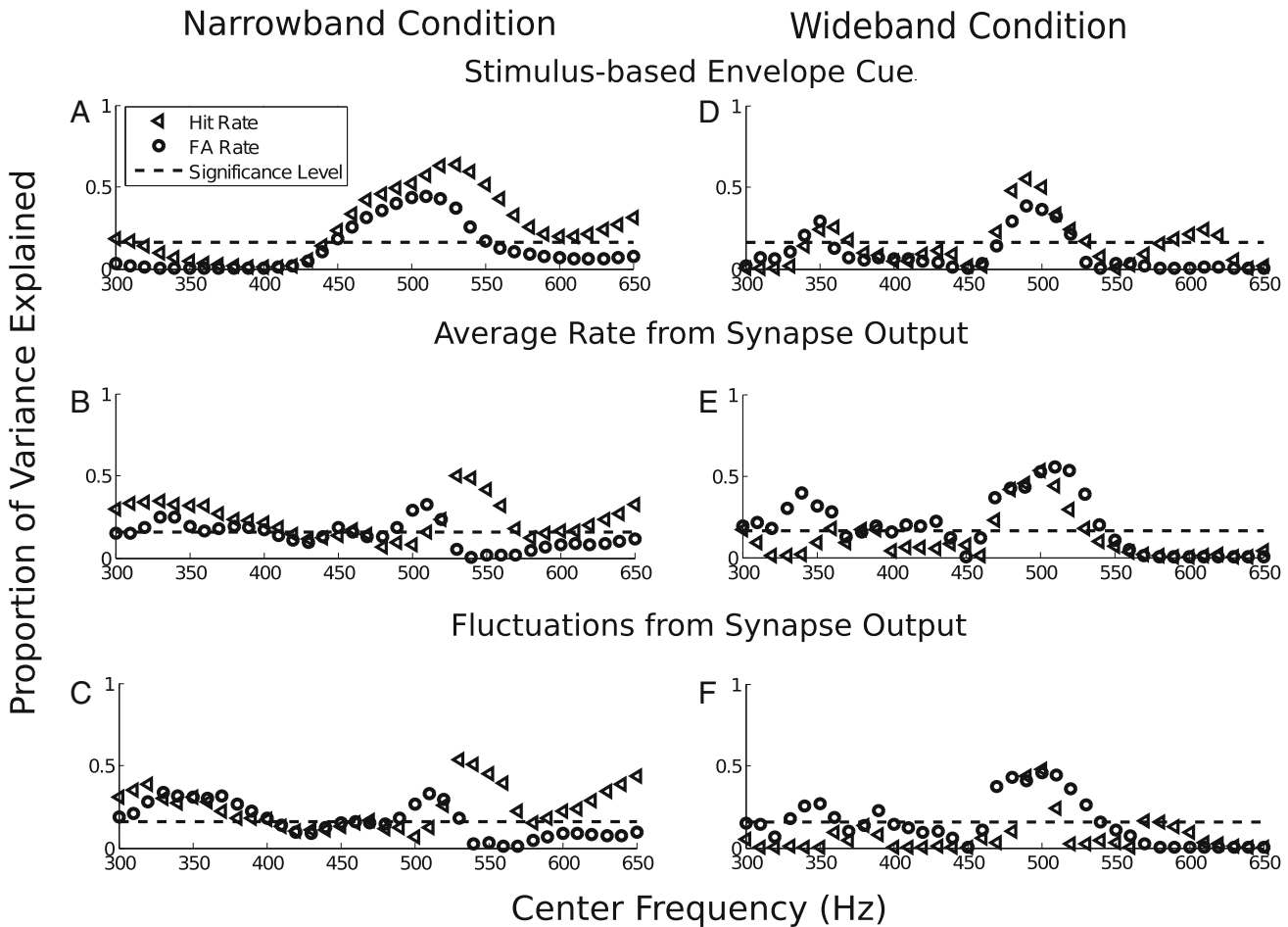


FIG. 5. Stimulus-based and physiological model predictions of the average listener's hit (*triangles*) and FA (*circles*) rates based on stimulus-based envelope cues (A narrowband, D wideband), average rate (B narrowband, E wideband), and fluctuations (C narrowband, F wide-

band) computed from the model IC cell responses. The *x-axis* shows the center frequencies of the model cells, and the *y-axis* shows the proportion of variance explained by the model. The *black dotted line* indicates the level required for significant predictions ($p < 0.05$).

the low correlations between listeners' detection patterns for this condition. In both narrowband and wideband conditions, model predictions were not symmetric around the matched center frequencies of 500 Hz. This is partly due to the fact that the physiological cues were obtained by combining positive contralateral and negative ipsilateral CN inputs; exchanging the contralateral and ipsilateral inputs yields slightly different results (i.e., they are not mirror symmetric) because of the differences across reproducible waveforms.

Predictions based on the model response fluctuations for the average listener in the wideband (Fig. 7(A, B)) and individual listeners in the narrowband condition [S1, Fig. 7(C, D); S3, (E, F); and S4, (G, H)] are shown in Figure 7. The overall trends in Figure 7 are similar to the results from Figure 6. Listeners' detection patterns were best predicted by combining frequency channels both tuned near 500 Hz in the wideband condition, whereas different combinations of frequency channels yielded better predictions of listeners' patterns in the narrowband condi-

tion. However, there are some detailed differences between the trends in Figures 6 and 7. In the wideband condition, for fluctuation cues, the best frequency channels are located closer to 500 Hz compared with rate cues, though predictions from 600-Hz contralateral and 400-Hz ipsilateral inputs also yielded good predictions. For the narrowband condition, model predictions for the dichotic conditions based on model response fluctuations yielded a substantially higher correlation for some listeners' patterns than predictions using the rate cue.

Comparisons of model predictions for the average listeners using stimulus-based and physiological cues are shown in Table 1. Table 2 shows predictions for individuals in the dichotic narrowband conditions, because the average listener was not used in this condition. Both hit and FA rates were predicted using the physiological envelope cues for both diotic and dichotic conditions. For the stimulus-based cues, hit and FA rates were predicted for the diotic condition, whereas only the hit rate for the dichotic condition could be predicted. As shown from the table, both types of physiological

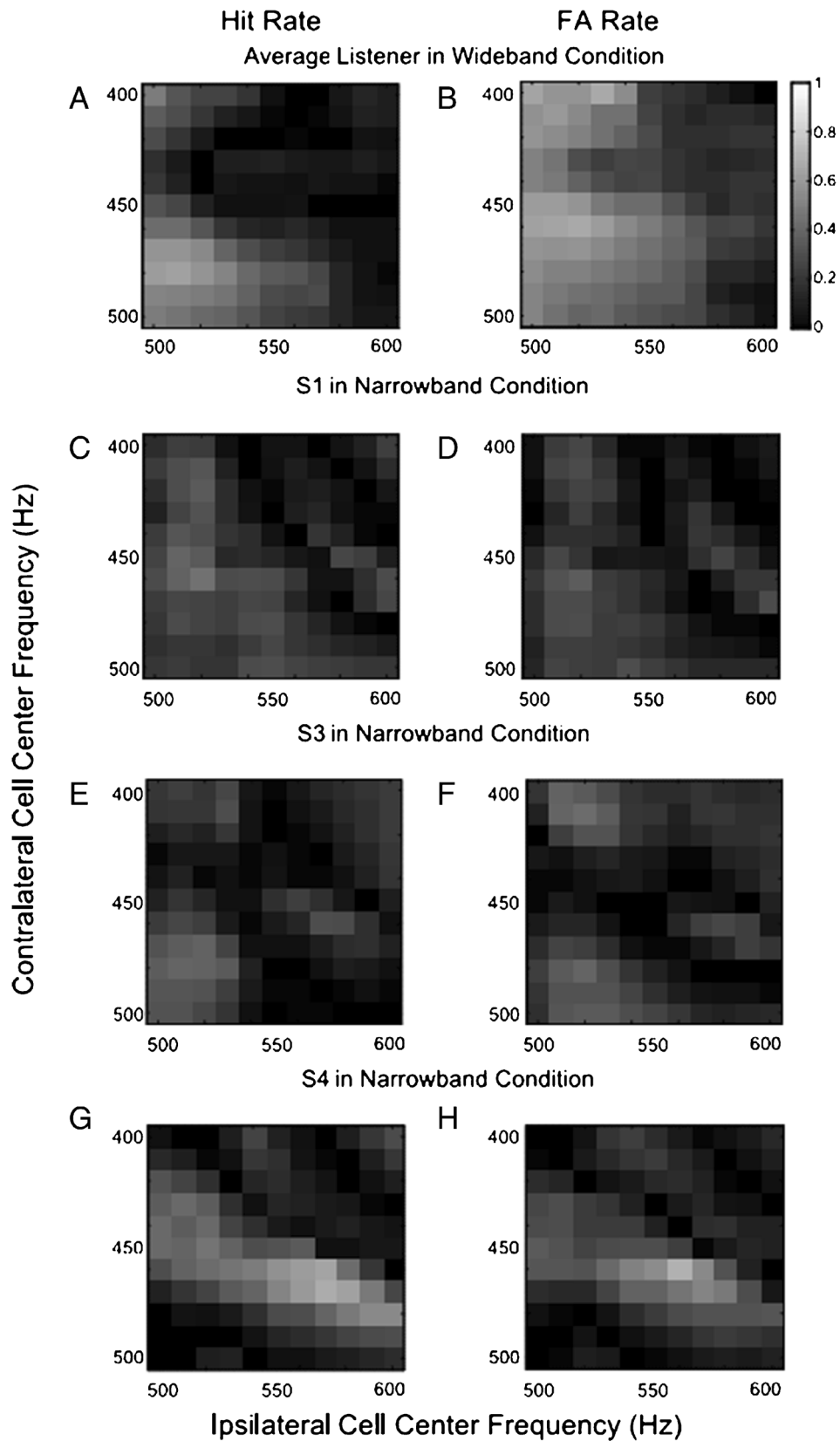


FIG. 6. Physiological model predictions of the average listener's wideband hit and FA rates (A, B) and individual listeners' narrowband hit and FA rates (S1: C, D; S3: E, F; and S4: G, H) based on the average rate of the model IC synapse output. The x-axis shows the

center frequencies of model cells receiving the stimuli presented to the *left ear*, and y-axis shows the difference of center frequencies of model cells at right ears with respect to the left ears. The *grayscale value* shows the proportion of variance explained by the model.

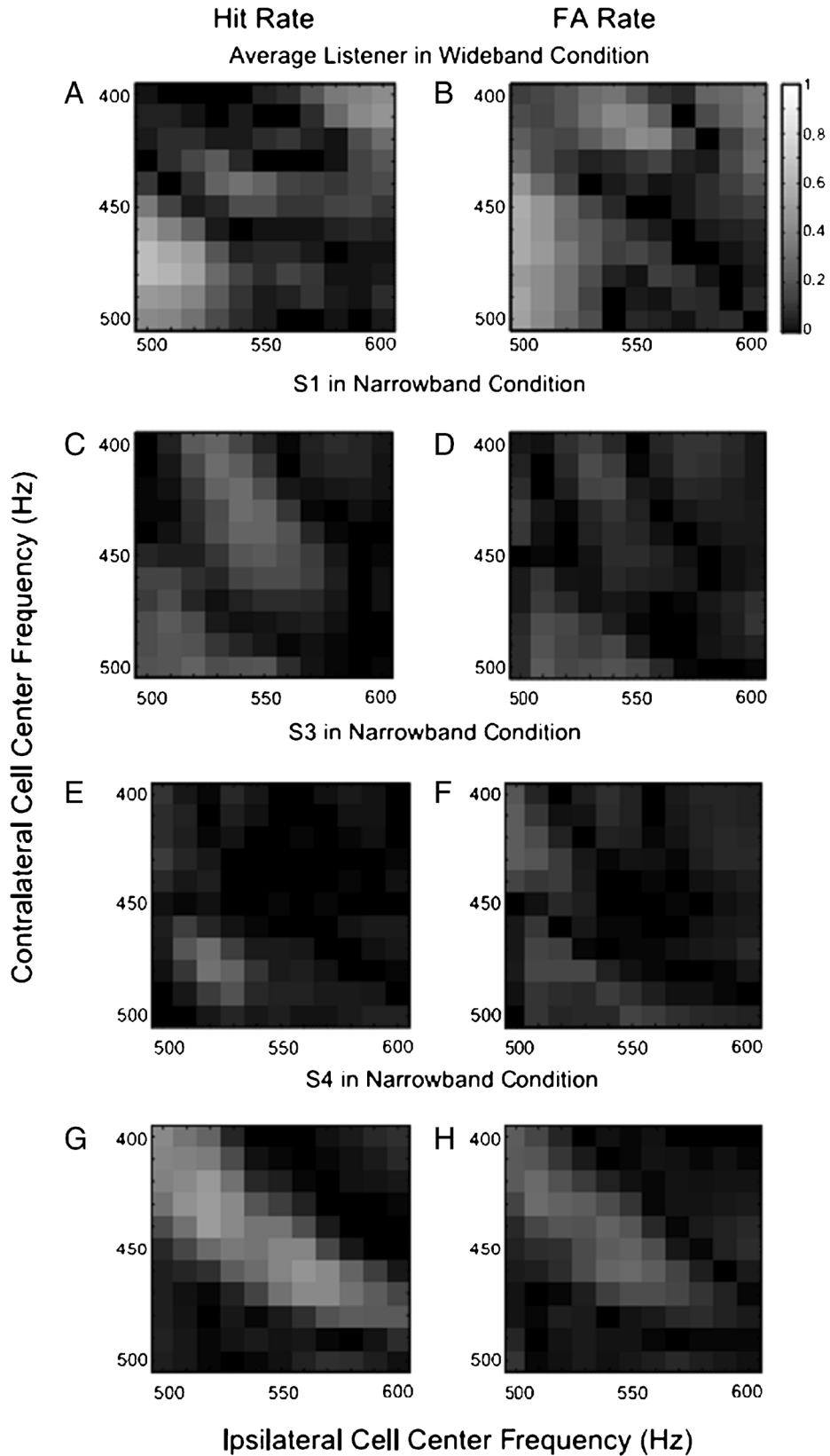


FIG. 7. Physiological model predictions of the average listener's wideband hit and FA rates (A, B) and individual listeners' narrowband hit and FA rates (S1: C, D; S3: E, F; and S4: G, H) based on fluctuations of the model IC synapse output. The *x*-axis shows the

center frequencies of model cells receiving the stimuli presented to the left ear, and *y*-axis shows the difference of center frequencies of model cells at right ears with respect to the left ears. The *grayscale value* shows the proportion of variance explained by the model.

envelope models predicted similar amounts of variance of the average listeners' detection patterns in the wideband condition, and these predictions were similar to the stimulus-based cues. For the narrowband dichotic condition, predictions from fluctuation cues were slightly, but not significantly better than rate cues for some listeners, and predictions for S3 and S4 from these physiological cues were significantly improved in comparison to the stimulus-based SIED cue. Predictions were comparable (S6) or slightly higher (S1 and S5) for three other listeners and lower (S2) for only one listener. The improved predictions of the physiological model are likely due to the modulation gain that is present in this model, but not in the stimulus-based models. Modulation gain, in either the synchrony or rate responses, is present at every level of the physiological model.

DISCUSSION

Physiological and stimulus-based models that rely on envelope cues were tested with listeners' detection patterns for diotic and dichotic reproducible noises. The physiological models included peripheral nonlinearities that distort envelope cues, such as cochlear compression and saturation in the responses of hair cells and auditory nerve synapses. These models also include mechanisms that distort, but enhance, the envelopes, including adaptation in the auditory nerve, dynamic range adaptation, and tuning for amplitude modulation rate at the level of the IC (Zilany et al. 2009; Zilany and Carney 2010; Nelson and Carney 2004). The physiological models explained comparable or greater amounts of variance in most listeners' detection patterns as compared to the stimulus-based models, for both diotic and dichotic conditions.

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

In the current study, physiological monaural and binaural envelope cues were analyzed, motivated by the robustness of stimulus-based envelope cues for predicting listeners' performance in both diotic and

dichotic conditions (Richards 1992; Davidson et al. 2006; Mao et al. 2013; Mao and Carney 2014). The physiological envelope cues studied here were based on the average rate and response fluctuations computed from a physiological model for the auditory pathway up to the level of the IC. The rate computation from the model IC responses can be interpreted as the response energy from the envelope-sensitive cell, and the fluctuations of the IC responses are similar to the envelope slope cues used in the stimulus-based models. In addition, the synchrony of IC responses to the 500-Hz tone and synchronized rate (the product of synchrony and rate) values were also computed, but the correlations between DVs based on these metrics and listeners' detection performance were not significant.

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

The results in Figures 6 and 7 indicate that for the dichotic wideband condition, for which listeners' patterns were highly correlated, similar frequency channels were used across listeners to detect tones in noise. In general, the frequency channels tuned near the target frequency of 500-Hz provided the best predictions for the wideband condition. In contrast, for the dichotic narrowband condition, different combinations of frequency channels yielded the highest correlations for different listeners' patterns. Similar results were observed when predicting dichotic detection patterns using the stimulus-based SIED cue with gammatone filters that had mismatched center frequencies (Mao and Carney 2014). Both results suggest that different listeners use different combinations of frequency channels for the dichotic narrowband condition.

Although predictions of listeners' detection patterns based on envelope cues explained a significant amount of listeners' performance in the current study, these predictions were lower than the predictable variance for the diotic condition (Mao et al. 2013). The predictable variance describes the proportion of the variation in detection patterns that is common

among all listeners and is used as a benchmark for model predictions. As shown in previous studies, predictions based on an optimal nonlinear combination of stimulus-based energy and temporal cues approach the predictable variance (Mao et al. 2013). In the current model, physiological envelope cues were analyzed from a single model IC neuron that received input from a single model AN channel. Assuming that physiological cues are Gaussian-distributed, it would be worth investigating model predictions based on an optimal combination of these cues across different frequency channels, using the likelihood ratio-based method as in Mao et al. (2013).

In this study, the hypothesis that envelope cues available in a physiological model were as reliable as stimulus-based envelope cues in predicting listeners' tone-in-noise detection patterns was tested. In conclusion, predictions from physiological cues were similar to stimulus-based cues in diotic wideband and narrowband and dichotic wideband conditions (Table 1). For the dichotic narrowband condition, in which different listeners apparently use different strategies, predictions from physiological cues explained comparable amounts, and in some cases significantly more, of the variance of listeners' detection patterns than the stimulus-based binaural envelope cue (Table 2). Based on these results, changes in physiological mechanisms that would affect the representation of envelopes, including sensorineural hearing loss and aging, would be expected to explain some of the deficits of these listeners in masked detection. It is interesting that these envelope-based cues may play the most important role for detection of low-frequency signals, such as the 500-Hz tone used here, given that fine structure is typically assumed to dominate coding of these signals.

ACKNOWLEDGMENTS

This work was supported NIH R01-DC010813.

Conflict of Interest The authors declare that they have no conflict of interest.

APPENDIX

The sixth-order band-pass filter (H) was computed by cascading three second-order band-pass filters (H_1 , H_2 , and H_3). The formula for each second-order band-pass filter (H_i) is

$$H_i(z) = \frac{1 - \alpha_i}{2} \frac{1 - z^{-2}}{1 - \beta_i(1 + \alpha_i)z^{-1} + \alpha_i z^{-2}} \quad (\text{A1})$$

where β is related to the center frequency, f_b of H_i by

$$\beta_i = \cos(2\pi f_i), \quad (\text{A2})$$

and α is related to the 3-dB bandwidth, W_b , by

$$\alpha_i = \frac{1 - \sin(2\pi W_i)}{\cos(2\pi W_i)}. \quad (\text{A3})$$

REFERENCES

- BLODGETT HC, JEFFRESS LA, TAYLOR RW (1958) Relation of masked threshold to signal-duration for interaural phase combination. *Am J Psychol* 71:283–290
- BLODGETT HC, JEFFRESS LA, WHITWORTH RH (1962) Effect of noise at one ear on the masked threshold for tone at the other. *J Acoust Soc Am* 34:979–981
- BREEBAART J, VAN DER PAR S, KOHLRAUSCH A (2001) Binaural processing model based on contralateral inhibition I. Model structure. *J Acoust Soc Am* 110:1074–1088
- CARNEY LH, HEINZ MG, EVILSIZER ME, GILKEY RH, COLBURN HS (2002) Auditory phase opponency: a temporal model for masked detection at low frequencies. *Acta Acust Acust* 88:334–347
- DAU T, PÜSCHEL D, KOHLRAUSCH A (1996) A quantitative model of the “effective” signal processing in the auditory system. I. Model structure. *J Acoust Soc Am* 99:3615–3622
- DAVIDSON SA, GILKEY RH, COLBURN HS, CARNEY LH (2006) Binaural detection with narrowband and wideband reproducible noise maskers. III. Monaural and diotic detection and model results. *J Acoust Soc Am* 119:2258–2275
- DAVIDSON SA, GILKEY RH, COLBURN HS, CARNEY LH (2009) An evaluation of models for diotic and dichotic detection in reproducible noises. *J Acoust Soc Am* 126:1906–1925
- DOLAN TR, ROBINSON DE (1967) Explanation of masking-level difference that result from interaural intensive disparities of noise. *J Acoust Soc Am* 42:977–981
- DURLACH NI (1963) Equalization and cancellation theory of binaural masking-level differences. *J Acoust Soc Am* 35:1206–1218
- EVILSIZER ME, GILKEY RH, MASON CR, COLBURN HS, CARNEY LH (2002) Binaural detection with narrowband and wideband reproducible maskers: I. Results for human. *J Acoust Soc Am* 111:333–345
- FLETCHER H (1940) Auditory patterns. *Rev Mod Phys* 12:47–65
- FRISINA RD, SMITH RL, CHAMBERLAIN SC (1990) Encoding of amplitude modulation in the gerbil cochlear nucleus: I. A hierarchy of enhancement. *Hear Res* 44:99–122
- GAI Y, CARNEY LH (2006) Temporal measures and neural strategies for detection of tones in noise based on responses in anteroventral cochlear nucleus. *J Neurophysiol* 96:2451–2464
- GAI Y, CARNEY LH (2008) Influence of inhibitory inputs on rate and timing of responses in the anteroventral cochlear nucleus. *J Neurophysiol* 99:1077–1095
- GILKEY RH, ROBINSON DE (1986) Models of auditory masking: a molecular psychophysical approach. *J Acoust Soc Am* 79:1499–1510
- GILKEY RH, ROBINSON DE, HANNA TE (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

- GOPELL MJ, HARTMANN WM (2007) Interaural fluctuations and detection of interaural incoherence. III. Narrowband experiments and binaural models. *J Acoust Soc Am* 122:1029–1045
- IBRAHIM RA, BRUCE IC (2010) Effects of peripheral tuning on the auditory nerve's representation of speech envelope and temporal fine structure cues. In: Lopez-Poveda EA, Palmer AR, Meddis R (eds) *The neurophysiological bases of auditory perception*. Springer, New York, pp 429–438
- ISABELLE SK (1995) Binaural detection performance using reproducible stimuli. Ph.D. dissertation, Dept of Biomedical Eng, Boston University, Boston, MA
- ISABELLE SK, COLBURN HS (1987) Effects of target phase in narrowband frozen noise detection data. *J Acoust Soc Am* 82:S109
- ISABELLE SK, COLBURN HS (1991) Detection of tones in reproducible narrow-band noise. *J Acoust Soc Am* 89:352–359
- ISABELLE SK, COLBURN HS (2004) Binaural detection of tones masked by reproducible noise: experiment and models. Report BU-HRC 04–01, Boston University, Boston, MA
- JORIS PX, SCHREINER CE, REES A (2004) Neural processing of amplitude-modulated sounds. *Physiol Rev* 84:541–577
- JORIS PX, VAN DE SANDE B, LOUAGE DH, VAN DER HEIJDEN M (2006) Binaural and cochlear disparities. *Proc Natl Acad Sci U S A* 103:12917–12922
- KIDD G JR, MASON CR, BRANTLEY MA, OWEN GA (1989) Roving-level tone-in-noise detection. *J Acoust Soc Am* 86:1310–1317
- KRISHNA BS, SEMPLE MN (2000) Auditory temporal processing: responses to sinusoidally amplitude-modulated tones in the inferior colliculus. *J Neurophysiol* 84:255–273
- MAO J, CARNEY LH (2014) Binaural detection with narrowband and wideband reproducible noise masker. IV. Models using interaural time, level, and envelope differences. *J Acoust Soc Am* 135:824–837. doi:10.1121/1.4861848
- MAO J, VOSOUGHI A, CARNEY LH (2013) Predictions of diotic tone-in-noise detection based on a nonlinear optimal combination of energy, envelope, and fine-structure cues. *J Acoust Soc Am* 134:396–406
- MOORE BCJ (2003) *An introduction to the psychology of hearing*, 5th edn. Academic Press, San Diego
- NELSON PC, CARNEY LH (2004) A phenomenological model of peripheral and central neural responses to amplitude-modulated tones. *J Acoust Soc Am* 116:2173–2186
- NELSON PC, CARNEY LH (2007) Neural rate and timing cues for detection and discrimination of amplitude-modulated tones in the awake rabbit inferior colliculus. *J Neurophysiol* 91:522–539
- PFÄFFLIN SM, MATHEWS MV (1966) Detection of auditory signals in reproducible noise. *J Acoust Soc Am* 39:340–345
- RICHARDS VM (1992) The detectability of a tone added to narrow bands of equal energy noise. *J Acoust Soc Am* 91:3424–3435
- SIEGEL RA, COLBURN HS (1983) Internal and external noise in binaural detection. *Hear Res* 11:117–123
- ZHANG X (2004) Cross-frequency coincidence detection in the processing of complex sounds. Ph.D. dissertation, Dept of Biomedical Eng, Boston University, Boston, MA
- ZILANY MSA, BRUCE IC (2006) Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery. *J Acoust Soc Am* 120:1446–1466
- ZILANY MSA, BRUCE IC (2007) Representation of the vowel /ε/ in normal and impaired auditory nerve fibers: model predictions of responses in cats. *J Acoust Soc Am* 122:402–417
- ZILANY MSA, BRUCE IC, NELSON PC, CARNEY LH (2009) A phenomenological model of the synapse between the inner hair cell and auditory nerve: long-term adaptation with power-law dynamics. *J Acoust Soc Am* 126:2390–2412
- ZILANY MSA, BRUCE IC, CARNEY LH (2014) Updated parameters and expanded simulation options for a model of the auditory periphery. *J Acoust Soc Am* 135:283–286
- ZILANY MSA, CARNEY LH (2010) Power-law dynamics in an auditory-nerve model can account for neural adaptation to sound-level statistics. *J Neurosci* 30:10380–10390