ACTA ACUSTICA UNITED WITH ACUSTICA
Vol. 88 (2002) 334 – 347

Auditory Phase Opponency: A Temporal Model for Masked Detection at Low Frequencies

Laurel H. Carney*, Michael G. Heinz†, Mary E. Evilsizer, Robert H. Gilkey‡, H. Steven Colburn
Boston University Hearing Research Center, Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts 02215, USA

Summary
We present a model for tone-in-noise detection at low frequencies that includes a physiologically realistic mechanism for processing the information in neural discharge times. The proposed model exploits the frequency-dependent phase properties of the tuned filters in the auditory periphery and uses cross-auditory-nerve-fiber coincidence detection to extract temporal cues. Information in the responses of model coincidence detectors is quantified and compared to human performance in a masked detection task. The responses of some cross-frequency coincidence detectors are reduced when a low-frequency tone is added to a noise because of phase differences between fibers tuned to different frequencies. We refer to this response reduction as "phase opponency." For super-critical masker bandwidths, the PO model succeeds in predicting detection of low-frequency tones in roving-level maskers, a psychophysical task for which the classical energy model fails. The PO model describes a physiologically realistic mechanism for extracting spatio-temporal information that can be applied to other sensory systems in which spatially overlapping and partially correlated temporal information is important.

PACS no. 43.64.Bt, 43.66.Ba

1. Introduction

Sensory systems encode information about objects in the presence of noise. The simple psychophysical task of detecting a pure tone in the presence of a random noise has been of fundamental importance in the development of auditory theory. Essentially all modern models of auditory processing include an initial array of narrowband filters, and for detection of a tone in noise, it is generally assumed that a filter tuned near the frequency of the tone is used. This model is based on Fletcher's [1] demonstration that the detectability of a tone is affected by the bandwidth of a masking noise only up to a certain value termed the "critical bandwidth"; addition of noise energy outside the critical band does not affect the detectability of a tone. Moreover, the performance of an energy detector operating on the output of a single critical-band filter tuned to the tone frequency (Figure 1a) depends on changes in signal-to-noise ratio in a manner similar to the performance of human listeners [2]. This critical-band energy-detector (CBE) model has served as the foundation of numerous empirical and computational investigations of audition.

Despite its prominence, the CBE model fails to predict several characteristics of detection (reviewed in [3, 4]). For example, listeners detecting a tone do not ignore masker information outside the critical band, whether that information is useful [5] or detrimental [6]. This evidence suggests that multi-band processing is important for tone-in-noise detection. Additionally, while the CBE model can explain the variation in human performance with signal-to-noise ratio, it cannot explain the finding that detection is only slightly disrupted when overall stimulus energy is randomized across intervals (e.g. [7, 8, 9]). In these studies, a roving-level masker was used; that is, the levels of both the target and the masker were randomly varied from interval to interval over a large range (while maintaining a fixed signal-to-noise ratio). If the listener based detection of the tone on the energy at the output of the auditory filter tuned to the frequency of the tone, performance would be considerably degraded by this stimulus manipulation, but this is not the case [7, 8, 9]. The CBE model also fails to explain performance when differences in stimulus energy across intervals are eliminated [10, 11]. The above results suggest that stimulus envelope and fine-structure timing, which typically co-vary with signal-to-noise ratio but not with overall level, provide important cues for tone-in-noise detection [12]. However, physiologically realistic mechanisms to extract this timing information have not

Received 10 September 2001, revised 13 December 2001, accepted 7 January 2002.

* Present address: Dept. of Bioengineering and Neuroscience, Institute for Sensory Research, 621 Skytop Rd., Syracuse University, Syracuse, NY

† Present address: Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, 21205, USA

‡ Present address: Department of Psychology, Wright State University, Dayton, OH, 45434, and Human Effectiveness Directorate, Air Force Research Laboratory, Wright-Patterson AFB, OH, USA

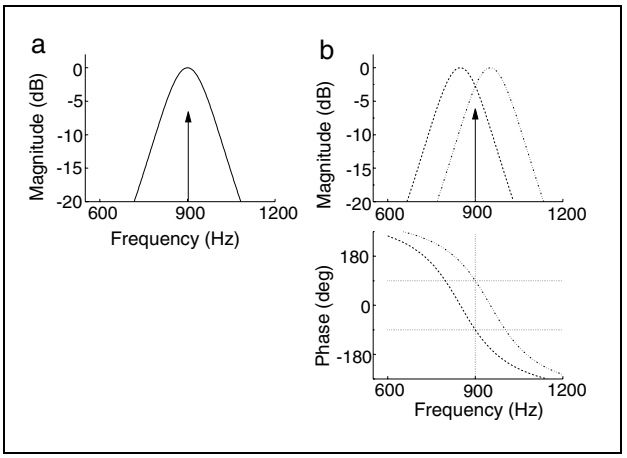

Figure 1. Filter characteristics important for a, the critical-band energy (CBE) model and for b, the phase-opponency (PO) model. a, A single filter centered at the frequency of a tone (arrow) is the fundamental signal-processing element of the CBE model. Only the magnitude is shown, since the phase spectrum is unimportant for the CBE model. b, Magnitude and phase spectra for two filters that illustrate the operation of the PO model. One filter is tuned below and one is tuned above the frequency of the tone. Fourth-order gammatone filters that have a 180° phase difference at the tone frequency are shown. The filters shown have CFs of 848.5 and 951.5 Hz and were used for the simulations in Figures 2 and 3.

been quantified and compared to human performance. In this report, we present the phase-opponency (PO) model, which includes a specific temporally based mechanism, and we demonstrate quantitatively that its properties are consistent with human performance for the detection of low-frequency tones in noise. This model applies particularly to detection of low-frequency tones because it depends upon the phase-locked responses of auditory-nerve (AN) fibers.

Physiological models of the CBE (also known as the power-spectrum model) typically depend upon the representation of stimulus energy in the discharge rates of AN fibers, but this representation is not easily implemented in ways that are consistent with available physiological data. Because the majority of AN fibers have a limited dynamic range, these models depend upon pooling information across low-threshold, high-spontaneous-rate AN fibers and high-threshold, wide-dynamic-range, low-spontaneous-rate fibers (e.g. [13, 14, 15, 16], reviewed in [17]). The strategy of pooling the information in AN discharge rates across fibers with different spontaneous rates may be relevant for high characteristic frequencies (CFs). However, the large dynamic range for low-spontaneous rate fibers depends upon the amount of cochlear compression, e.g. [18, 19]. Because the degree of cochlear nonlinearity is relatively small at low CFs, there are not large differences between the dynamic ranges of low- and high-spontaneous-rate fibers at low CFs (see discussion in [20, 21]. Fibers with straight (non-saturating) rate-level functions have not been observed at CFs below 1500 Hz in guinea pig [16] and are not observed at any CF in cat (e.g. [18, 22]). Although rate-level functions of AN fibers have not been measured in humans, psychophysical studies have shown that the amount of cochlear compression in humans is reduced at low frequencies [23], which is consistent with physiological observations in other species. Thus, it is reasonable to assume that low-frequency AN fibers in humans have discharge rates that saturate at mid to high levels, whether they are low or high spontaneous-rate fibers. Another strategy to overcome the limitations of the CBE model is to combine energy information across filters tuned to different frequencies (e.g. [24]). However, rate-based physiological models of this approach still require that the discharge rates of each channel encode energy, which is not the case at low frequencies due to saturation of AN fibers.

The PO model depends upon comparisons of the temporal response patterns across AN fibers tuned to different frequencies, rather than upon the energy of the filter outputs. The phase properties of filters tuned to different frequencies enable the filters to encode information about the stimulus waveform in the spatio-temporal patterns of a filter-bank response. Spatio-temporal patterns are the temporal discharge patterns across the population of neurons tuned to different frequencies. The spatial dimension is frequency and refers to the orderly tonotopic map of tuned neurons in most auditory nuclei. The general properties of the spatio-temporal response patterns are robust to changes in overall stimulus level and to non-linearities, such as saturation, that are inherent in neural responses [25, 26]. For the model considered here, the temporal response properties of the fibers are most important; there are not significant differences in the phase-locking properties of low- and high-spontaneous-rate fibers [27].

The PO model is expressed in physiological terms and uses a simple coincidence-detection mechanism to extract the reliable spatio-temporal cues embedded within the auditory-nerve (AN) population response. In particular, the PO model takes advantage of a reduction in the rate of some cross-frequency coincidence detectors in response to the presence of the target. This reduction is caused by the phase differences between fibers tuned to different frequencies (Figure 1b). This feature of the PO model, which is applied here to monaural masked detection, is similar to that used in models for binaural detection. The *reduction* in the response of binaural coincidence detectors to a tone that is out-of-phase to the two ears can explain binaural unmasking [28].

The addition of a tone to a noise systematically influences the relative timing across filters tuned to different frequencies. These spatio-temporal patterns can be processed by coincidence detectors that are sensitive to the relative timing of their inputs. We first consider a single coincidence detector that receives inputs from two filters tuned to frequencies above and below the tone frequency (Figure 1b). Because the two filters overlap in the frequency domain, their responses to a wideband noise are partially correlated, resulting in occasional coincidences in their outputs. This phenomenon is illustrated in Fig-

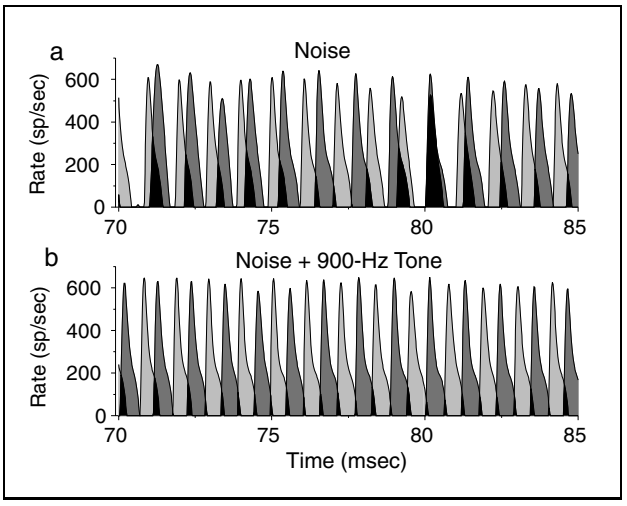

Figure 2. Responses of the two model AN fibers (with filters in Figure 1b) to a, wideband noise alone and to b, wideband noise plus a 900-Hz tone at 65 dB SPL. The noise spectral level was 35 dB SPL. The AN model was a linear fourth-order gammatone filter with equivalent rectangular bandwidth of 122 Hz, based on the auditory-filter bandwidth for humans at 900 Hz [29]. The filter was followed by models for the inner hair cell and AN synapse [21, 30]. The curves represent the time-varying AN discharge rate of the low-frequency (light gray) and high-frequency (dark gray) model AN fibers. The black areas represent the regions of overlap that would activate a model coincidence-detecting cell.

ure 2, in which the instantaneous firing rates of two model AN fibers are illustrated using two shades of gray; a coincidence-detecting cell that received these two AN inputs would be most likely to discharge when the inputs overlap (shaded in black). When a tone is added to the noise at a sufficient signal-to-noise ratio, the tone dominates the temporal responses and "pulls" the filter responses out of phase with each other (Figure 2b) because the phases of these two filters differ by 180° at the tone frequency (Figure 1b). The coincidence detector's response is *reduced* when its inputs have opponent phases and thus are least coincident with each other, even when the inputs are saturated. The model uses this reduction to detect the presence of the tone. The response properties required for the PO model are fundamental properties of any filter bank with overlapping filters having realistic phase properties, each followed by a saturating nonlinearity.

Because the PO model uses only relative timing across neurons, there is no need for the processor to have *a priori* knowledge of the tone or noise waveforms by the processor. As noted above, there are significant difficulties inherent in AN rate- or count-based models due to the rate-saturation observed in most AN fibers. In contrast, temporal information is robust over a wide range of levels and in background noise, because AN fibers phase-lock to low-frequency tones over a wide dynamic range even in the presence of noise [31, 32, 33].

Several models have been proposed for processing temporal information in the responses of the auditory periphery. Previous temporal processing models fall into three general classes: auto-correlation, cross-correlation, and lateral inhibition. Auto-correlation models were developed to extract the periodicities, or stimulus-related inter-spike intervals, in responses of phase-locked neurons and have been discussed primarily in the context of pitch processing (e.g. [34, 35, 36, 37]). Models for processing based on spike intervals have also been investigated for speech and a variety of other complex sounds (e.g. [38, 39, 40]). Limitations of the auto-correlation-based models include the lack of anatomical and physiological evidence for the required delay lines, lack of physiological evidence for the cells that are performing the auto-correlation, and the absence of certain psychophysical phenomena related to multi-tonal stimuli that would be predicted by auto-correlation-based models (reviewed in [41]). These limitations motivate the study of cross-correlation models for temporal processing.

Several cross-correlation models have been proposed that take advantage of the combination of phase-locked responses and systematic frequency-dependent delays along the cochlea that are associated with the traveling wave (e.g. [41], in the context of pitch perception, and [42], in the context of localization). These models proposed that cross-frequency correlation takes place in the medial superior olive, where cells had been previously demonstrated to be sensitive to small interaural time differences between their inputs (e.g. [43]). Deng and Geisler [44] proposed cross-channel correlation on the output of a composite auditory-nerve model as a processor for detecting formants in speech sounds in quiet and in noise. Their results showed a reduction in the cross-correlation between neighboring frequency channels that were tuned near spectral peaks due to phase differences between the channels, a phenomenon that is closely related to the PO model explored here. Deng and Geisler [44] described the similarities between their cross-correlation mechanism and lateral inhibition mechanisms, which will be discussed further below.

The PO model includes coincidence-detecting cells that receive inputs from AN fibers. Physiological studies have demonstrated that low-frequency cells in the anteroventral cochlear nucleus (AVCN) that receive convergent inputs from AN fibers have response properties that are consistent with cross-frequency coincidence detection [45], which is a form of cross-correlation [46]. Evidence for cross-frequency coincidence detection in low-frequency cells in the AVCN was reported to be strongest for highly phase-locked, or "hi-sync", cells, for primarylike-with-notch cells, and for onset cells, as well as for some chopper cells [45]. Cochlear nucleus primarylike-with-notch [47] and onset [48] cells at high frequencies have also been described as having responses consistent with cross-frequency coincidence detection. Primarylike-with-notch response types are associated with globular bushy cells [49] and also with Type II responses in the AVCN slice preparation [50]. Further support for these cells behaving as coincidence detectors comes from the fact the nonlinear membrane properties of Type II AVCN neurons [51] are

similar to those of cells in the medial superior olive [52], which have been demonstrated to be sensitive to coincidences of their inputs [43, 52].

In this report, the properties of the cross-frequency coincidence-detection mechanism and its potential role in masked detection are explored. After illustrating the nature of the PO mechanism in the context of a single coincidence detector, the responses of populations of model coincidence detectors that receive different combinations of AN inputs will be studied. In both cases, we will quantify the ability of the proposed mechanism to detect low-frequency tones in the presence of noise and directly compare the results to human performance. In particular, we will explore the PO mechanism in the context of a roving-level paradigm, which has been used in psychophysical studies to challenge the classical CBE model for detection. In this paradigm, the signal-to-noise ratio is tracked in a two-alternative, two-interval forced-choice (2A2IFC) task while the noise level is randomly varied across intervals.

2. Methods

The methods included generation of stimuli based on those used in psychophysical studies of masked detection, computation of the responses of model AN fibers and coincidence-detecting cells to these stimuli, and simulation of psychophysical tasks based on decision variables that were computed for each interval of the stimulus. Each of these steps will be described in this section.

2.1. Stimuli

The noise stimuli were matched to those used in Kidd *et al.*'s [7] study of masked detection of a 900-Hz tone. Specifically, the noise had a spectral level that was either fixed at 35 dB SPL or randomly varied from interval to interval over a 32-dB range centered at 35 dB SPL. Several different noise bandwidths were studied as in Kidd *et al.* [6]; each bandwidth was geometrically centered about the tone frequency, which was always 900 Hz. The tone and noise durations were 250 ms, with 20-ms rise/fall times.

2.2. Computation of Responses of Model AN Fibers

The model AN responses for the results presented here were all produced using a recently developed model for the auditory periphery in human [21]. This model is an extension of a nonlinear model for the responses of AN fibers in cat [30], modified to have bandwidths for human listeners [29] as well as fibers with different spontaneous rates and thresholds. The nonlinear model with compression and suppression (Model #1 in [21]) for high-spontaneous-rate fibers was used for all simulations presented here. Qualitatively similar results were obtained using linear gammatone filters followed by models for the inner-hair-cell (IHC) transduction and IHC-AN synapse (not shown). The output of the AN model is an estimate of the instantaneous firing rate as a function of time during the stimulus. The steady-state portion of the responses were used for all results presented here.

2.3. Computation of Responses of Model Coincidence-Detecting Cells

Two models involving coincidence-detecting cells will be considered here: a *single-channel* model and a *multi-channel* model. In both cases, each model coincidence detector received two AN inputs, for simplicity. In the case of the *single-channel model*, which will be illustrated first, the coincidence detector received one AN input with CF below the tone frequency and one AN input with CF above the tone frequency. These two AN CFs were chosen such that the phases of their fourth-order gammatone-filter transfer functions (with human bandwidths appropriate for the noise levels used) differed by approximately 180° at the frequency of the stimulus tone and at the mean level of the stimulus ensemble[1]; e.g., CFs of 848.5 Hz and 951.5 Hz were chosen for the 900-Hz stimulus tone. This choice of input CFs yielded a coincidence cell with maximal phase opponency.

The multi-channel model considers the responses of a population of coincidence detectors, each receiving a pair of convergent AN inputs. Given a set of AN fibers tuned to a range of CFs, coincidence-detecting cells including all possible pairwise combinations of the set of AN fibers were included in the multi-channel model. The multi-channel-model simulations presented here were all based on a set of 27 model AN fibers, with logarithmically spaced CFs ranging from 625 Hz to 1295 Hz; the range of CFs was chosen to be geometrically centered on the tone frequency, to encompass the range of CFs that had the greatest impact on the detection task, and to include the coincidence-detecting cell used for the single-channel model. The central (14th) fiber in the population was tuned to the frequency of the 900-Hz tone.

Based on the responses of the model coincidence-detecting cells, a decision variable was computed to allow simulation of the masked-detection task. This decision variable will be described here for both the single- and multi-channel models. We investigated both the case for which only the external noise due to the variability of the masker stimuli was included in the decision variable and the case for which internal noise due to the variability of neural discharge counts was also included. The results with no internal noise represent a lower bound for thresholds that could be obtained with a very large number of identical model cells, because the combination of responses across independent cells reduces the effect of the internal noise. After describing the calculation of the

[1] Because the AN model fibers included nonlinear peripheral filters with level-dependent bandwidth, the AN response phases varied with sound level. However, the phase-opponency mechanism was robust to these changes for wideband noise maskers, even over the 32-dB range of levels included in the roving-level paradigm. That is, the AN responses do not have to be perfectly 180° out of phase with each other to result in a reduction of discharge count of the coincidence detecting cell.

decision variable for each of these cases, the use of the decision variables in a tracking algorithm to simulate psychophysical results will be described. The decision variables used here were all based on the statistics of the discharge counts of model coincidence-detecting cells, which were computed using the theory developed for binaural cross-correlators [53]. Details of this theory were presented in a recent study of monaural cross-frequency coincidence detection related to intensity discrimination for tones in quiet [20]. Briefly, the instantaneous firing rate of the coincidence detector is proportional to the product of the instantaneous firing rates of the model AN fibers. The average rate of a model coincidence cell depends upon the size of the coincidence window; the size of the window must be small with respect to the period of a tonal input for this simple description of the coincidence cell's response to be appropriate [20, 53].

The expected value of the discharge counts given the stimulus waveform was computed for each stimulus interval. This value varies across noise samples (i.e. the "external noise") but does not include the variability associated with the random nature of the neural discharge patterns (i.e. the "internal noise"). For all of the results presented here, 20 independent Gaussian noise samples were used to compute the means and variances of the discharge counts that were used to derive the model decision variables. Thus, simulations that used this conditional expected value of the discharge counts as a decision variable provide an estimate of the limitations of performance based on the stimulus variability alone. Additional simulations that included the variability of neural discharge counts from trial to trial provide an estimate of thresholds based on a limited number of neurons with realistically variable discharge counts.

For the simulations that included internal noise, the decision variable was a random deviate generated from a probability distribution chosen to match the observed distributions of spike counts from AN fibers and AVCN cells. The simplest description of the spike counts of auditory neurons would be a Poisson distribution; however, this description has been shown to significantly over-estimate the variance of the counts (e.g. [14, 16]). For the results shown here, the distribution of the coincidence-detecting cells was assumed to be Gaussian with a variance calculated from the expected value using a formula derived from empirical AN data by Winter and Palmer [16][2]. This assumption was based on the fact that many AVCN cells (e.g., primarylike and transient chopper response types) have count variance similar to that of AN fibers, and other AVCN cells types (e.g., sustained choppers) have even lower variances [54].

To use this count-dependent description of the variance, it was important that the model counts were similar to those of relevant auditory neurons. The simple, two-input coincidence detector model with a small coincidence window resulted in very low average discharge rates. To achieve average discharge rates more consistent with those typical of auditory neurons, 10 identically driven inputs from each AN CF converged on each model coincidence-detecting cell in both the single-channel and multi-channel models. The 10 identically driven AN inputs were assumed to be statistically independent, consistent with the analysis of AN data by Johnson and Kiang [55]. In addition, for the simulations that included internal noise, 10 identical coincidence-detecting cells were simulated and the decision variable used for the task was an equally weighted combination of the 10 cells. The combination of the responses of the 10 cells reduced the effect of the internal noise to provide threshold estimates that were comparable to human thresholds. The term single-channel model will be used to refer to the model consisting of 10 identical coincidence-detecting cells, each of which receives 10 inputs from each of two AN CFs. The choice of 10 cells for this "unit" coincidence-detection model was based on the comparison of preliminary estimates of threshold for the 10-cell version of the single-channel model to human thresholds for detection of a tone in a 3 kHz bandwidth, 35 dB SPL spectrum-level fixed-level masker [7]. The multi-channel model is comprised of a population of these single-channel models.

The 20-μs coincidence-window duration was matched to the temporal resolution of the stimulus waveforms used in the simulations, which was convenient for computations; this window size was also comparable to the 10-μs coincidence window used in Colburn's binaural model [53]. The choice of the coincidence-window duration influences the estimated threshold in a complex manner; preliminary investigations of window sizes both smaller and larger than the 20-μs window used here were performed, and it was verified that threshold was not a strong function of window size at these durations.

For the simulated thresholds of the single-channel model, either the expected value of count (no internal noise case) or a random deviate based on the set of 10 identical cells (case with both external and internal noise) was compared across stimulus intervals; the interval with the lower count was chosen as the interval with the tone. In this case, a particular cell was chosen that had AN inputs with CFs that were optimal for the PO mechanism.

For the multi-channel model, the information in the responses of cells with different input CFs was combined in order to derive a decision variable. Assuming that the response counts could be described by Gaussian distributions, the optimal combination scheme for a population of statistically independent cells is a weighted sum of the counts, where the weight for each cell is the difference in mean counts across stimulus intervals divided by the variance of the count [56]. The weights are computed as

$$W_k = \frac{\overline{C}_{T+N} - \overline{C}_N}{(V_{T+N} + V_N)/2},$$

[2] Winter and Palmer's [16] expression for the standard deviation, y, of count as a function of count, x, is $y = 0.857x^{0.343}$.

where $\overline{C}$ represents the expected value of the count for either the tone-plus-noise or noise-alone interval, and V represents the variance. Each variance is computed, based on the assumption of Poisson discharge statistics, as $V = \overline{C} + \mathrm{var}(C)$, where the first term represents the internal variance due to Poisson variability (equal to the mean count across the noise samples used) and the second term represents the external variance in the counts across the set of noise samples. The simulations of the multi-channel model shown here were either based on the expected values of counts or on simulated counts computed in the same way as described above for the single-channel model; each coincidence-detecting cell in the population was represented by a set of 10 identical and statistically independent cells, similar to the single-channel model simulations that included internal noise.

For the case of a wideband stimulus, the responses of a population of model coincidence cells that receive different AN inputs that have overlapping filters would not be statistically independent. In this situation, the weighted sum described above is non-optimal but is still a reasonable scheme for deriving a single decision variable from the population response. The optimal scheme would require use of the cross-covariance matrix [56], which is an avoidable complication. Results presented here for the multi-channel model are all based on the non-optimal weighted-sum decision variable described above.

The multi-channel model weights for each stimulus condition were derived from responses to 20 independent noise samples using a tone level 6 dB higher than the single-channel threshold with internal noise. This choice of tone level was based on examination of the patterns of weights across the population (see below). At threshold, the patterns were relatively noisy, as expected; the 6-dB increase in tone level above threshold resulted in a relatively clean set of weights that was qualitatively consistent with the weight pattern at lower levels.

2.4. Simulation of Psychophysical Tasks

To determine the masked threshold for tones in either the fixed-level or roving-level noise, the two-interval, two-alternative, forced-choice (2IAIFC) tracking paradigm used by Kidd *et al.* [7] was simulated. Model threshold estimates were based on simulations of a two-down, one-up track with tone level varied in 2-dB steps, with the levels at the last 12 reversals of a 16-reversal track averaged to estimate the threshold for each track. Forty-two tracks were simulated to match the number of estimates in the human study [7] (6 repetitions $\times$ 7 subjects). Independent samples of Gaussian noise were generated for each stimulus interval of the simulated tracks. For the single-channel model without internal noise, the expected value of the count for the coincidence detector was used as the decision variable for the PO model to determine which stimulus interval contained the tone. For the single-channel model with internal noise, the decision variable was based on the average of 10 random deviates representing identical individual cells. For the multi-channel model, the weighted sum of either the expected values of count or of the simulated counts was used as the decision variable.

The use of simulated tracks to estimate the model detection threshold provided predictions that could be directly compared to experimental results (e.g. [57, 58, 59]). In addition, the use of tracking, instead of directly estimating a value for d', eliminated the difficulty of calculating the variance of the population response for cells that lacked independence due to the noise masker (see [60]). Also, the use of the tracking paradigm to estimate thresholds allowed the inclusion of random level variation across intervals in the same manner as in the experimental paradigms without requiring an extension of the analytical approach (e.g., see [61]).

3. Results

3.1. Single-channel model

Simulations were performed to estimate the threshold of the model based on a coincidence detector that received inputs from two AN CFs, as illustrated in Figure 1b (Table I). The threshold for the single-channel model without internal noise for a 900-Hz tone in a fixed-level noise with a 3-kHz bandwidth and a 35-dB-SPL spectral-level (N_0) was 17.83 dB re N_0. This threshold is comparable to the threshold reported for human listeners of 17.04 dB re N_0 [7]. The threshold for the same stimulus conditions for the model with internal noise was 19.36 dB re N_0. As expected, this threshold estimate, based on 10 identical statistically independent cells, is elevated with respect to the model with no internal noise[3]. When the simulations for these models were repeated with the masker level randomly roved over a 32-dB range, the threshold for the model with no internal noise increased to 19.82 dB re N_0, which is comparable to the human threshold for this paradigm of 18.96 dB re N_0 [7]. The roving-level threshold for the model with internal noise was 20.36 dB re N_0, representing only a 1-dB increase in threshold with respect to the non-roving condition. As has been reported in previous studies, the CBE model cannot explain the similar thresholds for the roving and fixed-level paradigms (e.g. [7, 8, 9]. For example, the CBE model would predict an increase in threshold of approximately 7.5 dB, assuming the energy in the 100-Hz band centered on 900-Hz was used as the decision variable [7].

These simulation results (Table I) indicate that estimates of threshold based on a single coincidence detector channel with well-chosen inputs can detect tones in the presence of fixed-level or roving-level wideband maskers at levels comparable to those just detectable by human listeners. The coincidence-detecting cell in this model was chosen to have AN inputs tuned to frequencies that maximized the phase opponency between the two inputs, resulting in sensitive detection of the tone. It is more realis-

[3] Note that the use of ten identical cells only affects the simulated threshold in the case with internal noise.

Table I. Thresholds for detection of 900-Hz tone in 3 kHz bandwidth noise. Fixed-level results are for a 35-dB SPL spectrum level noise (N_0 = 35 dB SPL); roving level results are for a 32-dB rove of masker level centered about 35 dB SPL. Mean threshold +/- standard deviation in dB re N_0 is shown for each case. For the human listeners [7] and for each model, means were based on 42 estimates of threshold. Models with no internal noise were based on expected values of coincidence-cell counts; models with internal noise were based on the responses of 10 identical cells for each coincidence-detecting cell in the population.

	Human Listeners [7]	Single-channel model	Multi-channel model
Fixed-level masker	17.04 +/- ∼2	Without internal noise: 17.83 ± 1.98, with internal noise: 19.36 ± 2.00	Without internal noise: 18.37 ± 2.33, with internal noise: 18.03 ± 2.13.
Roving-level masker	18.96 +/- ∼2 dB	Without internal noise: 19.82 ± 1.73, with internal noise: 20.36 ± 2.20	Without internal noise: 17.86 ± 1.84, with internal noise: 18.06 ± 2.11

tic to consider a population of cells that receive convergent input from an array of AN fibers tuned across a range of frequencies, as shown in the next section.

3.2. Multi-channel model

A population of model coincidence-detector cells was used in the following multi-channel model simulations. Each cell received two AN inputs; all possible pair-wise combinations of the set of 27 model AN fibers were included in the population of coincidence-detector cells, including cells with inputs that had matched CFs. All possible combinations of these fibers were included in the multi-channel model so as not to make any assumptions about which combinations of fibers would contribute to the detection task. The contributions of each coincidence-detecting cell were weighted in the decision variable and are illustrated below. As described above, the multi-channel model simulations were done with and without including the effects of internal noise. When internal noise was included, each combination of AN CFs was represented by 10 statistically identical and independent model coincidence-detecting cells. The weighted sum of the counts was used as the decision variable in determining simulated thresholds with the multi-channel model (see Methods). Using this strategy, the threshold for the multi-channel model without internal noise for detection of a 900-Hz tone in a 3-kHz wideband noise with a fixed spectral level of 35 dB SPL was 18.37 dB re N_0 (Table I). The multi-channel model with internal noise had a threshold of 18.03 dB re N_0. These model thresholds are comparable to the human threshold of 17.04 dB re N_0 for this stimulus [7], and to the single-channel model threshold (Table I). Simulations for the multi-channel model with no internal noise for a masker level that roved over a 32-dB range centered at 35 dB SPL resulted in a threshold of 17.86 dB re N_0. The roving-level threshold for the multi-channel model with internal noise was 18.06 dB re N_0, which is comparable to the human threshold of 18.96 dB re N_0 [7]. Note that for both the fixed- and roving-level paradigms, simulations based on a small set of coincidence-detecting model cells, with inputs chosen to yield maximal phase opponency, provided a reasonable estimate of the threshold based on the population response (Table I). For the single-channel and multi-channel models, both absolute threshold and the robustness of the thresholds in the presence of roving-level wideband maskers are well explained by the PO model.

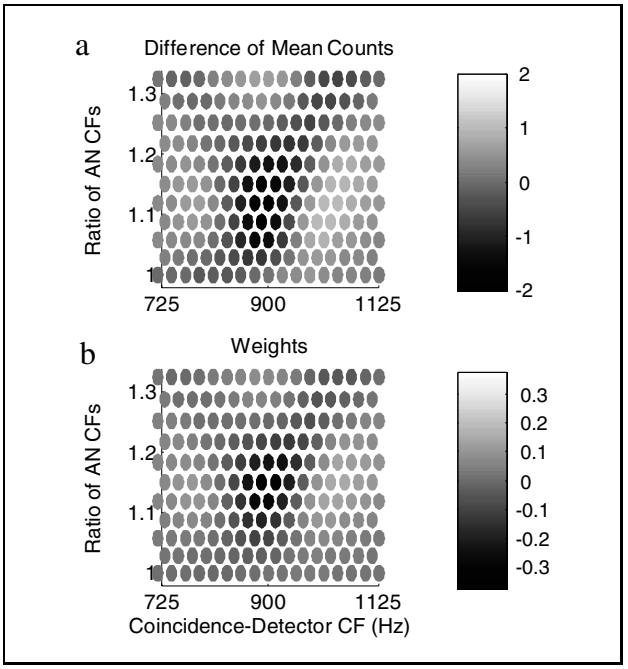

Figure 3. a) Differences in mean counts between the stimulus interval with a tone plus noise and the interval with noise alone for a population of coincidence-detecting cells. Dark shades represent a reduction in count upon addition of the tone to the masker, and light shades represent an increase in count upon addition of the tone to the masker. b) Weights for the population of coincidence-detecting cells (see Methods). Dark-shaded weights are negative, indicating cells that have reduced counts upon addition of the tone, and light-shaded weights are positive, indicating cells that have increased rates upon addition of the tone. These weights were derived from the same simulations as in Figure 3a; the weights include normalization by the variance of the difference in mean counts across intervals and thus provide an indication of the statistical significance of the changes in the count differences shown in Figure 3a.

Figure 3a illustrates the differences in mean counts across intervals used for the multi-channel model predictions above; the counts are shown for a tone level 6 dB above the single-channel model threshold (as described above). In this plot, each circle represents a model coincidence-detecting cell, with the shade of the circle representing the difference in mean counts for a cell between

the stimulus interval with the tone plus masker and the interval with masker alone (Figure 3a). The abscissa represents the CF of the model cell, which was computed as the geometric mean of the two AN CFs that provided the inputs to each coincidence-detecting cell. The ordinate represents the ratio between the two input CFs[4]. A subset of the total population of coincidence-detecting cells, centered about the tone frequency of 900 Hz, is illustrated in the plot. This plot illustrates the differences in mean counts across intervals based on an average across 20 independent samples of Gaussian noise.

Dark circles indicate cells that decreased their mean count when the tone was added; light colors indicate cells that increased their mean count when the tone was added to the masker noise. In this plot, it is clear that a number of coincidence-detector cells receiving a range of AN CFs are influenced by the PO mechanism. Thus, the careful selection of the single optimal cell (as was done above) is not a critical factor for this model because perfect phase opponency is not required to yield significant reductions in mean count upon addition of the tone to the masker. Also, the range of AN CFs in Figure 3a was selected to illustrate results for a population of cells tuned to frequencies near the 900-Hz tone; however, this plot of the differences in mean counts shows that the PO mechanism would allow identification of a tone with unknown frequency added to a masker, based on the location of cells with depressed rates within the tonotopic population.

The weights used to compute the multi-channel model decision variable that were derived from the responses in Figure 3a are illustrated in Figure 3b. Each model cell's weight was computed as the difference in mean counts across the intervals (as shown in Figure 3a) divided by the variance of the counts for that cell (see Methods). The most striking feature in this illustration of the weights is the dark region indicating negative weights (i.e. the decision variable increases when the output of these cells decreases) for cells that are influenced by phase opponency (AN CF ratio = 1.15). This region is centered on cells that receive one AN input tuned below the tone frequency and one tuned above it. Another important feature of the weight pattern is that the cells lying along the lowest row in the population, which receive matched-CF inputs (AN CF ratio = 1), have weights near zero. Because the discharge rates of the model AN fibers are saturated by the noise masker, these cells do not convey statistically significant information about the addition of the tone in their discharge rates. Cells that receive inputs tuned to *different* CFs can be sensitive to the addition of the tone because of the *relative* changes in the temporal patterns of their inputs.

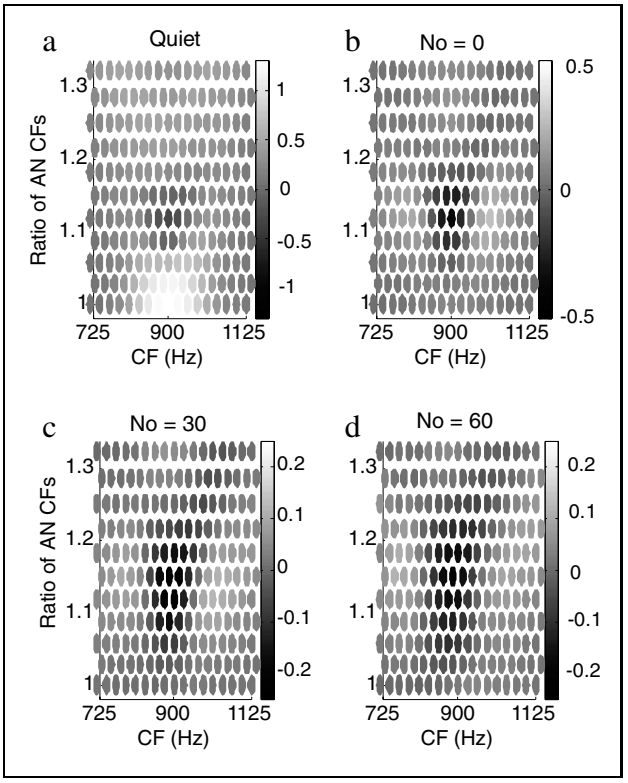

Figure 4. Multi-channel model weight patterns for tones in quiet (a) and for three different masker levels,b) 0, c) 30, and d) 60 dB SPL spectral level. Tone levels were set to 6 dB SPL in quiet and to 24 dB re N_0 for the three maskers. Note that the grayscale axis values vary across plots.

Changes in the response variance across the population of coincidence cells also influence the weight pattern. As a result, the positive weights of some of the cells that have increased response counts upon addition of the tone (e.g., model cells on the right-hand side of Figure 3b) are relatively low compared to the negative weights in the phase-opponent (dark) region. When thresholds were estimated for the multi-channel model, all cells were included, regardless of the sign or magnitude of their weights. However, for the wideband masker at 35 dB No (Figure 3b), the reduction in the responses of the phase-opponent cells was a more statistically reliable feature than was the increase in rate of the model cells with positive weights.

The details of the weights are influenced by several factors, including noise level, bandwidth, and tone level. The stimulus level influences the bandwidth of the nonlinear AN model, and changes in bandwidth are associated with changes in the phase properties of AN fibers (e.g. [20, 62, 63]). Thus, the locus of the strongest phase-opponent cells in the population shifts with changes in stimulus level (Figure 4). At low sound levels, the PO cells receive inputs with more similar CFs (AN CF ratio closer to 1), and at high sound levels, the locus shifts upwards toward cells that receive AN inputs with more widely separated CFs. In response to higher sound levels, for which the bandwidths of AN fibers are broad and phase functions

[4] The CFs of the two AN inputs to each coincidence-detecting model cell can be computed as follows:

$$\mathrm{AN_{CF1}} = \mathrm{CD_{CF}}/\sqrt{\mathrm{R}}, \text{ and } \mathrm{AN_{CF2}} = \mathrm{CD_{CF}}\mathrm{x}\sqrt{\mathrm{R}},$$

where R is the ratio the two ANCFs, $R = \mathrm{AN_{CF1}}/\mathrm{AN_{CF2}}$, and $\mathrm{CD_{CF}}$, the CF of the coincidence-detecting cell, is computed as the geometric mean of the two input CFs.

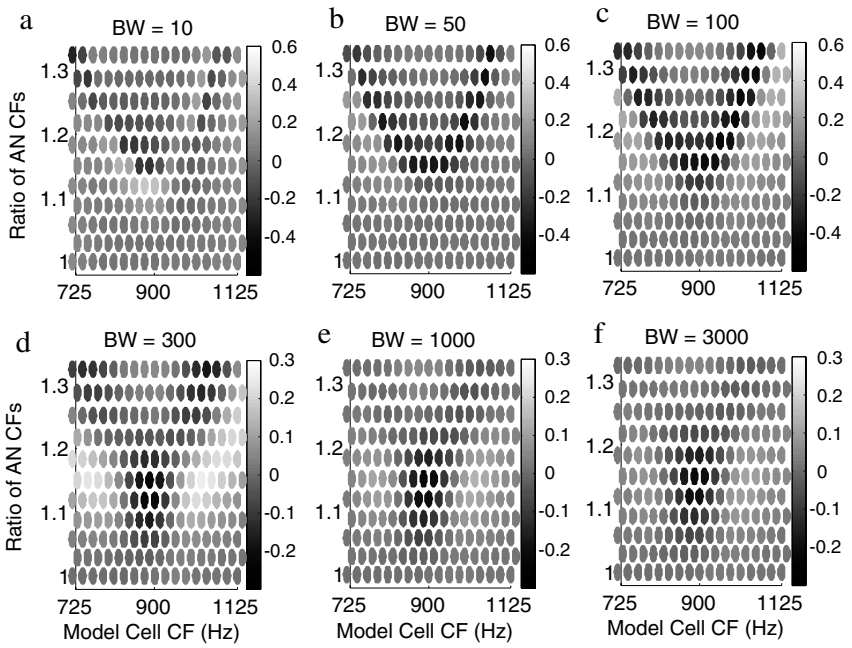

Figure 5. Multi-channel model weight patterns for six different masker bandwidths (a-f). The spectrum level of the masker was 35 dB SPL in all cases, and tone level was set to approximately 6 dB above threshold (19, 21, 23, 24, 24, 24 dB re N_0, for BWs 10 through 3000 Hz, respectively). Note that the grayscale axis values vary across plots.

are shallow, more distant CFs must be combined to yield the strongest phase opponency.

The masker bandwidth also influences the weight pattern. The patterns shown in Figure 5 are for several masker bandwidths, with spectral level fixed at N_0 = 35 dB SPL and fixed tone levels 6 dB above the multi-channel model threshold for each bandwidth. For narrowband maskers, less masking is produced with the fixed spectral level (as expected), and the PO mechanism begins to break down. This change can be seen in the weight pattern for the narrowest band masker (e.g. Figure 5a) by the strong positive weights for cells tuned near the tone frequency. Because the narrowband masker noise does not saturate the AN fibers, addition of the tone results in an increase in the response count for many coincidence detector cells. Note that the cells with the strongest positive weights are still not those that receive matched-CF inputs, but rather are those which receive inputs that are tuned to slightly different frequencies, with a ratio of about 1.1, centered about the tone frequency. Upon addition of the tone to the noise, there is an increase in the correlation of the temporal responses of the AN inputs to these cells. Coincidence-detecting cells that receive inputs tuned to slightly different frequencies can have wider dynamic ranges than AN fibers because they can take advantage of nonlinear phase cues [20, 63]; these cells can encode increases in stimulus energy at levels for which the AN fibers (and cells with matched-CF inputs) are saturated. In the limiting case of detection of a tone in quiet, the strategy would be to detect an increase in the responses of the most sensitive cells in the population, which are the cells along the lowest row that receive AN inputs with matched CFs. This is a classical energy-based strategy for detection.

As masker bandwidth increases to the point that cells tuned to 900 Hz are saturated, the PO mechanism begins to outperform the energy-based strategy, as indicated by the strong negative (dark) weights (Figure 5b–f). In this case, the strategy becomes detection of a reduction in the response of the phase-opponent cells.

The variation in threshold as a function of masker bandwidth is an important property of any masking model. The CBE model is based on the systematic increase in threshold as bandwidth increases (with fixed spectral level) up to a critical band, beyond which thresholds are stable [1]. The same general trend is expected for the PO model; detection thresholds increase as bandwidth is increased for narrowband maskers. Once the masker bandwidth (and level) results in saturated responses of the inputs to the coincidence detectors, the detection threshold remains stable. The thresholds across a range of bandwidths for the multi-channel model with internal noise are illustrated in Figure 6. For each bandwidth, the population weight pattern for that bandwidth (see Figure 5) was used to compute the decision variable during the simulated tracking algorithm. The thresholds for the multi-channel model without internal noise (not shown) are very similar to the thresholds with noise; this result is expected because the multi-channel model combines responses across a substantial number of model cells, which reduces the effects of internal noise on the threshold estimates.

For fixed-level maskers at super-critical bandwidths, the model thresholds (open triangles) are comparable to human thresholds (filled triangles, Figure 6) [7]. In addition, at these bandwidths the model's thresholds are not influenced by roving the masker level (open squares), similar to results for human listeners (filled squares, Figure 6). The change in model threshold as a function of masker bandwidth demonstrates a critical-band-like effect for the fixed-level maskers (open triangles); thresholds remain stable for masker bandwidths greater than approximately

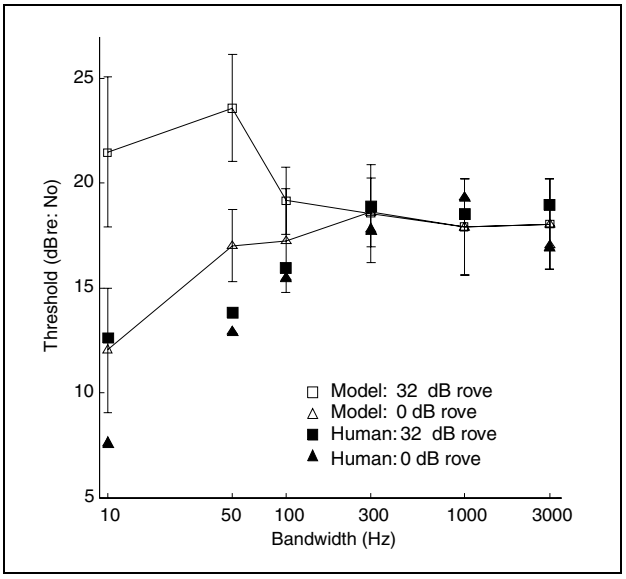

Figure 6. Simulated thresholds for the multi-channel model with internal noise as a function of masker bandwidth for fixed-level (0 dB rove) maskers at 35 dB SPL spectral level (open triangles), and for masker levels roved over a 32-dB range centered at 35 dB SPL (open squares). For comparison, thresholds of human listeners from Kidd *et al.* [7] are also shown for the 0-dB rove (filled triangles) and the 32-dB rove (filled squares) conditions.

50–100 Hz. This effect is not clear for the roving-level maskers (open squares), because the model thresholds for narrow (sub-critical band) maskers are strongly affected by roving the level of the masker. For sub-critical bandwidth maskers, for which AN fibers are not saturated, the model does not account for the human thresholds (filled squares), especially for roving-level maskers. Sub-critical band maskers are also the conditions for which human listeners are most affected by roving the masker level (Figure 6) (e.g. [7, 8]).

4. Discussion

The PO model demonstrates the successful application of a realistic temporal processing mechanism to the problem of masked detection. Responses of the model were directly compared to human performance in detection of a 900-Hz tone in noise. The first example focused on a simple case for the purpose of illustration: the use of a single-channel coincidence-detector model that received inputs from two model AN fiber CFs. The PO model was then extended to a multi-channel model using a weighted combination of model cell responses as a decision variable for simulations of the psychoacoustical task. The multi-channel model included the contributions of cells that had both negative and positive changes in response counts upon addition of the tone to the masker. For wideband maskers, the most statistically significant contributions to the population response were the reductions in response of the phase-opponent cells. For narrowband maskers, or for very low-level wideband maskers, cells with increased rates contribute more substantially.

4.1. Limitations of the PO Model

The detection thresholds for both the single-channel and multi-channel models are similar to those reported for human listeners in many respects. For super-critical band maskers, the absolute thresholds of the model are close to those of human listeners, both for the fixed-level and roving-level conditions. For a 3-kHz masker, thresholds were comparable to those of human listeners for roving- and fixed-level conditions for both the single-channel and multi-channel models (Table I). In the narrowband condition, the PO mechanism breaks down because the AN fibers tuned near the tone frequency are not saturated. In this condition, the most significant changes in the response of the multi-channel model were increased rates of coincidence cells that received AN inputs tuned to slightly different frequencies (e.g. Figure 5a). However, the encoding of the added tone by the increased rates of these cells is not robust to the roving-level paradigm, as seen in the model results for the roving-level condition at narrowbands (Figure 6). In order to make use of the increased rates of these cells, an independent estimate of masker level would be required. A more sophisticated model that varied the weights as a function of level during the course of the roving-level paradigm could make use of the combination of increased and decreased rates across the population of cells; such a model would be expected to show improved performance for the roving-level narrowband maskers. Narrowband stimuli also contain significant envelope fluctuations that are not taken into account by simple schemes based on rate cues. These envelope fluctuations influence the temporal responses of AN fibers because of cochlear nonlinearity; however, the PO mechanism alone cannot effectively take advantage of these cues for narrowband maskers because the AN fibers are not saturated. More complex models based on a combination of energy cues and cross-frequency temporal cues should be explored to test their ability to explain detection of tones in narrowband maskers.

The model thresholds provided reasonably accurate prediction of human performance except for the narrowband roving-level conditions (Figure 6). It is important to point out that none of the model parameters were adjusted to match the human thresholds. For example, no internal noise was added to adjust baseline thresholds, as is often done in psychophysical models (e.g. [24, 58, 59]). In this model, the internal noise was represented by statistics based on the discharge counts of actual auditory neurons (e.g. [14, 16, 54]). In addition to the internal noise, the most important model parameters for threshold determination were the size of the coincidence window (discussed above) and the bandwidth of the AN filters. These bandwidths were based on estimates of human auditory filters at appropriate noise levels [29] which are incorporated into the AN model used here [21]. It is likely that these psychophysically derived estimates of auditory-filter bandwidths are relatively broad as compared to physiological measures of AN tuning due to the influence of cochlear nonlinearities (see [60, 64, 65]). However, es-

timates of filter bandwidths based on the notched-noise masking paradigm seem to provide appropriate "effective" bandwidths for use in modeling studies of masked detection (see [66]).

4.2. Extension of the Model to other Masking Phenomena

Further tests of this model for masked detection at low frequencies should explore its ability to generalize to other phenomena involving noise maskers, such as critical ratios [1, 67], critical bands [68], and notched-noise masked thresholds (e.g. [69]). The model should also be tested with masking experiments that have been designed to quantify the relative contributions of different stimulus aspects on detection, such as manipulation of energy cues and/or temporal cues related to either fine-structure or envelope fluctuations (e.g. [8, 9, 10]. In addition, the model should be tested for detection tasks using non-Gaussian noise (e.g. [70]). Preliminary tests of the model's performance in predicting results for detection of tones in reproducible noise have been reported [71].

The multi-channel model could be extended to study masked detection of a tone at an unknown frequency based on identification of a localized reduction in the response of a population of coincidence cells to the wideband stimulus. For fixed-level maskers, a reduction in rate indicates the presence of a narrowband signal, and the place of the reduced response within the tonotopic map indicates the frequency of that signal. Human detection thresholds in repeatable noise with signal-frequency uncertainty are elevated by approximately 2–5 dB over fixed-frequency detection thresholds [72]. For a fixed-level masker, an estimate of masker level would provide the observer with information that could focus attention on a subset of the cells in the population. In contrast, detection of a roving-frequency tone in a roving-level masker paradigm would require monitoring of the entire population to identify the added tone, because the cells within the coincidence-detection population that provide the reduced response would vary with overall masker level due to cochlear nonlinearity (Figure 4).

In previous studies, we have used cross-frequency coincidence detectors to extract temporal information from AN responses to tones in quiet over a wide range of tone levels at both low and high frequencies [20, 63]. The coincidence-detection mechanism is quite general in that it provides a realistic processor that is sensitive to changes in temporal responses as well as to changes in average discharge rate [20]. Strategies for interpreting the responses of coincidence-detecting cells vary with the CFs of the AN inputs. These different strategies introduce an important distinction between the role of coincidence detection in the model for level discrimination of tones in quiet [20] and the model presented here for detection of tones in noise. Coincidence detectors that receive inputs with the same CFs are most sensitive to tones in quiet, which are detected by an increase in the response of the cell. Coincidence detectors that receive inputs of similar, but not identical, CFs are able to encode changes in tone level over a wide dynamic range based on the increase in cross-frequency correlation as level increases, which is due to changes in the phase of AN fibers with sound level [20]. These model cells were the most important coincidence-detecting cells in our model for level-discrimination of tones in quiet. As shown here, in the presence of a masking noise that saturates the rates of AN fibers tuned near the frequency of the tone, the most sensitive detection is based on a drop in the response of coincidence detectors that receive inputs with CFs that are well separated in frequency, the phase-opponent cells.

4.3. Physiological Evidence for Coincidence-detecting Cells

Physiological evidence for phase-opponent coincidence-detecting cells in the auditory brainstem is a topic for further study. Several of the elements suggested in these simulations have been observed in recordings from the AVCN. There is a diversity of rate-level functions in the AVCN in response to pure tones (e.g. [73, 74]); this diversity is expected for coincidence-detectors receiving different combinations of CF inputs from AN fibers. The sensitivity of AVCN cells to manipulations of the phase spectrum of transient stimuli (Huffman sequences, in [45]) varies from cell to cell, consistent with different cells receiving different combinations of CF inputs. Finally, in the gerbil, some cells in the AVCN that respond well to wideband noise show a decrease in response rate when a tone is added to the stimulus, which is appropriate for phase-opponent cells (unpublished observation). A comprehensive study of the properties of these cells in the context of this model has not yet been completed. It is clear from the variation of model thresholds as a function of bandwidth that results consistent with the critical-band interpretation of masked-detection data can be obtained by models, such as PO, that are based on comparisons of responses across different auditory filters as opposed to the response of a single auditory filter.

4.4. Extension of the Model to Higher Frequencies

The PO model can be extended to high frequencies, for which the temporal responses of AN fibers convey information about the envelope of complex sounds rather than their fine-time structure (e.g. [75]). Similar to low-frequency tone detection, thresholds for detection of high-frequency tones in wideband noise are not affected by roving the masker level [76]. A high-frequency model for detection must take into account not only the information carried in the temporal responses that are phase-locked to the envelope, but also the information carried in the average rate. The discharges of low-spontaneous-rate AN fibers are less effected by rate saturation at high CFs than at low CFs.

In addition, high-frequency extensions of this model should consider the potential role of lateral inhibition. Deng and Geisler [44] discussed the close relationship of

their cross-correlation model to the lateral inhibition network (LIN) scheme proposed by Shamma [77]. Lateral inhibition can be represented as a cross-channel differencing operation; thus, spectral peaks that produce a reduction in the output of cross-channel correlation produce a peak in the output of an LIN operation. Physiological evidence for lateral inhibition in the cochlear nucleus has been obtained mainly at high frequencies (e.g. [78, 79]). Based on extracellular recordings of responses to tones masked by noise, Rhode and Greenberg [78] reported that evidence for lateral inhibition was most clearly present in the responses of AVCN choppers and cells in the dorsal cochlear nucleus. However, Caspary *et al.* [79] used pharmacological techniques to show that the inhibitory responses of both chopper and primarylike cells in the AVCN were centered at the excitatory best frequency, as opposed to being strongest at frequencies lateral to CF. As they point out, this finding is consistent with anatomical and physiological observations related to the inhibitory inputs onto AVCN cells. The changes in the nature of temporal information and the relative importance of temporal and rate information at low and high frequencies may lead to the use of different neural mechanisms at different frequencies, e.g., coincidence detection, lateral inhibition, or combinations of these mechanisms.

4.5. Implications for Understanding Hearing Impairment

There may be implications of this model for understanding hearing impairment; damage to the auditory periphery not only affects the sensitivity of the ear, but also affects its frequency selectivity [80] and therefore the phase properties of its responses. Thus, neural processing mechanisms that involve spatio-temporal encoding schemes are likely to be influenced by hearing impairment. Models for detection and discrimination in noise are of particular interest as the difficulties of hearing-impaired listeners are most pronounced in noisy environments [80].

The PO model may also provide insight into the difficulty of noisy environments for cochlear-implant users (e.g. [81, 82]). Cochlear implants directly stimulate the AN with electrical current pulses, thus bypassing the damaged auditory periphery. Electrically stimulated AN fibers are extremely well synchronized [83] and do not contain the phase delays normally associated with the filters in the auditory periphery. Thus, although implanted ears have temporal information that might be considered superior to that of the healthy ear, they do not have normal spatio-temporal patterns. In particular, implant responses completely lack the phase delays critical for a mechanism such as phase opponency that takes advantage of the information in the spatio-temporal discharge patterns of the AN.

Acknowledgement

We thank David Cameron, Torsten Dau, Susan Early, J. Michael Harrison, Gerald Kidd, Jr., Christine Mason, Brian Moore, Ray Meddis, and Paul Nelson for their comments on previous versions of this paper. We thank Gerald Kidd, Jr., and Christine Mason for providing the human thresholds for comparison to model results. This work was supported by National Institutes of Health grants DC-01641 and DC-00100, National Science Foundation grant IBN 9983567, and AFOSR F49620-97-1-0231.

References

[1] H. Fletcher: Auditory patterns. Rev. Mod. Phys. **12** (1940) 47–65.

[2] D. M. Green, J. A. Swets: Signal detection theory and psychophysics. Wiley, New York, 1966. (reprinted 1988 by Peninsula, Los Altos, CA).

[3] B. C. J. Moore: An introduction to the psychology of hearing. Academic Press, San Diego, 1997.

[4] S. Buus: Auditory masking. Encyclopedia of Acoustics. John Wiley and Sons, New York, 1997.

[5] D. M. Green: Profile analysis. Oxford University Press, New York, 1988.

[6] D. L. Neff, D. M. Green: Masking produced by spectral uncertainty with multicomponent maskers. Percept. Psychophys. **41** (1987) 409–415.

[7] G. Kidd Jr., C. R. Mason, M. A. Brantley, G. A. Owen: Roving-level tone-in-noise detection. J. Acoust. Soc. Am. **86** (1989) 1310–1317.

[8] V. M. Richards, L. M. Heller, D. M. Green: Detection of a tone in a narrowband of noise: The energy model revisited. Q. J. Psychol. **43A** (1991) 481–501.

[9] M. F. Spiegel, M. C. Picardi, D. M. Green: Signal and masker uncertainty in intensity discrimination. J. Acoust. Soc. Am. **70** (1981) 1015–1019.

[10] J. W. Hall III, J. H. Grose: Comodulation masking release: Evidence for multiple cues. J. Acoust. Soc. Am. **84** (1988) 1669–1675.

[11] V. M. Richards, R. D. Nekrich: The incorporation of level and level-invariant cues for the detection of a tone added to noise. J. Acoust. Soc. Am. **94** (1993) 2560–2574.

[12] V. M. Richards: The detectability of a tone added to narrow bands of equal-energy noise. J. Acoust. Soc. Am. **91** (1992) 3424–3435.

[13] H. S. Colburn: Intensity perception: relation of intensity discrimination to auditory-nerve firing patterns. Internal memorandum, Research Laboratory of Electronics, 1981.

[14] B. Delgutte: Peripheral auditory processing of speech information: implications from a physiological study of intensity discrimination. – In: The Psychophysics of Speech Perception. M. E. H. Schouten (ed.). Nijhoff, Dordrecht, 1987, 333–353.

[15] N. F. Viemeister: Psychophysical aspects of intensity discrimination. – In: Auditory Function: Neurobiological Bases of Hearing. G. M. Edelman, W. E. Gall, W. M. Cowan (eds.). Wiley, New York, 1988, 213–241.

[16] I. M. Winter, A. R. Palmer: Intensity coding in low-frequency auditory-nerve fibers of the guinea pig. J. Acoust. Soc. Am. **90** (1991) 1958–1967.

[17] B. Delgutte: Psychophysiological models for basic auditory percepts. – In: Auditory Computation. H. L. Hawkins, T. A. McMullen, A. N. Popper, R. R. Fay (eds.). Springer-Verlag, New York, 1996, 157–220.

[18] M. B. Sachs, P. J. Abbas: Rate versus level functions for auditory-nerve fibers in cats: tone-burst stimuli. J. Acoust. Soc. Am. **56** (1974) 1835–1847.

[19] G. K. Yates, B. M. Johnstone, R. B. Patuzzi, D. Robertson: Mechanical preprocessing in the mammalian cochlea. Trends Neurosci. **15** (1992) 57–61.

[20] M. G. Heinz, H. S. Colburn, L. H. Carney: Rate and timing cues associated with the cochlear amplifier: Level discrimination based on monaural cross-frequency coincidence detection. J. Acoust. Soc. Am. **110** (2001) 2065–2084.

[21] M. G. Heinz, X. Zhang, I. C. Bruce, L. H. Carney: Auditory nerve model for predicting performance limits of normal and impaired listeners. ARLO **2** (2001) 91–96.

[22] R. L. Winslow, M. B. Sachs: Single-tone intensity discrimination based on auditory-nerve rate responses in backgrounds of quiet, noise, and with stimulation of the crossed olivocochlear bundle. Hear. Res. **35** (1988) 165–189.

[23] M. L. Hicks, S. P. Bacon: Psychophysical measures of auditory nonlinearities as a function of frequency in individuals with normal hearing. J. Acoust. Soc. Am. **105** (1999) 326–338.

[24] N. I. Durlach, L. D. Braida, Y. Ito: Towards a model for discrimination of broadband signals. J. Acoust. Soc. Am. **80** (1986) 63–72.

[25] L. H. Carney, T. C. T. Yin: Temporal coding of resonances by low-frequency auditory nerve fibers: Single fiber responses and a population model. J. Neurophysiol. **60** (1988) 1653–1677.

[26] L. H. Carney: Spatiotemporal encoding of sound level: Models for normal encoding and recruitment of loudness. Hearing Research **76** (1994) 31–44.

[27] D. J. Johnson: The relationship between spike rate and synchrony in response of auditory-nerve fibers to single tones. J. Acoust. Soc. Am. **68** (1980) 1115–1122.

[28] H. S. Colburn: Theory of binaural detection based on auditory-nerve data. II. Detection of tones in noise. J. Acoust. Soc. Am. **61** (1977) 525–533.

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

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

[31] W. S. Rhode, C. D. Geisler, D. K. Kennedy: Auditory-nerve fiber responses to wide-band noise and tone combinations. J. Neurophysiol. **41** (1971) 692–704.

[32] P. J. Abbas: Auditory-nerve fiber responses to tones in a noise masker. Hear. Res. **5** (1981) 69–80.

[33] M. I. Miller, P. E. Barta, M. B. Sachs: Strategies for the representation of a tone in background noise in the temporal aspects of the discharge patterns of auditory-nerve fibers. J. Acoust. Soc. Am. **81** (1987) 665–679.

[34] J. C. R. Licklider: A duplex theory of pitch perception. Experientia **7** (1951) 128–133. Reprinted in: Psychological Acoustics, edited by E. D. Schubert (Dowden, Hutchinson and Ross, Stroudsburg, PA, 1979).

[35] J. L. Goldstein: An optimum processor theory for the central formation of the pitch of complex tones. J. Acoust. Soc. Am. **54** (1973) 1496–1516.

[36] J. L. Goldstein, P. Srulovicz: Auditory-nerve spike intervals as an adequate basis for aural frequency measurement. – In: Psychophysics and Physiology of Hearing. E. F. Evans, J. P. Wilson (eds.). Academic Press, New York, 1977, 337–347.

[37] G. Langner: Neuronal mechanisms for pitch analysis in the time domain. Exp. Brain. Res. **44** (1981) 450–454.

[38] E. D. Young, M. B. Sachs: Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. J. Acoust. Soc. Am. **66** (1979) 1381–1403.

[39] B. Delgutte, N. Y. S. Kiang: Speech coding in the auditory nerve: I. Vowel-like sounds. J. Acoust. Soc. Am. **75** (1984) 866–878.

[40] R. D. Patterson: The sound of a sinusoid: Time-interval models. J. Acoust. Soc. Am. **96** (1994) 1419–1428.

[41] G. E. Loeb, M. W. White, M. M. Merzenich: Spatial cross-correlation. A proposed mechanism for acoustic pitch perception. Biol. Cybern. **47** (1983) 149–163.

[42] S. A. Shamma, N. M. Shen, P. Gopalaswamy: Stereausis: binaural processing without neural delays. J. Acoust. Soc. Am. **86** (1989) 989–1006.

[43] J. M. Goldberg, P. B. Brown: Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization. J. Neurophysiol. **32** (1969) 613–636.

[44] L. Deng, C. D. Geisler: A composite auditory model for processing speech sounds. J. Acoust. Soc. Am. **82** (1987) 2001–2012.

[45] L. H. Carney: Sensitivities of cells in anteroventral cochlear nucleus of cat to spatiotemporal discharge patterns across primary afferents. J. Neurophysiol. **64** (1990) 437–456.

[46] T. C. Yin, J. C. Chan, L. H. Carney: Effects of interaural time delays of noise stimuli on low-frequency cells in the cat's inferior colliculus. III. Evidence for cross-correlation. J. Neurophysiol. **58** (1987) 562–583.

[47] P. X. Joris, L. H. Carney, P. H. Smith, T. C. Yin: Enhancement of neural synchronization in the anteroventral cochlear nucleus. I. Responses to tones at the characteristic frequency. J. Neurophysiol. **71** (1994) 1022–1036.

[48] I. M. Winter, A. R. Palmer: Level dependence of cochlear nucleus onset unit responses and facilitation by second tones or broadband noise. J. Neurophysiol. **73** (1995) 141–159.

[49] W. S. Rhode, D. Oertel, P. H. Smith: Physiological properties of cells labelled intracellularly with horseradish peroxidase in cat ventral cochlear nucleus. J. Comp. Neurol. **213** (1983) 448–463.

[50] S. H. Wu, D. Oertel: Intracellular injection with horseradish peroxidase of physiologically characterized stellate and bushy cells in slices of mouse anteroventral cochlear nucleus. J. Neuroscience **4** (1984) 1577–1588.

[51] P. B. Manis, S. O. Marx: Outward currents in isolated ventral cochlear nucleus neurons. J. Neuroscience **11** (1991) 2865–2880.

[52] P. H. Smith: Structural and functional differences distinguish principal from nonprincipal cells in the guinea pig mso slice. J. Neurophysiol. **73** (1995) 1653–1667.

[53] H. S. Colburn: Some physiological limitations on binaural performance. PhD Dissertation. Massachusetts Institute of Technology, 1969.

[54] L. H. Carney, M. A. Burock: Encoding of sound level by discharge rates of auditory neurons. Comments on Theoretical Biology **4** (1997) 315–337.

[55] D. J. Johnson, N. Y.-S. Kiang: Analysis of discharges recorded simultaneously from pairs of auditory-nerve fibers. Biophys. J. **16** (1976) 719–734.

[56] H. L. van Trees: Detection, estimation, and modulation theory. John Wiley and Sons, New York, 1968.

[57] M. G. Heinz, J. Goldstein, M. H., C. Formby: Temporal gap detection thresholds in sinusoidal markers simulated with a multi-channel, multi-resolution model of the auditory periphery. Aud. Neurosci. **3** (1996) 35–56.

[58] T. Dau, D. Püschel, A. Kohlrausch: A quantitative model of the "effective" signal processing in the auditory system: II. Simulations and measurements. J. Acoust. Soc. Am. **99** (1996) 3623–3631.

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

[60] M. G. Heinz: Quantifying the effects of the cochlear amplifier on temporal and average-rate information in the auditory nerve. PhD Dissertation. Massachusetts Institute of Technology, 2000.

[61] M. G. Heinz, H. S. Colburn, L. H. Carney: Evaluating auditory performance limits: Ii. one-parameter discrimination with random level variation. Neural Computation **13** (2001) 2317–2339.

[62] D. J. Anderson, J. E. Rose, J. E. Hind, J. F. Brugge: Temporal position of discharges in single auditory nerve fibers within the cycle of a sinewave stimulus: frequency and intensity effects. J. Acoust. Soc. Am. **49** (1971) 1131–1139.

[63] L. H. Carney: Spatiotemporal encoding of sound level: Models for normal encoding and recruitment of loudness. Hear. Res. **76** (1994) 31–44.

[64] M. G. Heinz, H. S. Colburn, L. H. Carney: Quantifying the implications of nonlinear cochlear tuning for auditory-filter estimates. J. Acoust. Soc. Am. **111** (2002) 996–1011.

[65] C. A. Shera, J. J. Guinan, A. J. Oxenham: Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. Proceedings of the National Academy of Science **99** (2002) 3318–3323.

[66] R. P. Derleth, T. Dau: On the role of envelope fluctuation processing in spectral masking. J. Acoust. Soc. Am. **108** (2000) 285–296.

[67] D. M. Green, M. J. McKey, J. C. R. Licklider: Detection of a pulsed sinusoid as a function of frequency. J. Acoust. Soc. Am. **31** (1959) 1446–1452.

[68] E. Zwicker: Subdivision of the audible frequency range into critical bands (Frequenzgruppen). J. Acoust. Soc. Am. **33** (1961) 248.

[69] B. C. J. Moore, B. R. Glasberg: Auditory filter shapes derived in simultaneous and forwrad masking. J. Acoust. Soc. Am. **69** (1981) 1003–1014.

[70] A. Kohlrausch, R. Fasel, M. van der Heijden, R. Kortekas, S. van de Par, A. Oxenham, D. Püschel: Detection of tone in low-noise noise: Further evidence for the role of envelope fluctuations. Acustica – acta acustica **83** (1997) 659–669.

[71] L. Carney, M. E. Evilsizer, R. H. Gilkey, M. G. Heinz, H. S. Colburn: Modelling diotic and dichotic detection of tones in reproducible noise. Abstract, Association for Research in Otolaryngology **24** (2001) 86.

[72] M. F. Spiegel, D. M. Green: Signal and masker uncertainty with noise maskers of varying duration, bandwidth, and center frequency. J. Acoust. Soc. Am. **71** (1982) 1204–1210.

[73] I. M. Winter, A. R. Palmer: Responses of single units in the anteroventral cochlear nucleus of guinea pig. Hearing Research **44** (1990) 161–178.

[74] T. R. Bourk: Electrical responses of neural units in the anteroventral cochlear nucleus of cat. PhD Dissertation, MIT, 1976.

[75] P. X. Joris, T. C. T. Yin: Responses to amplitude-modulated tones in the auditory nerve of the cat. J. Acoust. Soc. Am. **91** (1992) 215–232.

[76] P. Allen, S. Turpin: Childrens detection of tonal signals in roving level noise maskers. J. Acoust. Soc. Am. (Abstract) **109** (2001) 2464.

[77] S. A. Shamma: Speech processing in the auditory system. II: Lateral inhibition and the central processing of speech evoked activity in the auditory nerve. J. Acoust. Soc. Am. **78** (1985) 1622–1632.

[78] W. S. Rhode, S. Greenberg: Lateral suppression and inhibition in the cochlear nucleus of the cat. J. Neurophysiol. **71** (1994) 493–514.

[79] D. M. Caspary, P. M. Backoff, P. G. Finlayson, P. S. Palombi: Inhibitory inputs modulate discharge rate within frequency receptive fields of anteroventral cochlear nucleus neurons. J. Neurophysiol. **72** (1994) 2124–2133.

[80] B. C. J. Moore: Perceptual consequences of cochlear damage. Oxford University Press, Oxford, 1995.

[81] R. C. Dowell, P. M. Seligman, P. J. Blamey, G. M. Clark: Speech perception using a two-formant 22-electrode cochlear prosthesis in quiet and in noise. Acta Otolaryngol. **104** (1987) 439–446.

[82] I. Hochberg, A. Boothroyd, M. Weiss, S. Hellman: Effects of noise and noise suppression on speech perception by cochlear implant users. Ear. Hear. **13** (1992) 263–271.

[83] C. van den Honert, P. H. Stypulkowski: Temporal response patterns of single auditory nerve fibers elicited by periodic electrical stimuli. Hear. Res. **29** (1987) 207–222.