Predictions of formant-frequency discrimination in noise based on model auditory-nerve responses

Qing Tan
Boston University Hearing Research Center, Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts 02215

Laurel H. Carney[a)]
Boston University Hearing Research Center, Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts 02215, and Departments of Biomedical & Chemical Engineering and Electrical Engineering & Computer Science, Institute for Sensory Research, 621 Skytop Road, Syracuse University, Syracuse, New York 13244

(Received 18 August 2005; revised 19 June 2006; accepted 21 June 2006)

To better understand how the auditory system extracts speech signals in the presence of noise, discrimination thresholds for the second formant frequency were predicted with simulations of auditory-nerve responses. These predictions employed either average-rate information or combined rate and timing information, and either populations of model fibers tuned across a wide range of frequencies or a subset of fibers tuned to a restricted frequency range. In general, combined temporal and rate information for a small population of model fibers tuned near the formant frequency was most successful in replicating the trends reported in behavioral data for formant-frequency discrimination. To explore the nature of the temporal information that contributed to these results, predictions based on model auditory-nerve responses were compared to predictions based on the average rates of a population of cross-frequency coincidence detectors. These comparisons suggested that average response rate (count) of cross-frequency coincidence detectors did not effectively extract important temporal information from the auditory-nerve population response. Thus, the relative timing of action potentials across auditory-nerve fibers tuned to different frequencies was not the aspect of the temporal information that produced the trends in formant-frequency discrimination thresholds. © *2006 Acoustical Society of America*
.[DOI: 10.1121/1.2225858]

PACS number(s): 43.64.Bt [WPS] Pages: 1435–1445

I. INTRODUCTION

The auditory system has a remarkable ability to extract speech information in the presence of noise. The study presented here was intended to improve our understanding of signal-processing and encoding mechanisms in the peripheral auditory system. Formant-frequency discrimination in background noise was studied using simulations of cat model auditory-nerve (AN) responses, and predictions of discrimination performance based on the AN model responses were compared with behavioral data in cat (Hienz *et al.*, 1998). Analyses of the AN model responses were based on either the information present in the discharge rate alone or on the combined information in both the rate and timing of AN discharges. In addition, one simple neural mechanism for decoding temporal information, coincidence detection, was investigated to determine whether cross-frequency coincidences of AN discharges contributed to the temporal encoding of formant frequencies.

Human thresholds for frequency discrimination of the second formant (F2) have been reported in the range of 1%–5% (Flanagan, 1955) to 6.8% (Mermelstein, 1978) in earlier studies, and as low as 1.9% (Hawks, 1994) and 1%–2% (Kewley-Port and Watson, 1994; Kewley-Port *et al.*, 1996) in more recent studies. The latter studies included highly trained listeners in paradigms with low stimulus uncertainty. These thresholds are roughly an order of magnitude higher than pure-tone frequency discrimination thresholds; the pure-tone frequency discrimination threshold for a tone in the F2 frequency region (e.g., 2000 Hz) is 3.2 Hz (0.16%) (Wier *et al.*, 1977). However, formant-frequency discrimination differs fundamentally from pure-tone frequency discrimination. Changes in the formant frequency are not associated with changes in the frequencies of the harmonics, but rather with a pattern of level changes distributed across the harmonic frequencies in the region of the formant peak (Fig. 1). Thus, understanding formant-frequency discrimination presumably requires an understanding of how the combined changes in the amplitudes of the harmonics associated with changes of the spectral envelope's center frequency affect the response patterns of the auditory periphery.

Quantitative tools are available for predicting behavioral thresholds based on the responses of AN fibers. Siebert (1965, 1968) developed a strategy for predicting limits of level discrimination for pure tones based on a very simple description of the auditory periphery. Heinz *et al.* (2001a) extended Siebert's approach to allow the use of recent computational models for AN responses. Since both approaches

[a)]Author to whom correspondence should be addressed: Institute for Sensory Research, 621 Skytop Road, Syracuse University, Syracuse, NY 13244. Electronic mail: Lacarney@syr.edu

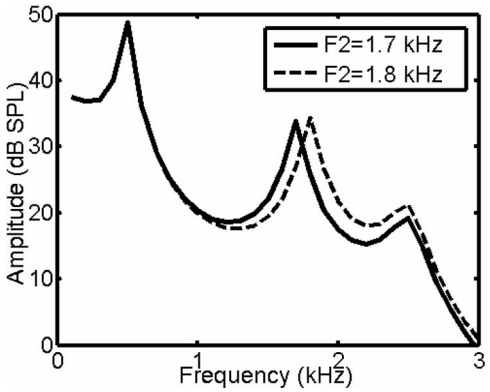

FIG. 1. Examples of spectral envelopes of synthesized vowels, illustrating the vowel with the standard second formant frequency (1700 Hz, solid) and with a frequency shift in the second formant frequency (1800 Hz, dashed).

include analyzing population responses and combining information across fibers tuned to different frequencies, they provide appropriate tools to study formant-frequency encoding. Furthermore, Heinz (2000) developed a computational strategy for predicting psychophysical performance on tasks that involve tones in noise maskers. We have applied and extended his strategies in this study of formant-frequency discrimination of vowels in quiet and in the presence of noise maskers.

Understanding speech coding in the presence of masking noise is a real challenge for both physiologists and psychophysicists. A central question in this puzzle is what aspects of AN responses convey information about complex sounds to the central nervous system. Most studies of speech sounds, including studies of speech recognition, focus on spectral energy, and the most straightforward coding strategy for energy is typically assumed to be average discharge rate. However, intense noise maskers and the relatively high levels of harmonics near formant peaks tend to dominate the rates of AN fibers, reducing the sensitivity of rate to small changes in the levels of neighboring harmonics. Hienz *et al.* (1996) concluded that behavioral results in cat for vowel discrimination in the presence of high-level maskers could not be explained by the rates of high-spontaneous-rate (HSR) AN fibers, which are severely degraded in this stimulus condition (Sachs and Young, 1979; Delgutte and Kiang, 1984b; Sachs *et al.*, 1983). Although low-spontaneous-rate (LSR) fibers may provide a wider dynamic range based on rate, it is clear from our results that there is ample information in the timing of small populations of HSR fibers to explain the behavioral results. The importance of the temporal responses of small populations of fibers tuned near the frequency of interest is consistent with previous studies (Young and Sachs, 1979; Sachs *et al.*, 1983; Delgutte and Kiang, 1984a, b; Miller *et al.*, 1987; Tan and Carney, 2005).

One concern is that quantitative predictions of psychophysical performance based on physiological responses often require comparisons across species. Most knowledge of AN responses is based on the cat, whereas most psychophysical knowledge of vowel discrimination is based on studies with human listeners. In the case of formant-frequency discrimination, however, behavioral results in the cat (Hienz *et al.*, 1998) can be used to directly compare predictions based on cat model AN responses to psychophysical results for F2 discrimination. Hienz *et al.* (1998) measured behavioral formant-frequency discrimination thresholds in cat, for both vowels in quiet and with noise maskers at various signal-to-noise ratios. Their results show that formant-frequency-discrimination thresholds of cat at low and medium noise levels [signal-to-noise ratios (SNRs) of 23 and 13 dB, respectively] are similar to the threshold measured in quiet (approximately 2.5% of the F2 frequency), whereas the threshold in the presence of a high noise level (SNR of 3 dB) is significantly elevated (approximately 5.8%). In human listeners, a similar trend in thresholds is seen as a function of SNR, with thresholds increasing very gradually for high SNRs and then increasing more significantly at low SNRs (Liu and Kewley-Port, 2004).

In the current study, the Hienz *et al.* (1998) formant-frequency discrimination task was simulated using a population of cat AN model fibers (Tan and Carney, 2003). This AN model was used in a previous study (Tan and Carney, 2005) of a related but simpler task involving center-frequency discrimination for harmonic complexes (Lyzenga and Horst, 1995). In the previous study, the prediction methods assumed that an ideal central processor used detailed knowledge of the AN population response pattern for each waveform. However, this assumption is not realistic when a random noise masker is included in the stimulus, because knowledge of the details of the noise response allows an ideal processor to perform at extremely low thresholds (this issue is discussed further below). The study presented here represents an application of the approach of Heinz (2000) and Heinz *et al.* (2002), in which it was assumed that the central processor only has knowledge of the *averaged* response patterns to the stimuli that include noise (i.e., the AN model response was averaged across an ensemble of stimuli that included different tokens of random noise maskers).

Given the result that temporal information is sufficient to explain behavioral performance for formant-frequency discrimination, it is interesting to explore what aspects of the temporal responses are most important, and what neural mechanisms might be used to extract them. In this study, predictions for formant-frequency discrimination thresholds were also made using a coincidence-detection mechanism, which has achieved some success in predicting psychophysical thresholds for tone detection in noise (Carney *et al.*, 2002). The best predictions in the current study utilized both average-rate and timing information of a relatively small population of model auditory-nerve fibers tuned near the formant frequency. Predictions based on the average rates of cross-frequency coincidence detectors were not as successful as predictions based directly on temporal responses of model AN fibers or as predictions based on the timing of the responses of coincidence detectors. This comparison suggests that the relative timing of AN responses tuned to different frequencies was not the critical temporal aspect in the peripheral responses for the formant-frequency discrimination task.

II. METHODS[1]

A. Stimuli

Vowel stimuli were based on those used in Hienz *et al.* (1998): the synthetic vowel /ɛ/ was produced by a cascade Klatt Speech Synthesizer (Klatt, 1980) with a sampling rate of 10 kHz. The fundamental frequency of the vowel was 100 Hz. The first, third, fourth, and fifth formant frequencies of the synthesized speech signals were fixed at 500, 2500, 3300, and 3750 Hz, respectively. The second formant frequency varied between 1700 and 1900 Hz. Figure 1 shows the spectral envelopes of two sample vowels with different second-formant frequencies: 1700 Hz (solid line) and 1800 Hz (dashed line). The discrimination threshold was computed for changes in the second formant frequency from the standard frequency of 1700 Hz based on the response pattern of the AN model fibers, as described below. The vowels were 250 ms in duration, with rise/fall times of 20 ms. Sound pressure levels of the synthetic vowels were computed on the basis of the total power in the first 30 harmonics (100 Hz to 3 kHz). Results are presented in two level ranges; vowel levels were randomly varied (roved) over a range of 10 dB (±5 dB, in 1-dB increments) around 50 dB and 70 dB SPL rms. These levels and rove ranges were chosen to match the stimulus conditions used in the Hienz *et al.* (1998) behavioral study.

Noise maskers were added to the vowels to create stimuli with three different signal-to-noise ratios: 23, 13, and 3 dB SNR, selected to match the Hienz *et al.* (1998) experiments. The masking noise had a flat spectrum; signal-to-noise ratios for vowels in background noise were computed based on the total noise power up to 3 kHz, and noise levels were varied with respect to the vowel level to adjust the SNR. The overall levels of the noise masker and vowel were roved while maintaining the SNR at the indicated value for each condition.

B. AN model

The Tan and Carney (2003) AN model was used for all simulations presented here. The time-varying rate function of the AN model response was used for all calculations; individual action potentials were not simulated. The time-varying rate function is proportional to the probability of AN discharges throughout the timecourse of the stimulus waveform. Its name derives from the fact that it represents the time-varying average arrival rate for a nonhomogenous Poisson process that describes the AN response. The units of the rate function are spikes/s, and it can be thought of as a prediction of the peri-stimulus-time (PST) histogram of an AN fiber.

Characteristic frequencies (CFs, the frequency to which a model AN fiber is most sensitive) were evenly distributed on a log frequency scale. To study second formant frequency encoding, CFs from 1200 to 2500 Hz were included, representing 3.1 mm along the cochlear frequency map in cat (Liberman, 1982). Based on Keithley and Schreiber's (1987) measurement of the frequency map of the spiral ganglion in cat, this population of CFs comprised approximately 2000 neurons/mm $\times$ 3.1 mm, for a total of 6200 AN fibers.

Approximately 61% of the total AN population are HSR fibers (Liberman, 1978). Because our AN model only applies to HSR fibers, only 61% of the AN population was included in our threshold predictions. Therefore, 3782 (0.61×6200) model AN fibers with CFs logarithmically distributed from 1200 to 2500 Hz were used in these simulations. Fifty CFs were simulated; therefore, 76 AN fibers represented each CF channel for a total of 3782 fibers.

C. Threshold predictions based on AN model responses

Predictions of formant-frequency-discrimination thresholds were computed from AN model responses using techniques introduced to our field by Siebert (1965, 1968, 1970). His approach was to determine the threshold of an optimal detector based on observations (i.e., AN discharge patterns) that included the randomness associated with a Poisson process. The randomness was assumed to describe the variability in AN responses. Siebert's studies focused on frequency and level discrimination of pure tones in quiet; however, Heinz (2000) extended this approach to study tone detection in the presence of a masker noise.

It is not realistic to assume that an ideal central processor takes advantage of every detail of the response to every waveform when the stimulus includes random noise. For example, a subtle aspect of a vowel-plus-noise waveform might change when the formant frequency is slightly shifted. An ideal processor can take advantage of such subtle changes because it can directly compare the response to the noise plus a vowel with the standard formant frequency with the response to the same noise waveform plus the vowel with the shifted formant frequency. Realistically, the randomness of the noise waveforms, and their variation from trial to trial, precludes the use of subtle details of these waveforms in behavioral experiments. To better model the experimental situation, a suboptimal processor can be considered, such that detailed knowledge of each masker waveform is not available to the processor. Heinz (2000, see also Heinz *et al.*, 2002) used this approach to explore psychophysical performance limits for detection of tones in the presence of random noise by assuming that the central processor uses knowledge of the *expected* response to a given type of masker, rather than knowledge of the responses to individual noise waveforms on each trial. In other words, the central processor uses the expected response pattern (the averaged response to an ensemble of waveforms that include random noises) as the *a priori* information for each trial.

In this study, we adopted Heinz's (2000) approach and extended it to include the stimulus variations associated with the roving-level paradigm. Thus, the suboptimal detector used *a priori* information that was based on the expected response to an ensemble of stimuli that included several noise masker waveforms and a range of vowel and masker levels. In particular, the suboptimal detector's decision was based on the comparison of the model AN response to a particular vowel-plus-noise stimulus with *a priori* information consisting of averaged model AN responses to 220 different stimuli. Responses to vowels were simulated for 20

noise masker waveforms at 11 different levels (±5 dB in 1 -dB increments).

A straightforward decision variable based on observations of the AN model *population* response is an equally weighted combination of the decision variables based on *single* AN model responses, i.e.,

$$Y(\tau) = \sum_{i=1}^{M} Y_i(\tau^i), \tag{1}$$

where $Y_i(\tau^i)$ is the decision variable based on the observation of the set of AN discharge times, τ, of the ith fiber's response pattern, and M is equal to 3782, the total number of fibers in the population. Thus, AN model fibers with relatively large changes in their responses as formant frequency is varied, and thus large changes in their decision variables [i.e., large changes in $Y_i(\tau^i)$], contribute more to the changes in the total value of the decision variable $Y(\tau)$.

In Heinz (2000, Eq. 5.1), the sensitivity index Q for the central processor is defined by

$$Q(f, f+\Delta f) = \frac{\{E_{w,\tau}[Y(\tau)|f+\Delta f] - E_{w,\tau}[Y(\tau)|f]\}^2}{\mathrm{Var}_{w,\tau}[Y(\tau)|f]}, \tag{2}$$

where $E_{w,\tau}[Y(\tau)|f+\Delta f]$ and $E_{w,\tau}[Y(\tau)|f]$ are the *expected* values of the decision variable used by the central processor across an ensemble of waveforms, w, and across the set of AN discharge times, τ, with and without the formant-frequency shift (Δf). The ensemble of waveforms for the current study includes the set of different vowel plus noise levels, including all of the different random masker tokens, that were included to simulate the roving-level paradigm. $\mathrm{Var}_{w,\tau}[Y(\tau)|f]$ is the variance of the decision variable due to both the random noise of the masker waveform (external noise) and the randomness of the AN discharges (internal noise), which was assumed to be described by Poisson statistics. Note that the metric Q is very similar to the commonly used measure of sensitivity, d'. (The two measures are equivalent in the case of Gaussian distributions with equal variance.) It is convenient to work with the normalized sensitivity squared, given by

$$[\delta'(f)]^2 = \frac{Q(f, f+\Delta f)}{(\Delta f)^2}. \tag{3}$$

Combining Eqs. (1)–(3) and using the description of sensitivity based on a single AN fiber rate function [e.g., Siebert (1970); see Heinz *et al.* (2001a) for derivation], the normalized sensitivity squared for the AN model population is [Eq. 5.34 in Heinz (2000)]

$$[\delta'(f)]^2 = \frac{\{\sum_{i=1}^{M} \int_0^T [1/\bar{r}_i(t|f)][\bar{\dot{r}}_i(t|f)]^2\, dt\}^2}{(\sum_{i=1}^{M} \int_0^T [1/\bar{r}_i(t|f)][\bar{\dot{r}}_i(t|f)]^2\, dt + \mathrm{Var}_w\{\sum_{i=1}^{M} \int_0^T [[\bar{\dot{r}}_i(t|f)]/\bar{r}_i(t|f)]r_i(t|f,w)dt\})}, \tag{4}$$

where M is the total number of fibers in the population; $r_i(t|f,w)$ is the response of the ith AN model fiber to the wth vowel-plus-masker with the second formant frequency, f. T is the duration of the stimulus waveform. The partial derivative of the rate function with respect to second formant frequency is indicated by a dot over the symbol, e.g., $\dot{r}_i(t|f)$. The overbar, e.g., $\bar{r}_i(t|f)$, indicates the averaged response of the ith model fiber across an ensemble of different masker waveforms and across a set of vowel and masker levels (for the roving-level paradigm), i.e., $\bar{r}_i(t|f) = E_w[r_i(t|f,w)]$, and $\bar{\dot{r}}_i(t/f) = E_w[\dot{r}_i(t|f,w)] = E_w[\partial\{r_i(t|f,w)\}/\partial f]$ is the averaged partial derivative of the response of the ith fiber. The partial derivative was estimated computationally as the difference between the rate functions in response to two slightly different second formant frequencies, divided by the difference between the formant frequencies. (A formant-frequency difference of 1 Hz was used in these calculations.) In the simulations presented here, an ensemble of 220 noise tokens (20 different noises at each of 11 SPLs within the rove range) was used to estimate the expected responses of the AN fibers to the vowel plus noise stimuli.

The complete derivation of Eq. (4) is rather extensive (see Chap. 5 of Heinz, 2000); however, it is possible to provide some intuitive explanation for the equation's form. The numerator of Eq. (4) is proportional to the change in the AN fiber's rate for small changes in formant frequency; thus, a larger partial derivative of the rate function with respect to formant frequency leads to higher sensitivity. The final sensitivity, though, depends not only on changes in the decision variable but also on its variance [see Eq. (2)]. The denominator in Eq. (4) thus represents the total variance of the decision variable; the first term in the denominator of Eq. (4) corresponds to the variance due to the Poisson randomness of the AN discharges (internal noise), and the second term corresponds to the variance introduced by the random noise maskers that vary across waveforms (external noise).

Finally, the relationship between the central processor's sensitivity and behaviorally measured thresholds must be established to compare model predictions to behavioral results. The form of the equation for Q [see Eq. (2)] is equivalent to

that for the square of the commonly used sensitivity index, d'. A conventional value of the sensitivity index associated with threshold (e.g., in a two-alternative forced-choice experiment) is $d'=1$, or equivalently $Q=1$. By setting Q equal to 1 in Eq. (3), the threshold, or just-noticeable difference (jnd), for the formant frequency can be identified. The model threshold corresponding to $Q(f, f+\Delta f_{\text{jnd}})=1$ is

$$\Delta f_{\text{AN}} = \frac{1}{\delta'(f)}. \tag{5}$$

This model threshold can then be directly compared to psychophysical thresholds for discrimination of formant frequency.

The above equations included both the rate and temporal information from observations of AN responses. Predictions can also be computed with only the average-rate information of the AN responses by assuming that the central processor knows only the number of discharges for each trial. In this case, the response of the ith fiber was assumed to be a stationary Poisson process with an average discharge rate of $R_i(f,w)$. The predicted performance based only on the information in the average response rate of the AN model fibers is given by [Eq. 5.36 in Heinz (2000)]:

$$[\delta'(f)]^2 = \frac{\{\sum_{i=1}^{M}([1/\overline{R_i}(f)][\overline{\dot{R}_i}(f)]^2)\times T\}^2}{(\sum_{i=1}^{M}([1/\overline{R_i}(f)][\overline{\dot{R}_i}(f)]^2)\times T + \text{Var}_w\{\sum_{i=1}^{M}[[\overline{\dot{R}_i}(f)]/\overline{R_i}(f)]\times R_i(f,w)\times T\})}. \tag{6}$$

The form of Eq. (6) is the same as that of Eq. (4), except that the time-varying rate functions have been replaced by average rates. The model threshold obtained with only average rate information was computed using Eqs. (5) and (6), and this model threshold was then directly compared to psychophysical thresholds.

D. Threshold predictions based on coincidence detection

Cross-frequency coincidence detection has been proposed as a mechanism to detect temporal cues encoded in sound stimuli (Heinz *et al.*, 2001b; Carney *et al.*, 2002; Colburn *et al.*, 2003). It has been shown that cross-fiber coincidence counts do not successfully predict trends in psychophysical thresholds for discriminating the center frequency of harmonic complexes (Tan, 2003). However, the coincidence-detection mechanism has been reported to successfully predict psychophysical thresholds for detection of tones in the presence of background noise (Carney *et al.*, 2002). Thus the ability of the coincidence-detection mechanism to discriminate formant frequencies in the presence of a noise masker was quantitatively evaluated here.

For simplicity, the outputs of the coincidence detectors were assumed to be independent nonstationary Poisson processes. The quantification methods were adapted from those used for the AN model fiber predictions [Eq. (4)], except that the total sensitivity was accumulated over the population of coincidence detectors instead of the population of AN model fibers:

$$[\delta'(f)]^2 = \frac{\{\sum_i\sum_j\int_0^T[1/\overline{C_{ij}}(t|f)][\overline{\dot{C}_{ij}}(t|f)]^2dt\}^2}{(\sum_i\sum_j\int_0^T[1/\overline{C_{ij}}(t|f)][\overline{\dot{C}_{ij}}(t|f)]^2\,dt + \text{Var}_w\{\sum_i\sum_j\int_0^T[\overline{\dot{C}_{ij}}(t|f)/\overline{C_{ij}}(t|f)]C_{ij}(t|f,w)dt\})}. \tag{7}$$

$C_{ij}(t|f,w)=\sum_{l=1}^{K_i}\sum_{m=1}^{K_j}F(t_l^i-t_m^j,w)$ describes the coincidence detector that receives the responses to waveform w of the ith and the jth AN model fibers, where F is a brief (e.g., 20 μs) rectangular coincidence window with unity height. K_i and K_j are the total number of discharges on the i^{th} and j^{th} model fibers. As above, the partial derivatives with respect to the second formant frequency are indicated by dots in Eq. (7), and averages across different vowel-plus-masker waveforms and levels (for the roving-level paradigm) are indicated by overbars. The time-varying rate function of each coincidence detector $[C_{ij}(t)]$ was estimated based on the products of the time-varying rate functions of the convergent model AN fibers (Colburn, 1969, 1977; Heinz *et al.*, 2001b). The product of the rate function and the time step used in the simulations is equal to the probability of a discharge within a time step, and the

probability of a response in the coincidence detector is the product of the response probabilities of its two inputs, thus

$$C(t) = p_{ij}(t)/\Delta t = [p_i(t)p_j(t)]/\Delta t = [r_i(t)\Delta t][r_j(t)\Delta t]/\Delta t = r_i(t)r_j(t)/\Delta t, \quad (8)$$

where $p(t)$ is the time-varying probability of discharge, $C(t)$ and $r(t)$ are time-varying rate functions, and Δt is the time step used in the simulation. (Note that these time-varying functions are actually implemented in discrete time, but are shown as continuous-time functions for simplicity.)

The population of coincidence detectors consisted of model cells that received all pairwise combinations of the range of model AN CFs used in the simulation. In this study, each AN fiber innervated a single coincidence detector and the response of each model coincidence detector was simulated using one pair of model AN fibers. Thus the 3782 HSR AN fibers with CFs in the frequency range of interest innervated 1891 coincidence detectors. As stated above, the responses of these 3782 AN fibers were simulated by AN models of 50 different CFs, and these 50 AN models resulted in $(50^2+50)/2=1275$ distinct pairwise CF combinations as the inputs to the model coincidence detectors. This expression excludes the redundant pairwise combinations of AN CFs but retains coincidence detectors with same-CF inputs. Each of these simulated CF combinations represented about one or two coincidence detectors $(1891/1275=1.5)$.

Formant-frequency discrimination thresholds were predicted by using the rate-plus-timing information in the responses of coincidence detectors and by using only the rate information in these responses. For the rate-only predictions, coincidence-detector counts were analyzed in a manner similar to that used for the AN responses [see Eq. (6)]. The interpretation of the coincidence-detection predictions was as follows: Coincidence detectors are driven by temporal coincidences across fibers. If this form of the temporal information in the AN responses was most important for encoding formant frequencies, then the counts (or rate) of the model coincidence-detector cells should be as successful as the AN rate-plus-timing predictions in explaining the behavioral results. Thus, if the coincidence-detectors counts are *not* as successful as the AN rate-plus-timing predictions, then the coincidence detectors are *not* extracting the temporal information that is critical for explaining the thresholds. Predictions based on both rate and timing of the coincidence detectors illustrate the extent to which timing information is preserved (but not "extracted" into response rates) by these model cells.

III. RESULTS

Figure 2 illustrates model and behavioral thresholds as a function of SNR. The goal of this work was to model the trends in performance, not absolute threshold values. Thresholds for optimal detectors are often significantly lower than empirical thresholds, but if the correct trends across different stimulus conditions are observed, then straightforward methods can be used to elevate the thresholds to match the data (e.g., by making the model less optimal). However, if a model predicts thresholds that are either higher than observed thresholds or have incorrect trends across conditions, then that model can be eliminated.

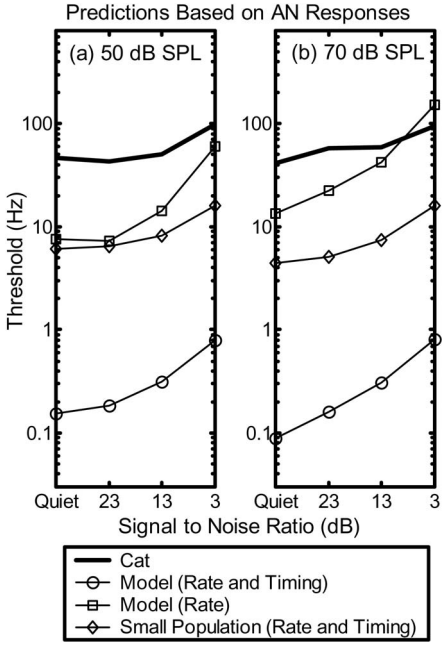

FIG. 2. Formant-frequency discrimination thresholds at various SNRs. Stimulus (combination of vowel and noise) levels were randomly varied (roved) with level uniformly distributed over a ±5-dB range in increments of 1 dB. Average vowel levels were (a) 50 dB SPL rms and (b) 70 dB SPL rms. Thick lines are cat performances, no lesions [(Hienz *et al.*, 1998), open bars of their Fig. 3]. Model threshold predictions were based on rate-only information for the population of model fibers with CFs between 1200 and 2500 Hz (squares), on both rate and timing information for the same population of model fibers (circles), and on both rate and timing information of a small group ($N=76$) of model fibers all having the same CF of 1720 Hz (diamonds).

Cat performance (thick line in Fig. 2, from Hienz *et al.*, 1998[2]) shows similar thresholds in quiet, in low-level noise, and in medium-level noise (the curve is almost flat), and the threshold in high-level noise is about twice as high as for the other noise levels. In contrast, the model's predicted thresholds based on the average response rate of the AN model fibers (squares) increased progressively as noise level increased across the entire range of noise levels tested. Performance based on both average rate and timing information of the AN model fibers (circles, based on all the model fibers with CFs between 1200 and 2500 Hz) also showed an increasing trend in thresholds across all noise levels. However, the predicted thresholds based on a single frequency channel (diamonds, the set of AN fibers with CF equal to 1720 Hz) were similar for quiet and for low and medium noise levels, and then increased for the high noise level, in agreement with the cat performance. This improved similarity between performance predicted by the responses of a subset of AN model fibers and psychophysical data is consistent with previous studies (Young and Sachs, 1979; Delgutte and Kiang, 1984a, b; Miller *et al.*, 1987; Tan and Carney, 2005).

The threshold predictions shown in Fig. 2 were made using the roving-level paradigm used by Hienz *et al.* (1998) for the behavioral tests. Model thresholds were also estimated using fixed-level stimuli to investigate the influence of

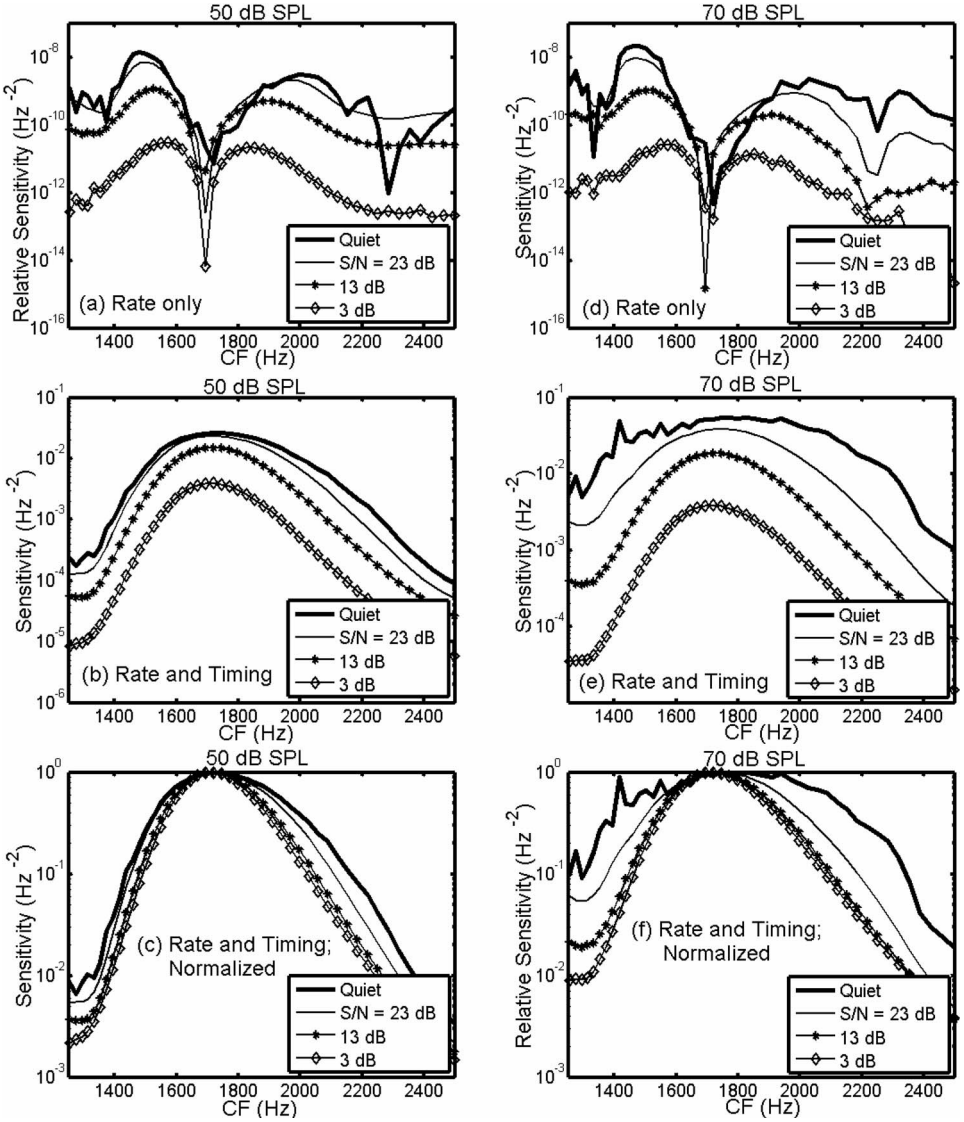

FIG. 3. Sensitivities computed for individual CF channels [Eqs. (4) and (6) computed for one CF at a time]. Roving vowel plus masker levels averaged 50 (left column) and 70 dB SPL (right column), with SNRs of 3 dB (diamonds), 13 dB (asterisks), 23 dB (thin lines), and quiet (bold). Results are based on the model using average rate information (a and d), or using both rate and timing information [(b) and (e)]. Panels (c) and (f) show the sensitivity metric (rate and timing information) normalized by the peak values to better illustrate changes in the width of the curves across SNRs.

the rove on model predictions (not shown). As expected, thresholds based on rate were most affected by the rove for vowels in quiet (thresholds were increased by about 150% due to the rove), and the elevation of threshold due to the rove dropped to about 10% for the lowest SNR studied (3 dB) for both 50 and 70 dB SPL vowels. Model thresholds based on both rate and timing were relatively unaffected by the rove; thresholds were elevated by approximately 1% or less for all SNRs for predictions based on the AN population, and by less than 1% for predictions based on the subset of fibers tuned near the formant frequency for the 50-dB vowels. The effect of the rove was slightly larger for the 70-dB SPL vowels (up to 2% threshold elevation for vowels in quiet based on rate and timing of the subset of model fibers).

The sensitivity of each AN model fiber for predictions made with average-rate information was calculated by evaluating Eq. (6) for individual CF channels (i.e., sensitivity was computed separately for each of the 50 different AN model CFs) and examining the results across CF [Figs. 3(a) and 3(b)]. (Note that Fig. 3 illustrates how sensitivity varies as a function of CF, but the actual contributions of the individual channels to the sensitivity predicted with the population response [Fig. 2; Eq. (4)] also depend upon the combined variance computed from the responses of all the included AN channels, i.e., the second term in the denominator of Eq. (4).) In general, Fig. 3 illustrates that sensitivity decreased quickly as the noise level increased, although there was a smaller difference between the sensitivities in quiet and low-level noise.

The drop in sensitivity for model responses based on rate at CFs around 1700 Hz [Figs. 3(a) and 3(d)] was due to the greater sensitivity of AN rates to changes in energy on the skirts of their tuning curves as compared to the peaks, where rates begin to saturate at mid to high stimulus levels (e.g., Whitfield, 1967; Siebert, 1968). The response rates for CFs tuned near the formant frequency are relatively high, as expected [Figs. 4(a) and 4(c)]; however, the *changes* in rate for fibers tuned near the formant peak as formant frequency is varied are relatively small [Figs. 4(b) and 4(d)]. The largest changes in rate for small changes in formant frequency occur for fibers tuned 100–200 Hz below or above the formant frequency [Figs. 4(b) and 4(d)]; therefore, these fibers contribute the most average-rate information [Figs. 3(a) and 3(d)]. The rates near the formant frequency are not com-

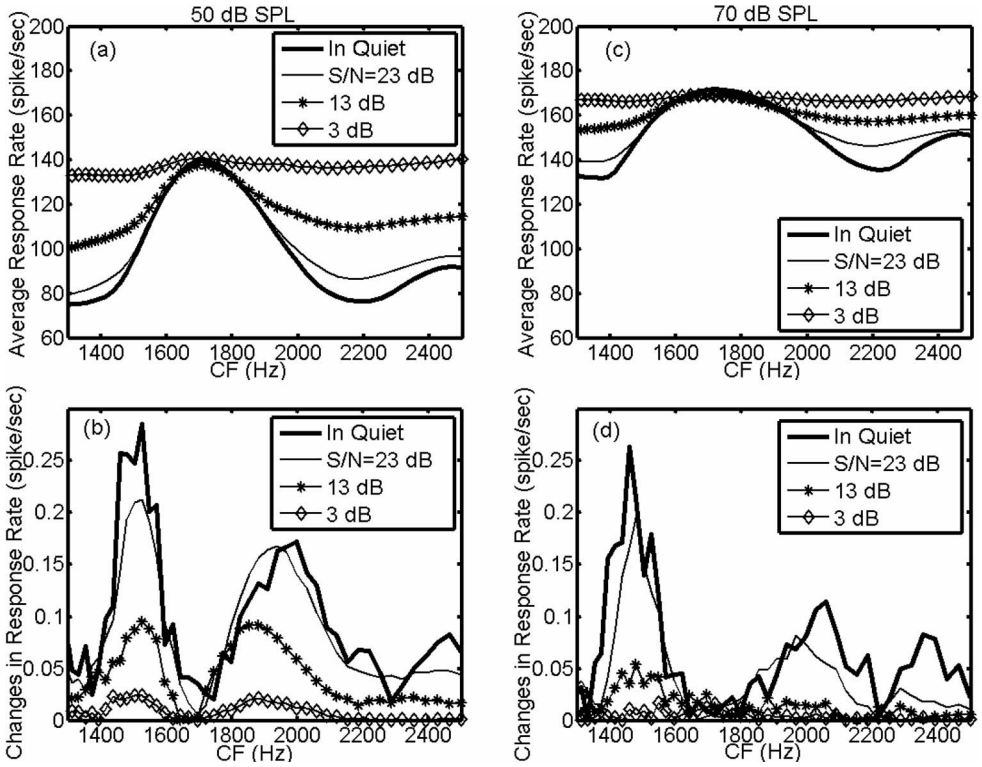

FIG. 4. Average model AN response rates [(a) and (c)] and changes in these rates [(b) and (d)] at SNRs of 3 dB (diamonds), 13 dB (asterisks), 23 dB (thin line), and in quiet (bold line) with fixed vowel levels of 50 dB SPL [(a) and (b)] and 70 dB SPL [(c) and (d)]. Changes in average AN response rates are for a 1-Hz increase (from 1700 to 1701 Hz) in the second formant frequency.

pletely saturated (i.e., rates for CFs near the formant frequency are higher in response to the 70-dB SPL vowel than in response to the 50-dB SPL vowel); however, the decreasing slope of AN rate-level functions as level increases toward saturation means that the sensitivity of these fibers decreases as level increases. In addition to this effect of saturation on fibers stimulated by the higher levels near formant peaks, the higher rates of these fibers are associated with higher variance, which also limits the sensitivity of the average discharge rate (Colburn *et al.*, 2003).

The sensitivity of each AN model fiber when using both average rate and timing information at each signal-to-noise ratio is shown in Figs. 3(b) and 3(e) and is found by computing the sensitivity using Eq. (4) for each CF channel. The sensitivity of the AN model fibers near F2 (1700 Hz) was always higher than that of other model fibers at each noise level. The sensitivity curve was wider in quiet (bold line) or in low-level noise (thin line) than in high-level noise (diamonds). The change in width of the relative sensitivity plots is made clearer by showing results normalized to the peak sensitivity [Figs. 3(c) and 3(f)]. The responses of the model AN fibers with CFs away from F2 were increasingly dominated by the masker noise as noise level increased [see Figs. 4(a) and 4(c)]. This result is consistent with the suggestion that when the whole population of AN model fibers was used in the predictions, more fibers contributed significantly in quiet or at low noise levels than at high noise levels (Miller *et al.*, 1987).

Figure 5 shows predicted thresholds based on the coincidence-detection mechanism [Eq. (7)]. If coincidence detection extracts the timing information that is relevant for this formant-frequency discrimination task, then the rate (or counts) of coincidence detectors should provide better predictions of the trends in the thresholds than do the rates of

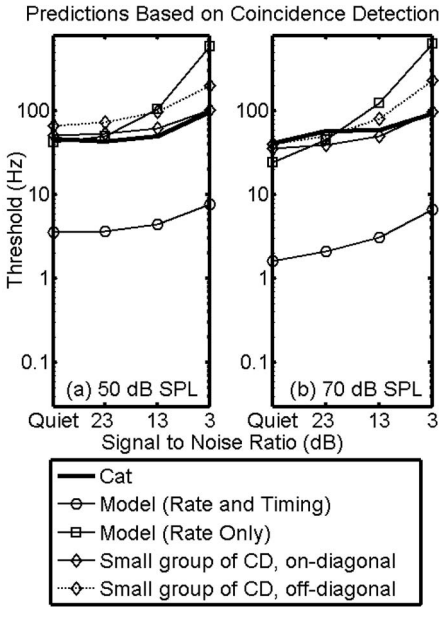

FIG. 5. Thresholds for formant-frequency discrimination based on the coincidence-detection mechanism [Eq. (7)] at various SNRs. Stimulus (combination of vowel and noise) levels were randomly varied (roved) with level uniformly distributed over a ±5-dB range, in increments of 1 dB. Average vowel levels were (a) 50 and (b) 70 dB SPL rms. Cat performance is shown with thick lines [Hienz *et al.* (1998), open bars of their Fig. 3]. Model threshold predictions used only rate information for the population of coincidence detectors that received inputs from model AN fibers with CFs between 1200 and 2500 Hz (squares), used both rate and timing information for the same population of coincidence detectors (circles), and used both rate and timing information of small groups of coincidence detectors [on- and off-diagonal, see Fig. 6(a)] (diamonds).

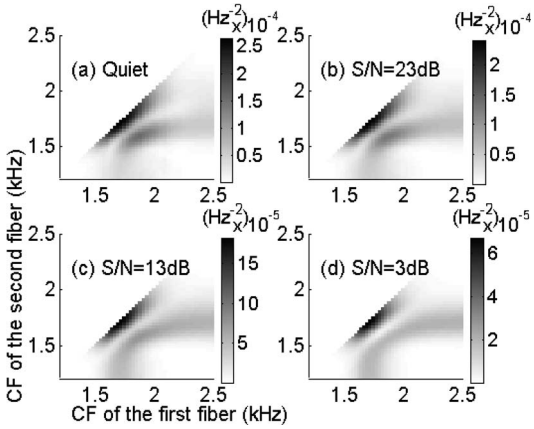

FIG. 6. Sensitivities of coincidence detectors to change in the second formant frequency. Roving-level vowel and masker levels were uniformly distributed over a ±5-dB range, with 1-dB increments and an average level of 50 dB SPL rms. Each panel corresponds to one SNR: (a) quiet, (b) 23 dB, (c) 13 dB, and (d) 3 dB.

AN fibers. The coincidence-detection threshold predictions based on count (Fig. 5, squares), however, are very similar to those based on AN rate (Fig. 2), suggesting that the coincidence detectors are not using the temporal information in the AN population (i.e., converting AN timing information into coincidence detector counts).

The trend in predicted thresholds obtained using both count and timing information of a large population of coincidence detectors (Fig. 5, circles) is nearly flat for quiet, low, and medium background noise levels, and the threshold at the high noise level is about two times the threshold at the medium noise level. This trend agrees with the psychophysical data, which simply suggests that the coincidence detectors preserve timing information present in the AN response. Thus, the timing information that contributes to the predictions based on AN responses is not lost in the process of coincidence detection, but it is apparently not "used" by the coincidence detectors, either (in which case it would be reflected in the coincidence-detector counts).

Thresholds were also predicted using the rate and timing information of subsets of the coincidence-detector population. The selection of the smaller groups of coincidence detectors was based on the results shown in Fig. 6, which demonstrates the sensitivity of each coincidence detector to the formant-frequency change. In general, each panel of Fig. 6 (corresponding to one signal-to-noise ratio) depicts two groups of coincidence detectors with relatively high sensitivity. The first group is the one on the diagonal, which represents coincidence detectors that receive inputs from two AN fibers that have matched CFs. The other group is away from the diagonal, representing coincidence detectors that receive inputs from two AN fibers with different CFs, and it has lower sensitivity than the first group. The coincidence detector with the highest sensitivity in each of these two groups was used to compute the formant-frequency discrimination thresholds. In Fig. 5, the solid line with diamonds corresponds to the performance of the most sensitive coincidence detector from the first group (on-diagonal, with matched AN CFs of 1745 Hz), and the dashed line with diamonds corresponds to the performance of the coincidence-detector from the second group (off-diagonal, convergence of ANs with CFs of 1595 and 1881 Hz). The performance of each of these coincidence detectors, based on both count and timing information, shows a trend that is similar to the cat performance and to the prediction with the original population of coincidence detectors (line with circles in Fig. 5). As shown above for the AN model predictions, these results suggest that a subset of the model coincidence-detector cells can predict the trends in the behavioral performance; however, the fact that both rate and timing of the coincidence detectors had to be included again illustrates that the coincidence detectors were not extracting the critical temporal information from the AN responses. Predictions that used only the rate of a subset of the coincidence detectors (not shown) would have been even higher than the rate-based predictions for the whole population of model coincidence detectors, and therefore higher than the behavioral thresholds, especially for the lower SNRs.

IV. DISCUSSION

This study provides predictions of performance limits in formant-frequency discrimination that used the response patterns of AN model fibers and responses of cross-frequency coincidence detectors. Quantitative tools developed to evaluate coding strategies for simple acoustic stimuli, such as tones and tones in noise (Heinz, 2000), were applied here to complex stimuli. The results of this study provide insight into the relative contributions of rate and timing information to speech coding in noise. When the background noise level increased, the thresholds predicted by using average-rate information of model AN fibers increased more than behavioral thresholds did, whereas model thresholds based on both rate and timing information increased in a manner more consistent with the behavioral data (Fig. 2). This result suggests that timing information was required to explain trends in psychophysical performance in this task.

A prominent feature of the AN model sensitivity patterns based on rate information was the dip at CFs near 1700 Hz [Figs. 3(a) and 3(d)], in contrast to the simple peak at 1700 Hz in the sensitivity pattern based on temporal and rate information [Figs. 3(b) and 3(e)]. The dip in the rate-based sensitivity was further explored by examining the average-rate responses (Fig. 4). The dominant factor in the decrease in sensitivity for fibers tuned near the formant frequency was the reduced *change* in rate for different formant frequencies for these fibers due to rate saturation. The model fiber sensitivity to the formant-frequency change was significantly reduced by high-level background noise [Figs. 3(a), 3(d), 4(b), and 4(d)] when predictions were based on only average-rate information (Young and Barta, 1986). Thus, as has been seen in previous studies (e.g., Heinz *et al.*, 2001a, b; Colburn *et al.*, 2003), the sensitivity pattern across a population of AN fibers is typically more complex for rate-based information [Figs. 3(a) and 3(d)] than it is for temporally based information [Figs. 3(b) and 3(e)].

For the predictions based on both rate and timing information, the sensitivity of AN model fibers as a function of

CF [Figs. 3(b) and 3(d)] indicated that the model fibers with CFs near but not too close to the second formant frequency may have contributed to the overall sensitivity at low noise levels, yet they contributed relatively less when the noise level was high because their response was overwhelmed by the high-level noise. Because of this behavior, the prediction based on a large population of model fibers degraded more for high-level noise than did the prediction based on a small subset of fibers [Figs. 3(c) and 3(f)]. As suggested by our previous results and by others, it is realistic and reasonable to predict psychophysical thresholds with a small population of AN model fibers with CFs close to the signal frequency when the task involves discrimination of spectral changes over a small frequency range (Young and Sachs, 1979; Sruloviczand Goldstein, 1983; Delgutte and Kiang, 1984a, b; Miller *et al.*, 1987; Tan and Carney, 2005).

The coincidence-detection mechanism did not specifically extract the temporal cues from the AN model responses related to the formant-frequency discrimination task. If it had, then the coincidence detection predictions based on rate (or count) would have explained the behavioral results as well as the AN predictions based on both rate and timing. In fact, the threshold trend of the coincidence-detector count prediction degraded more with decreasing SNR than did the prediction based on both rate and timing information in AN responses, and it was similar to the threshold trend obtained using only the average rate of the AN responses. Future studies must therefore consider other temporal aspects of the AN responses, such as information conveyed by phase locking to the envelope of the stimulus after narrow-band filtering in the auditory periphery (Tan and Carney, 2005).

An interesting question regarding these results concerns the change in the predicted thresholds, and thus in the model AN responses, at the highest noise level, that is, why does the threshold based on both temporal and rate measures increase noticeably at the highest noise level, especially as compared to the threshold based on rate only, which changes more gradually across all noise levels? Synchrony information is generally robust across a wide range of sound levels, even in response to signals in the presence of noise (Rhode *et al.*, 1978). However, the temporal responses related to the envelope of the vowel complex would be expected to change as noise level increases, because the bandwidth of the peripheral filters increases with sound level and because adaptation shapes the responses of AN fibers. The degree of modulation in response to the signal in the AN responses is reduced as level increases due to rate saturation (Smith and Brachman, 1980; Joris and Yin, 1992), and filter broadening due to the added noise tends to further reduce the largest envelope modulations in the AN responses [i.e., wideband noises have a broader amplitude-modulation spectra, with lower amplitudes at low frequencies, as compared with narrow-band noises, which have narrower and higher amplitude-modulation spectra, especially at low modulation frequencies (Lawson and Uhlenbeck, 1950; Dau *et al.*, 1997)]. The representation of modulation cues and the influence of masking noise on these cues in AN responses to speech sounds are topics for future study.

At every SNR, the threshold predicted by the coincidence-detection mechanism was higher than the prediction based on the response patterns of the AN model fibers, as expected, because the coincidence-detection process decreased the total amount of information carried by the AN model fiber response patterns. Similarly, for both the AN model fiber and the coincidence-detector predictions, thresholds based on smaller populations were always higher than the thresholds based on the full set of model fibers or coincidence detectors. This increase in thresholds was simply due to the fact that the amount of information is related to the number of fibers or cells included in the population response [Eqs. (1), (4), and (6)].

The work presented here was all based on formant-frequency discrimination of F2 in synthesized speech stimuli. It would be interesting to test predictions for lower (first formant) or higher (third formant) frequencies. To extend the study to higher frequencies, it would be important to include nonsaturating low-spontaneous-rate AN fibers in the simulations, because temporal information conveyed by the synchronization of AN fibers rolls off at high frequencies (Johnson, 1980). The distinct roles of rate and temporal coding in these frequency ranges and the diverse contributions of different spontaneous-rate populations of AN fibers (e.g., Heinz *et al.*, 2001b) suggest that different aspects of the AN model responses may dominate threshold predictions for speech cues in different frequency ranges.

ACKNOWLEDGMENTS

This work was supported by NIH-NIDCD Grant No. R01-01641. We gratefully acknowledge the comments of Sean Davidson, Michael Heinz, Paul Nelson, and two very helpful reviewers on a previous version of this manuscript and the editorial assistance of Susan Early.

[1]Computer code used for the simulations presented here is available at http://web.syr.edu/~lacarney

[2]Cat thresholds illustrated in the figures for comparison to model thresholds were all based on Hienz *et al.*'s (1998) results for cats without olivocochlear bundle (OCB) lesions. The AN model (Tan and Carney, 2003) was based on AN recordings made in cats without OCB lesions. The role of the OCB (or lesions of the OCB) in AN responses to speech, in either awake or anesthetized animals or in behavioral responses to speech, is still not completely understood; therefore, it seemed most straightforward to focus comparisons on results for intact animals.

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

Colburn, H. S. (1969). "Some physiological limitations on binaural performance," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Colburn, H. S. (1977). "Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise. Supplementary material," AIP Document No. PAPS JASMA-61–525–98.

Colburn, H. S., Carney, L. H., and Heinz, M. G. (**2003**). "Quantifying the information in auditory-nerve responses for level discrimination," J. Assoc. Res. Otolaryngol. **04**, 294–311.

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

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

Delgutte, B., and Kiang, N. Y.-S. (**1984b**). "Speech coding in the auditory

nerve: V. Vowels in background noise" J. Acoust. Soc. Am. **75**, 908–918.
Flanagan, J. L. (**1955**). "A difference limen for vowel formant frequency," J. Acoust. Soc. Am. **27**, 613–617.
Hawks, J. W. (**1994**). "Difference limens for formant patterns of vowel sounds," J. Acoust. Soc. Am. **95**, 1074–1084.
Heinz, M. G. (2000). "Quantifying the effects of the cochlear amplifier on temporal and average-rate information in the auditory nerve," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.
Heinz, M. G., Colburn, H. S., and Carney, L. H. (**2001a**). "Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve," Neural Comput. **13**, 2273–2316.
Heinz, M. G., Colburn, H. S., and Carney, L. H. (**2001b**). "Rate and timing cues associated with the cochlear amplifier: Level discrimination based on monaural cross-frequency coincidence detection," J. Acoust. Soc. Am. **110**, 2065–2084.
Heinz, M. G., Colburn, H. S., and Carney, L. H. (**2002**). "Quantifying the implications of nonlinear cochlear tuning for auditory-filter estimates," J. Acoust. Soc. Am. **111**, 996–1011.
Hienz, R. D., Aleszczyk, C. M., and May, B. J. (**1996**). "Vowel discrimination in cats: Acquisition, effects of stimulus level, and performance in noise," J. Acoust. Soc. Am. **99**, 3656–3668.
Hienz, R. D., Stiles, P., and May, B. J. (**1998**). "Effects of bilateral olivocochlear lesions on vowel formant discrimination in cats," Hear. Res. **116**, 10–20.
Johnson, D. H. (**1980**). "The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones," J. Acoust. Soc. Am. **68**, 1115–1122.
Joris, P. X., and Yin, T. C. T. (**1992**). "Responses to amplitude-modulated tones in the auditory nerve of the cat," J. Acoust. Soc. Am. **91**, 215–232.
Keithley, E. M., and Schreiber, R. C. (**1987**). "Frequency map of the spiral ganglion of cat," J. Acoust. Soc. Am. **81**, 1036–1042.
Kewley-Port, D., and Watson, C. S. (**1994**). "Formant-frequency discrimination for isolated English vowels," J. Acoust. Soc. Am. **95**, 485–496.
Kewley-Port, D., Li, X., Zheng, Y., and Neel, A. (**1996**). "Fundamental frequency effects on thresholds for vowel formant discrimination," J. Acoust. Soc. Am. **100**, 2462–2470.
Klatt, D. H. (**1980**). "Software for a cascade/parallel formant synthesizer," J. Acoust. Soc. Am. **67**, 971–995.
Lawson, J. L., and Uhlenbeck, G. E. (**1950**). *Threshold Signals*, Vol. **24** of *Radiation Laboratory Series* (McGraw-Hill, New York).
Liberman, M. C. (**1978**). "Auditory-nerve response fromcats raised in a low-noise chamber," J. Acoust. Soc. Am. **63**, 442–455.
Liberman, M. C. (**1982**). "The cochlear frequency map for cat: Labeling auditory-nerve fibers of known characteristic frequency," J. Acoust. Soc. Am. **72**, 1441–1449.
Liu, C., and Kewley-Port, D. (**2004**). "Formant discrimination in noise for isolated vowels," J. Acoust. Soc. Am. **116**, 3119–3129.
Lyzenga, J., and Horst, J. W. (**1995**). "Frequency discrimination of bandlimited harmonic complexes related to vowel formants," J. Acoust. Soc. Am. **98**, 1943–1955.
Mermelstein, P. (**1978**). "Difference limens for formant frequencies of steady-state and consonant-bound vowels," J. Acoust. Soc. Am. **63**, 572–580.
Miller, M. I., Barta, P. E., and Sachs, M. B. (**1987**). "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**, 665–679.
Rhode, W. S., Geisler, C. D., and Kennedy, D. K. (**1978**). "Auditory-nerve fiber responses to wide-band noise and tone combinations," J. Neurophysiol. **41**, 692–704.
Sachs, M. B., and Young, E. D. (**1979**). "Encoding of steady-state vowels in the auditory nerve: Representation in terms of discharge rate," J. Acoust. Soc. Am. **66**, 470–479.
Sachs, M. B., Voigt, H. F., and Young, E. D. (**1983**). "Auditory-nerve representation of vowels in background noise," J. Neurophysiol. **50**, 27–45.
Siebert, W. M. (**1965**). "Some implication of the stochastic behavior of primary auditory neurons," Kybernetika **2**, 206–215.
Siebert, W. M. (**1968**). "Stimulus transformation in the peripheral auditory system," in *Recognizing Patterns*, edited by P.A. Kolers and M. Eden (MIT, Cambridge, MA), pp. 104–133.
Siebert, W. M. (**1970**). "Frequency discrimination in the auditory system: place or periodicity mechanisms?" Proc. IEEE **58**, 723–730.
Smith, R. L., and Brachman, M. L. (**1980**). "Response modulation of auditory-nerve fibers by AM stimuli: Effects of average intensity," Hear. Res. **2**, 123–133.
Srulovicz, P., and Goldstein, J. L. (**1983**). "The central spectrum: A synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum," J. Acoust. Soc. Am. **73**, 1266–1276.
Tan, Q. (2003). "Computational and statistical analysis of auditory peripheral processing for vowel-like signals," Ph.D. dissertation, Boston University.
Tan, Q., and Carney, L. H. (**2003**). "A phenomenological model for the responses of auditory-nerve fibers: II. Nonlinear tuning with a frequency glide," J. Acoust. Soc. Am. **114**, 2007–2020.
Tan, Q., and Carney, L. H. (**2005**). "Encoding of vowel-like sounds in the auditory nerve: Model predictions of discrimination performance," J. Acoust. Soc. Am. **117**, 1210–1222.
Whitfield, I. C. (**1967**). *The Auditory Pathway* (Edward Arnold, London).
Wier, C. C., Jesteadt, W., and Green, D. M. (**1977**). "Frequency discrimination as a function of frequency and sensation level," J. Acoust. Soc. Am. **61**, 178–184.
Young, E. D., and Barta, P. E. (**1986**). "Rate responses of auditory-nerve fibers in noise near masked threshold," J. Acoust. Soc. Am. **79**, 426–442.
Young, E. D., and Sachs, M. B. (**1979**). "Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers," J. Acoust. Soc. Am. **66**, 1381–1403.