Adaptive Stimulus Optimization for Auditory Cortical Neurons

Kevin N. O’Connor,1,2 Christopher I. Petkov,1 and Mitchell L. Sutter1,2
1Center for Neuroscience and the 2Section for Neurobiology, Physiology and Behavior, University of California, Davis, California

Submitted 14 January 2005; accepted in final form 24 August 2005

O’Connor, Kevin N., Christopher I. Petkov, and Mitchell L. Sutter. Adaptive stimulus optimization for auditory cortical neurons. J Neurophysiol 94: 4051–4067, 2005. First published August 31, 2005; doi:10.1152/jn.00046.2005. Despite the extensive physiological work performed on auditory cortex, our understanding of the basic functional properties of auditory cortical neurons is incomplete. For example, it remains unclear what stimulus features are most important for these cells. Determining these features is challenging given the considerable size of the relevant stimulus parameter space as well as the unpredictable nature of many neurons’ responses to complex stimuli due to nonlinear integration across frequency. Here we used an adaptive stimulus optimization technique to obtain the preferred spectral input for neurons in macaque primary auditory cortex (AI). This method uses a neuron’s response to progressively modify the frequency composition of a stimulus to determine the preferred spectrum. This technique has the advantage of being able to incorporate nonlinear stimulus interactions into a “best estimate” of a neuron’s preferred spectrum. The resulting spectra displayed a consistent, relatively simple circumscribed form that was similar across scale and frequency in which excitation and inhibition appeared about equally prominent. In most cases, this structure could be described using two simple models, the Gabor and difference of Gaussians functions. The findings indicate that AI neurons are well suited for extracting important scale-invariant features in sound spectra and suggest that they are designed to efficiently represent natural sounds.

INTRODUCTION

It is commonly thought that auditory cortical (AC) neurons encode certain sound features, but exactly what these features are remains uncertain. Determining what stimulus features are important to a cortical neuron can be a very difficult task given the potentially enormous size of the relevant parameter space. This problem is even more difficult because of the complex stimulus interactions that might affect the cell’s response. For example, a neuron’s response to a spectrally complex sound is not necessarily equal to the linear sum of its responses to the individual components taken in isolation (Abeles and Goldstein 1972; Kadia and Wang 2003; Katsuki et al. 1999; Nelken et al. 1999; Theunissen et al. 2000). Conventional methods such as reverse correlation (Blake and Merzenich 2002; deCharms et al. 1998; Depireux et al. 2001; Klein et al. 2000; Miller et al. 2002; Rutkowski et al. 2002; Theunissen et al. 2000) that essentially compute a spike-triggered average estimate of the best stimulus are not well equipped to deal with this problem, readily revealing only the linear interactions among stimulus components.

Here we describe and employ an adaptive stimulus optimization technique that manipulates the spectral composition of a stimulus, a multi-tone complex, in an attempt to maximize responses from a single neuron. It relies on feedback from the neuron to explore a multi-dimensional parameter space, searching for the best stimulus in that space. It estimates a neuron’s preferred spectral input, regardless of the spectral complexity of the stimulus. It also has the advantage of significantly reducing the size of the potential parameter space because it focuses on the portion of the space that is important to the cell. In this study, adaptive stimulus optimization provided a direct and efficient way to determine the relevant stimulus features and basic functional properties for AC neurons. The results also demonstrate the tremendous potential of the adaptive stimulus optimization technique.

The resulting preferred spectra displayed a structural simplicity and consistency not previously reported for AC neurons. The spectra possessed a nearly scale-invariant, prototypical form having a relatively simple quantitative description. This structure appears well suited for identifying important spectral features and for efficiently representing the information in natural sounds.

METHODS

Subjects

Two adult rhesus monkeys, (Macaca mulatta; 1 male, 1 female) with normal hearing, on a restricted water access protocol, were subjects. All procedures performed on the subjects conformed to the PHS policy on experimental animal care and were approved by the UC Davis animal care and use committee.

Electrophysiological recording and data acquisition

Each monkey was implanted with a head post and chronic recording chamber for access to auditory cortex. Recordings were made while the monkeys were comfortably restrained and sitting quietly in an acoustically “transparent” primate chair within a sound-attenuated, foam-lined booth (IAC: 9.5 × 10.5 × 6.5 foot). Subjects received diluted fruit juice or water intermittently. High-impedance tungsten microelectrodes (FHC) were inserted into the brain using a remotely controlled hydraulic microdrive (FHC) through guide tubes held by a plastic grid (Crist Instrument) in the recording chamber. Extracellular potentials were amplified and filtered (0.3–5 kHz; AM Systems 1800) and selected using a dual (amplitude-time) window discriminator (Bak RP-1). Auditory cortex was identified by single- and multiunit responses to pure-tone pips, broad- and narrow-band noise bursts, and clicks. Primary (core) auditory cortex was identifiable by the vigor and selectivity of single-unit responses to pure tones and the latency of these responses and also from the gradient of best frequency obtained along rostrocaudal and mediolateral anatomical coordinates. During adaptive optimization sessions, counts of well-isolated single-unit potentials were made during presentation of the 170-ms stimulus
perturbation in the \(i^{th}\) set, \(\bar{r}\) the average response over the \(i^{th}\) set, \(\Delta r_{\text{max}}(i)\) the maximum absolute difference from the mean across the \(i^{th}\) set, \(\nu_{ij}\) the parameter vector representing the \(j^{th}\) stimulus variant in the \(i^{th}\) set, and \(\alpha\) a weighting factor determining the magnitude of parameter perturbation. This rule can be summarized as follows: For each iteration, it determines the differences between the response to each stimulus in the set and the average response, normalizes these differences, weights each stimulus variant by its corresponding normalized response, averages across the set of stimuli and, finally, determines the new base parameter vector by weighting the average and adding the result to the previous base parameter vector. The resulting parameter vectors were then used to synthesize a new base stimulus. The base stimulus is thereby moved along a gradient in multi-dimensional parameter space, at each step moving toward the form of the stimulus that evokes the largest response.

The base stimulus and set of randomized variants (48 stimuli) were presented in random order within blocks, several times (range: 2–6) on one presentation (iteration). Iterations continued until the amplitude vector (spectrum) stabilized (1 session). The experiment was then repeated for a second optimization session, starting from different initial conditions. The mean inter-stimulus interval was \(-1.2\) s with a random uniform variation of \(\pm 0.25\) s. The time required to complete one experimental session was usually \(-1.5\) h.

For the starting base stimulus on each session, the amplitudes of all frequency components were set to the same level and all phases were randomized. For the first presentation and every iteration, a set of stimulus variants was generated by randomly perturbing the amplitudes and phases of the base tone components. The second testing session began with a different set of randomized variants than the first. The phase of each tone component was randomly advanced or delayed by \(36^\circ\) (or, in some experiments, \(45^\circ\)) from the base value. The amplitude of each frequency component was randomly increased or decreased by a constant magnitude on an iteration. The magnitude of the amplitude perturbations was gradually increased over the course of an experimental session. Typically beginning at \(6\) dB, amplitude perturbations were usually increased in \(2\)-dB steps to \(12\) dB (step size was controlled by the weighting factor \(\alpha\) in Eq. 1). This was done to avoid local maxima (the gradient ascended toward the stimulus evoking the largest response), and to counteract response habituation by presenting stimulus variants sufficiently different from the base.

In the initial experiments, the amplitude of each tone in the complex was independently varied. However, we found that better results could often be obtained if the perturbations initially occurred in segments or bins of adjacent frequencies, particularly if the spectral sensitivity of the cell was broad (see Fig. 2A). When employing this “coarse-search” technique, the number of frequency segments independently varied in amplitude was gradually increased during the session from a small set (e.g., 5 or 6) to the total number of frequencies comprising the stimulus (typically 24–36). Prior to using the coarse search strategy, the probability of obtaining a pair of optimization sessions from a cell for which the amplitude vectors were significantly correlated was slightly less than one in two (0.44) after its implementation the odds increased to greater than one in two (0.65).

The sound signals were generated using a digital signal processor (AT&T DSP32C) with 16-bit output resolution at a sampling rate of 100 kHz and a D/A converter (TDT Systems DA1), passed through a programmable attenuator (TDT Systems PA4), and then through a passive attenuator (Leader LAT-45). The signal was then amplified (Radio Shack MPA-200) and delivered through a speaker (Radio Shack PA-110, 10-in woofers and piezo-horn tweeter, 0.038–27 kHz or Radio Shack 40-1310B tweeter, 5–50 KHz) positioned at ear level 1.5 m in front of the subject. System calibration (Brueil & Kjaer 2231 meter, unfiltered calibration, 1/4-in 4133 condenser microphone) revealed that the PA-110 speaker gave a flat frequency response to within \(\pm 10\) dB from the average level between 0.05 and 25 kHz. The output from the tweeter speaker varied at most \(\pm 8\) dB from mean level between 5 and 40 kHz and declined 10 dB between 40 to 50 kHz and for 100 ms immediately after.

Experimental control and data collection and analysis was accomplished using customized C-language and Matlab (MathWorks) programs running on a personal computer.

**Stimulus generation and presentation**

The optimization stimulus was a multi-tone complex created by summing a large number of pure tones (Fig. 1A) with randomized phases. Each tone complex comprised either 12 or 16 tones per octave (typically 24–36 tones) spaced at equal log-frequency intervals. The range of frequencies was usually three octaves, but was adjusted if needed to suit the pure-tone frequency selectivity of the cell (range: 2–6 octaves). An attempt was made to center the range on the neuron’s preferred frequency as determined by pure-tone stimulation, although the range of frequencies able to evoke strong responses was often quite broad. The tone complex was temporally shaped with a Gaussian amplitude envelope with a width (at one-half-amplitude) of \(-50\) ms, producing a temporal Gabor stimulus (Fig. 1A). The intensity of the stimulus was adapted to a moderate level within the cell’s best-intensity region as estimated from the initial search. The intensity range across experiments was between \(27\) and 66 dB SPL (mean = 46.4 dB; SD = 8.1 dB; Brueil & Kjaer 2231 meter, unfiltered calibration).

Starting from an initial or base stimulus, a set of stimulus variants was created by randomly perturbing the individual amplitudes and phases of each tone component in the complex from its base value (Fig. 1B). A neuron’s response (spike count) to each stimulus in the set was used to modify the base stimulus parameters (amplitudes and phases) in the direction of the stimulus generating the largest responses. The base stimulus was updated using a rule similar to one used successfully for visual cortical neurons (Foldiak 2001)

\[
b_{i+1} = b_i + \alpha\frac{(r_i - \bar{r})/\Delta r_{\text{max}}(i))}{|b_i|}
\]

where \(b_i\) is a vector representing the values of base (amplitude or phase) parameters in stimulus set \(i\), \(r_i\) the response to the \(j^{th}\) stimulus

\[1\] The updating rule is identical to that used by Foldiak except for the normalization term, \(\Delta r_{\text{max}}(i)\).

---

**FIG. 1.** Optimization stimuli. A: initial base stimulus (bottom) was created by summing a set of shaped pure-tone pulses \((f_i - f_j)\) of equal amplitude spaced at equal log-frequency intervals. Phases were randomized. B: schematic depicts the component amplitudes of the initial base stimulus (black), along with 3 stimulus variants (red, green, and yellow). All stimuli presented during an optimization session were variants of the base formed by randomly perturbing the amplitude of each frequency in the base by some constant value, and advancing or delaying its phase (only a subset of the total number of frequency components and stimuli presented on an iteration are shown for clarity).
kHz; this measured decline may have been due to the high sensitivity limit (roll-off point) of 40 kHz for the 4133 microphone.

The stimuli were normalized with respect to digital (16-bit) signal peak amplitude. This provided a limit to overall energy level because historically the term “receptive field” has referred to the stimulus space to which a neuron is sensitive (Hartline 1938). However, recently the term has become associated with the notion of a quantitative model of a spectral or spectral-temporal receptive field (e.g., Theunissen et al. 2000). Therefore we have used the term “preferred spectrum” in this paper to avoid confusion.

In three early experiments, a slightly different set of frequencies was used for the two runs. This was done because either the initial estimated best frequency was not well centered within the frequency range or there was an attempt to obtain greater frequency resolution on the second session. In these cases, correlation coefficients were computed between amplitude pairs closest in frequency, and the final preferred spectrum was obtained by combining the resulting vectors (interleaving the amplitudes at the appropriate frequencies) rather than averaging them. In the case of two other neurons, the amplitude spectra resulting from the two sessions were closely similar, but slightly shifted in frequency (by 1 or 2 frequency steps), suggesting that, although the structure was consistent, there was some small degree of variation in its relation to absolute frequency. In these cases the correlations were computed using the shifted amplitude vectors.

Widths of the preferred spectra were estimated by measuring the distance (in Hz) between the peaks and troughs of the amplitude spectrum. If the spectrum comprised a center peak surrounded by two flanking troughs, the difference between the frequencies at the upper and lower troughs was taken as the spectral width. If a neuron’s preferred spectrum was symmetrical, the center frequency of the receptive field (RF) corresponded to the maximum peak (or minimum trough) frequency. If the spectrum was asymmetrical, then the center frequency of the receptive field was used for the two runs. This was done because either the initial operation was performed for the cells with center troughs). If only one lower trough met this criterion, the difference between the peaks and trough was taken as the spectral width. If a neuron’s preferred spectrum was obtained by combining the resulting vectors (interleaving the amplitudes at the appropriate frequencies) rather than averaging them. In the case of two other neurons, the amplitude spectra resulting from the two sessions were closely similar, but slightly shifted in frequency (by 1 or 2 frequency steps), suggesting that, although the structure was consistent, there was some small degree of variation in its relation to absolute frequency. In these cases the correlations were computed using the shifted amplitude vectors.

Data analysis

A final estimate of the preferred spectrum was obtained by averaging the amplitude vectors from the two experimental sessions, provided that the vectors were significantly correlated (Pearson $r$, $P < 0.05$, 1-tailed). As Table 1 shows, most (26) of these correlations were highly significant ($P < 0.005$). The preferred spectrum can be considered an estimate of the neuron’s spectral receptive field in the sense that it reveals the frequencies that influence a neuron’s response, and because historically the term “receptive field” has referred to the

TABLE 1. Vector correlations

<table>
<thead>
<tr>
<th>$P$ level</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>6</td>
</tr>
<tr>
<td>0.025</td>
<td>2</td>
</tr>
<tr>
<td>0.01</td>
<td>2</td>
</tr>
<tr>
<td>0.005</td>
<td>26</td>
</tr>
</tbody>
</table>
from both optimization sessions for each cell. The Gabor function, 
\[ y = a \left[ \frac{1}{\sigma(2\pi)^{1/2}} \right] \exp \left\{ -\left( x - \mu \right)^2 / 2\sigma^2 \right\} \sin(\omega x + \phi), \]
where the parameters \( \mu \), \( \sigma \), \( \omega \), and \( \phi \) are the center frequency and SD of the Gaussian, and the frequency and phase of the sinusoid, and \( a \) is a scale factor. In the DoG function 
\[ y = a/\sigma \left( 2\pi^{1/2} \right) \exp \left\{ -\left( x - \mu_1 \right)^2 / 2\sigma_1^2 \right\} - b/\sigma_2 \left( 2\pi^{1/2} \right) \exp \left\{ -\left( x - \mu_2 \right)^2 / 2\sigma_2^2 \right\}, \]
the parameters \( \mu_1, \mu_2, \sigma_1, \) and \( \sigma_2 \) are the center frequencies and SDs of the two Gaussians, and \( a \) and \( b \) are scale factors. The fits were performed using an iterative, nonlinear (reflective-Newton) least-squares algorithm. Care was taken to avoid local minima by performing the fit several times using different starting parameters and choosing the best fit.

For statistical evaluation of the fitted Gabor functions, the analysis was limited to that portion of the obtained amplitude vector falling within the width of the Gaussian window (equal to twice the 1/2 power width centered over the Gaussian mean). For the DoG function, the analysis was limited to the amplitude vector falling within the bounds of the upper and lower Gaussians, calculated in the same way. The \( t^2 \) statistic (representing the proportion of total variance accounted for by the fitted function) was determined as a measure of the goodness of fit. An \( F \) statistic (the ratio between the variation from the dependent variable and the residual variation about the regression) was then calculated giving the statistical significance for each fitted function (the degrees of freedom were \( K - 1 \) and \( N - K - 1 \), where \( N \) was the number of elements in the vector and \( K \) the number of parameters in the function) (Daniel and Wood 1980). Twenty-three (77%) of the Gabor fits were significant at the \( P < 0.05 \) level (21 at the \( P < 0.025 \) level, and 16 at the \( P < 0.01 \) level); 24 (80%) of the DoG fits were significant at the \( P < 0.05 \) level (20 at the \( P < 0.025 \) level and 18 at the \( P < 0.01 \) level).

RESULTS

We were able to isolate 95 single units. Five cells did not undergo any observable change in response to acoustic stimulation, and isolation was not maintained for 18 of the units; the data from these 23 cells were not subjected to further analysis. Thirty-six of the remaining 72 cells successfully met the criterion of having significantly correlated amplitude vectors from the two sessions. We could discern no obvious relationship between cortical location and optimization success. The following results and discussion concerns these 72 neurons.

Optimization and convergence

The evolution of stimulus optimization is illustrated for one of these successful cases in Fig. 2A, which depicts the relative amplitude of each frequency component in the multi-tone complex at three stages in the process. The resulting amplitude vector is the neuron’s preferred spectrum. It reflects the neuron’s affinity (positive level) or aversion (negative level) for each frequency when simultaneously present in the stimulus.

Before proceeding further there are two obvious questions concerning the optimization process that should be addressed: does the process converge toward a global rather than a local optimum and, if so, how quickly? For the first question, we used the criterion that, for each cell, two independent sessions produce spectra that were essentially alike, i.e., significantly correlated (Pearson \( r \)), meeting a minimal criterion of at least \( P < 0.05 \) (e.g., Fig. 2B). As Table 1 shows, most of the significance values were considerably smaller than this criterion with 72% (26/36) of the correlations significant at the smallest level (\( P < 0.005 \)). Figure 2C shows the probability distribution for these correlation coefficients, as well as that for the 21 neurons for which two sessions were completed but the resulting amplitude vectors were not significantly correlated (only a single session was performed for the remaining 15 units either because there was no evident change in the base spectrum or progress toward any discernable spectral pattern during this first session). As the plot shows, there was no overlap in these distributions.

To examine the rate of convergence, we tracked the similarity of spectra from successive iterations as optimization progressed. To do this, we computed the direction cosine of the angle (the normalized dot product) between the base vectors of adjacent pairs of iterations, which is equivalent to the correlation coefficient. Correlation coefficients computed from unit data for each session were plotted against iteration number and a negative exponential growth function was fit to the points. Thirty of the 36 units (83%) reaching criterion displayed at least one significant negative exponential fit (\( P < 0.05 \)); the fits from both sessions were significant for 20 (56%) units. In the case of the amplitude spectra, most units displayed rapid pattern convergence as a negatively accelerated function of iteration number (Fig. 3, A and B). Greater than half of the growth functions fit to individual unit data had time constants of two iterations or less, and asymptotic values derived from these fits cluster close to one, indicating that convergence was occurring quickly on most sessions (Fig. 3, C and D).

There was little evidence for convergence or optimization of the phase parameters. The phase vectors from the pair of optimization sessions were significantly positively correlated (\( P < 0.05 \), 1-tailed) in the case of only two neurons. This result is not unexpected by chance given the number \( n = 36 \) of correlation coefficients computed.

Preferred spectrum measurement and structure

The plots in Fig. 2, A and B, indicate the characteristic form of the preferred spectra obtained: a circumscribed, antagonistic multi-lobed organization in which positive and negative regions—suggestive of excitation and inhibition—appear to be about balanced. Figure 4 demonstrates the variation in spectral structure found within this form, which includes spectra with centered (positive) peaks (e.g., Fig. 4, A and B) and centered (negative) troughs (e.g., Fig. 4, E and F) as well as those of intermediate symmetry. All 36 neurons exhibited a preferred spectrum having this type of basic structure.

We contend in the preceding text that obtaining a significant correlation between amplitude vectors from two independent optimization sessions is strong evidence for arriving at a global optimal estimated spectrum. The fact that these spectra have in common the particular form described would appear to be independent support for this claim. As one test of this argument, we compared the spectrum estimates from the successful cases to those in which the final amplitude spectra were not significantly correlated (\( P > 0.05 \)); when the two amplitude vectors are not significantly correlated, the resulting spectra should not be as obviously structured. Ten of these spectra are shown in Fig. 5. These spectra appear to be “noisier” than those in Fig. 4, and the presence of unifying structure is less evident. A small number, however, do show some sign of the characteristic form in Fig. 4 (e.g., Fig. 5, C and F), evidently due to the convergence of the base stimulus toward a preferred form.
for the successful cases (35.8 dB) is significantly greater than values for the successful and unsuccessful cases. The median difference between the peak and trough in each spectrum confirmed this difference by comparing the spectral contrast (the result of averaging two vectors of low correlation. We confirmed this difference by comparing the spectral contrast (the difference between the peak and trough in each spectrum) values for the successful and unsuccessful cases. The median for the successful cases (35.8 dB) is significantly greater than that for the unsuccessful ones [15.9 dB; Mann-Whitney T(21,36) = 357.0; P < 0.001]. The spectra resulting from optimization sessions that failed to reach statistical significance, then, display only a coarse similarity to those that did, though there is evidence of a common form in a few cases, which may be due to a tendency to converge toward a preferred spectrum.

The fundamental form of the spectra displayed in Fig. 4 seems quite consistent, being relatively independent of size or shifts in center frequency. We examined this consistency by measuring the spectra (e.g., histograms in Fig. 4) and their excitatory and inhibitory subfields. Spectrum widths (in kHz) are plotted against spectrum center frequency in Fig. 7A. The average change in spectrum width as a function of frequency is well described by the regression line in Fig. 7A with a slope (exponent) close to one. This shows that spectrum width and center frequency increase at roughly the same rate; that is, that the ratio of spectrum width to center frequency varies about some constant value. This means that relative spectrum size (size in octaves) remains, on average, unchanged as a function of frequency. This is illustrated in Fig. 7B, which plots spectrum width in octaves (median = 0.69) over frequency. This scaling relationship also holds well for the subfields within the spectrum as is shown in Fig. 7C. The widths of the upper and lower subfields change at nearly the same rate with subfield peak (trough) frequencies changing at about the same rate as center frequency. The plots reveal striking constancy in spectrum structure, the spacing of the subfields closely maintaining their relationship over a large frequency range. Figure 7C suggests that the upper and lower frequency subfields have roughly the same relative width. This relationship is examined in Fig. 7D, which compares the width of the upper and lower...
subfields of individual preferred spectra in octaves. The points cluster near the positive diagonal, with lower bands, on average, slightly larger than upper bands.

Models of preferred spectrum structure

The structure of these preferred spectra is suggestive of the RFs found for simple cells in primary visual cortex (VI) which, like the AI preferred spectra found here, display a circumscribed, antagonistic multi-lobed organization. Two functions that have been used to provide a simple quantitative description of this sort of structure are the Gabor and DoG (Hawken and Parker 1987; Jones and Palmer 1987). The Gabor function has also been used to describe the temporal and spectral components (profiles) of spectrotemporal RFs (STRFs) for many neurons in the central nucleus of the inferior colliculus (Qiu et al. 2003), though as we note in the following text, with different results than our own. To test these functions against our data, we fit a Gabor and DoG function to each neuron’s optimized spectrum (Fig. 4). The fits of both functions track the large undulations in the spectra, and both are able to account for a significant proportion of the variance for a majority of the spectra [Gabor: 78% (28/36); DoG: 83% (30/36)]. Both fits also appear to provide good estimates of the extent of inhibitory and excitatory subregions (Fig. 4).

The preferred spectra resulting from adaptive stimulus optimization appear to be generally similar in shape despite large variations in the undulations of each spectrum.

FIG. 4. Preferred amplitude spectra from stimulus optimization. A–J: amplitude vectors (histograms) resulting from the optimization process for ten cells successfully meeting criterion (see Fig. 2 and RESULTS). Each vector represents an estimate of the neuron’s preferred spectrum (its estimated spectral receptive field), the result of averaging the 2 final amplitude vectors from the pair of optimization sessions performed for each neuron. Gabor (solid blue) and difference of Gaussian (DoG) (dashed red) functions were fit to the amplitude vector from each cell. Asterisks (Gabor) and crosses (DoG) indicate significance levels of the fits ($P < 0.05; P < 0.025; P < 0.01$).
changes in their size (scale) or in center frequency (translation),
that is, they seem to be scale invariant. To further quantify the
structure of the spectra and substantiate this scale invariance, we
examined the parameters obtained from significant fits of the
Gabor and DoG functions. For the DoG, scale invariance implies
that the widths of the two Gaussian functions, corresponding to
excitation and inhibition, change at equal rates as a function of
spectrum width. This is supported by Fig. 8A, which shows that
the mean excitatory and inhibitory Gaussian widths (in kHz)
change at nearly the same rate when plotted against spectrum
width. In the case of the Gabor function, for scale invariance to be
preserved, there should be an inverse relationship between the
Gaussian width and the frequency of the sine-spectral profile
function. That is, as width increases there should be a correspond-
ing decrease in frequency, such that the number of lobes or cycles
within the spectrum remains relatively unchanged. Conversely,
the period of the Gabor sine function and the Gaussian width
should be in direct proportion. These relationships are supported
by the plots in Fig. 8, B and C. Gabor width and sine-spectral
profile frequency tend to be inversely related (Fig. 8B), while
width and period tend to be directly proportional (Fig. 8C). The
remaining parameter defining Gabor function shape, the phase of
the sine profile, determines the relative symmetry of the spec-
trum. A phase of 0° corresponds to a center peak with flanking
troughs (e.g., Fig. 4, A and B), a phase of 180° to a center
trough with flanking peaks (e.g., Fig. 4, E and F). The
histogram in Fig. 8D shows that the phase parameter is
about evenly distributed between 0 –360°, indicating that
spectral troughs are about as prevalent as peaks. These
results differ strikingly from Qiu et.al.’s results showing
that the phase of the spectral Gabors fitted to inferior
colliculus (IC) STRFs was bounded between approximately
0–90°. We do not yet know, of course, to what degree this
difference is due to processing effects between IC and
cortex or to the differences in method (adaptive search vs.
linear estimation).

FIG. 5. Amplitude spectra from unsuccessful optimization
sessions. Amplitude vectors resulting from the optimization
process for 10 cells that did not meet criterion. Each spectrum
is the result of averaging the 2 final amplitude vectors from
the pair of optimization sessions performed for each neuron.
The parameter analysis, then, supports the finding that the spectra resulting from adaptive optimal stimulus search have a scale-invariant, multi-lobed antagonistic structure, in which inhibition dominates as often as excitation.

Stimulus optimization and neural responses

Because cortical neurons are stochastically nonstationary and habituate to repeated stimulation, we would not predict their responses to simply monotonically increase during the adaptive optimization process. Rather their actual behavior is an important empirical question that must be examined to understand the optimization technique. Figure 9A depicts the response dynamics for three neurons with significantly correlated final vectors. These cells’ responses (spikes/trial), averaged over all stimuli (E) as a function of iteration for the first (left) and second (right) optimization sessions, illustrate the degree of response variability encountered. In most cases the base stimulus responses (○) tracked those of the mean response (Fig. 9A). This supports the notion that the stimulus set—designed to randomly explore the multidimensional parameter space immediately surrounding the base stimulus—did so, given that this set and the base produced positively correlated responses in almost all cases.

Response variance (computed over each session for each neuron) increases more quickly than the mean (Fig. 9B). One common gauge of response variance is the Fano factor—the ratio of response variance to mean. For a homogeneous Poisson process (the rate of which does not change over time), this value is 1. Fano factor values computed over all neurons and sessions were greater than one in almost all cases (median = 1.90). These findings exemplify the unpredictable nature of auditory cortical neurons’ responses and underscore that this uncertainty tends to increase with a rise in response rate.

To assess the general response trends during the optimization process, we performed linear regressions on the mean

**Fig. 6.** Simulated amplitude spectra. These spectra are obtained from 2 simulations using random weightings (see RESULTS). Each spectrum is the result of averaging 2 vectors from pairs of runs given they were significantly correlated. Left: spectra from a 24-tone simulation; right: spectra from a 36-tone stimulation.
FIG. 7. Preferred spectrum measurements. A: relationship between spectrum center frequency and width. The regression line is a power function with a slope (exponent) of 0.996 ($r = 0.874; P < 0.01; n = 36$). B: scatterplot showing that mean spectrum width is invariant with respect to center frequency when width is expressed in octaves. C: scatterplots showing the relationship between the peak or trough frequencies of the spectrum subfields and center frequency. The best-fitting power functions for the lower bands (---, $r = 0.975; P < 0.01; n = 31$) and upper bands (-----, $r = 0.986; P < 0.01; n = 22$) both have exponents of 0.96, indicating that the mean relative sizes of the spectrum subfields remain nearly constant as a function of center frequency. D: relationship between the widths of the upper and lower bands when expressed in octaves. Data in D are from spectra with a center peak (or trough) and both an upper and lower subfield ($n = 20$).

FIG. 8. Preferred spectrum parameters from Gabor and DoG fits. A: relationship between spectrum width and excitatory ($r = 0.851; P < 0.01; n = 30$) and inhibitory ($r = 0.822; P < 0.01; n = 30$) Gaussian (1/2 power) widths obtained from fits of the DoG function (see Fig. 4 and METHODS); power functions both have exponents of 1.03. B: scatterplot between Gaussian spectral width and sine-profile frequency obtained by fits of the Gabor function. The best-fitting line shows an inversely proportional relation ($r = 0.770; P < 0.0001; n = 28$). C: scatterplot between Gaussian spectral width and sine-profile period. The regression line ($r = 0.915; P < 0.01; n = 28$) has a slope of 1.32 and y intercept of 0.17 octaves/cycle. D: distribution of the phase parameter obtained from the Gabor fits ($n = 28$). Inset: symmetry of preferred spectrum shapes for the corresponding phases. The parameters in (A–D) were taken only from significant ($P < 0.05$) fits.
base stimuli only. In contrast, the correlations for the remaining cases tended to be negative [median $r = -0.319$; Wilcoxon $T(57) = 476.5$; $P < 0.01$], indicating that overall response strength tended to decrease during sessions for neurons that did not successfully meet criterion. There was a slight but nonsignificant negative bias for the corresponding base stimulus correlations (median $r = -0.098$).

The fact that response rates were more likely to decrease on sessions where optimization was unsuccessful might mean that an increase in response rate contributed to success, even if it was not actually instrumental. This view is supported by examination of the response rates for neurons meeting criterion and those that did not. These are compared in Table 2, which displays the medians over sessions for the responses averaged over all stimuli and for the base stimuli alone. The response rates for successful optimization sessions tended to be larger than the unsuccessful ones for all stimuli as well as for base stimuli, although neither of these differences quite reached significance. The median maximum response rates for successful optimization sessions were significantly larger for all stimuli as well as for the base stimuli (Table 2). It is important to point out that responses measured on the first and second sessions tended to be highly significantly correlated and did not differ significantly in magnitude. The correlation was highest for the maximum responses made to base stimuli and (in the only nonsignificant case), lowest for averages over all stimuli for the unsuccessful sessions (Table 3). Despite the rather high degree of variance in neural response, then, there was a high degree of retest reliability over session in the optimization process. These results suggest that, although high response rates were not critical to successful stimulus optimization, they did play a role, perhaps in attenuating the effects of habituation.

**Discussion**

In this work, we used an adaptive stimulus optimization technique to search for the preferred spectral input for macaque primary auditory cortical (AC) neurons. This approach revealed preferred spectra having a similar structure with scaling of excitatory and inhibitory bandwidth as well as inhibition that appears about as prominent as excitation.

<table>
<thead>
<tr>
<th>TABLE 2. Median response values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Spikes/trial</td>
</tr>
<tr>
<td>All stimuli</td>
</tr>
<tr>
<td>Base stimuli</td>
</tr>
<tr>
<td>Maximum spikes/trial</td>
</tr>
<tr>
<td>All stimuli</td>
</tr>
<tr>
<td>Base stimuli</td>
</tr>
</tbody>
</table>

Median response values are shown for all successful ($n = 72$) and remaining ($n = 57$) optimization sessions. Each score in the top- and left-most 2 rows and columns represents the number of spikes/trial either averaged over all stimuli or over base stimuli only. The scores in the bottom- and left-most 2 rows and columns are the median maximum response rates averaged over either all stimuli or base stimuli. The Mann-Whitney rank-sum scores ($T$) for comparison of the successful and unsuccessful cases are shown in the 3rd column, and the corresponding probability values for the scores are given in the last; significant values are indicated by asterisks ($P < 0.05$).
The preferred spectra resulting from adaptive optimization have a circumscribed, antagonistic multi-lobed organization that appears scale invariant. How does this relatively simple prototypical form come about? It seems reasonable that both linear and nonlinear interactions contribute. The extent to which AC neurons summate linearly over frequency is a matter of current controversy. There is, however, strong recent evidence from studies using multi-tone complexes (Calhoun and Schreiner 1998; Nelken et al. 1994a) and linear RF estimation techniques (Barbour and Wang 2003; Machens et al. 2004; Sahani and Linden 2003) that many AC neurons behave in a substantially nonlinear manner in response to complex spectral input. It certainly seems possible, then, that nonlinear spectral interactions played a significant role in determining the preferred spectra found here. It is also important to note that this form and organization may arise before the level of auditory cortex; recent work using reverse correlation on neurons in the IC reveals that the Gabor functions approximating their spectral RFs show a trade-off between the widths of the Gabors and their sine-profile frequency (Qiu et al. 2003). An obvious question is: what purpose might this structure serve for audition? One possibility is that these RFs are designed to extract local features in sound spectra (deCharms et al. 1998), such as the peaks and notches in power that are important for identifying and localizing natural sounds (Middlebrooks and Green 1991; Reiss and Young 2005), that is, that they might operate as “edge” and “line” detectors for a spectrographic-like sound representation. It has also been suggested (Shamma et al. 1994) that AI RFs act as local linear filters, essentially performing a Fourier analysis of a power spectrum, analogous to the Fourier analysis on images that has been proposed for VI neurons (De Valois and De Valois 1988).

Another possibility is that this type of structure is designed for the analysis of natural acoustic scenes. Our preferred spectra show some properties consistent with Lewicki’s predictions of efficient filters for the analysis of natural sounds (Lewicki 2002). These filters were shown to have proportional increases in bandwidth, along with decreases in temporal extent, with increasing frequency—and so represent an efficient trade-off for locating sounds in both the time and frequency domains. The preferred spectra obtained in our study obey the same relationship between bandwidth and center frequency and so may be efficiently designed for the analysis of environmental sounds.

Furthermore, natural sounds are not randomly organized but display local correlations in structure in both time and frequency (Attias and Schreiner 1997; Hoth 1941; Singh and Theunissen 2003; Voss and Clark 1977). The task of the auditory system, then, may be to decorrelate overlapping signal and background sounds. Psychophysical evidence from studies of comodulation masking release (Hall et al. 1984), and neurophysiological evidence from the responses of cat AI cells to stimuli composed of either modulated or unmodulated background noise (Nelken et al. 1999) support that the auditory system may be designed for this function.

A localized, antagonistic spectral organization, such as found in our study, may efficiently perform this function because it is responsive to features that stand out against a uniform background (Daugman 1989). Interestingly, algorithms designed to optimize the efficiency of model visual RFs using natural images derive localized, antagonistic multi-lobed forms (Hyvarinen and Hoyer 2000; Olshausen and Field 1996). The resulting model RFs are similar in shape despite large changes in scale or translation. In the auditory system, this type of structure would permit AI cells to extract feature information from noisy backgrounds at multiple bandwidth levels or scales, with equal fidelity.

**Spectral RF size and structure: comparisons with previous methods**

Our results showing that preferred spectrum width scales with frequency might not seem surprising, given previous evidence that the frequency response areas (FRA) of neurons tend to broaden (in kHz) as a function of frequency in the lemniscal tonotopic pathway. However, the precise quantitative nature of this relationship has been more difficult to establish. The most common way of assessing this relation has been to measure the width of a neuron’s excitatory FRA at an intensity level just above threshold (typically 10 dB above). The best frequency (BF) divided by the width (in Hz) yields the familiar $Q$ (quality) factor. If RF width scales with frequency, then $Q$ values should remain constant relative to frequency. This is generally not true, however, for values of $Q$ measured at all levels of the auditory pathway from the auditory nerve to cortex. There is a tendency for $Q$ to increase as a function of frequency indicating that excitatory FRAs are relatively narrower (in octaves) at high frequencies, at least near threshold (Aitkin and Webster 1972; Aitkin et al. 1972, 1975; Batzri-Izraeli and Wollberg 1992; Cheung et al. 2001; Egorova et al. 2001; Ehret and Møffat 1985; Evans 1972; Kiang et al. 1965; Nuding et al. 1999; Pelleg-Toiba and Wollberg 1989; Phillips and Irvine 1981; Recanzone et al. 1999).

These results are consistent with studies reporting that AI excitatory bandwidth measured in octaves declines by about a factor of two across a 10-fold increase in BF (Evans and Whitfield 1964; Kowalski et al. 1995; measured 15 and 20 dB above threshold, respectively). Another study, however, has
found no significant relationship between excitatory octave-bandwidth (10 and 40 dB above threshold) and BF, although this might be accounted for by the fact that only one neuron with a BF <4 kHz was included in the analysis, as the authors remark (Schreiner and Sutter 1992). Still other studies have found a more modest though still significant ($p < 0.05$) decline of $\sim 0.25$ octave in bandwidth (40 dB above threshold) per decade increase in frequency (Loftus and Sutter 2001; Sutter and Loftus 2003; unpublished findings). Overall the measurements of excitatory FRAs suggest a relative decline in octave width as a function of BF, though this result may, to at least some extent, be dependent on the type of measure used as well as other experimental variables.

Another measure of excitatory FRA width is the square-root transformation. This is defined as the difference between the square roots of the high- and low-frequency bounds of the FRA just (10–20 dB) above threshold (Whitfield and Purser 1972). For FRAs that scale with frequency, square-root transform values should be linearly related to frequency when plotted on log–log coordinates with a best-fit regression slope (exponent) of 0.5. However, measurements made from the IC, medial geniculate and AI have not revealed any particular relationship between square-root transform values and frequency (Batzri–Izraeli and Wollberg 1992; Calford and Webster 1981; Calford et al. 1983; Pelleg-Toiba and Wollberg 1989; Whitfield and Purser 1972). We analyzed our own data using a variant of this technique, defining the upper and lower bounds of the preferred spectra as the peak and trough frequency points. A log–log plot of the square-root transform of these values against frequency revealed the best-fitting regression line to have a slope of 0.47 ($p < 0.01$), quite close to the predicted value of 0.5 for scale-invariant spectra.

One difference between the current and previous studies using $Q$ and the square-root transformation is that the previous studies recorded excitatory responses to single tones, whereas the present study incorporates inhibitory areas and possible nonlinear interactions including facilitation. Several studies have examined the role of excitation and inhibition on cortical FRA width using tone-plus-tone or tone-plus-noise stimulus combinations. Some of the two-tone studies have focused on the size and shape of excitatory and inhibitory response areas and their relative symmetry, in AI. Two-tone experiments conducted in cat AI have shown most neurons to have single excitatory FRAs and a smaller number (~20%) to exhibit multi-peaked excitatory bands, primarily in dorsal AI (Sutter and Schreiner 1991). The most common FRA type consists of a single excitatory band with two inhibitory flanking sidebands, ~50% in ventral AI (Sutter et al. 1999). In general agreement with these findings, recordings made in ferret AI using two-tone stimuli revealed most neurons to have excitatory centers with some degree of flanking inhibition (Kowalski et al. 1995; Shamma et al. 1993). It is interesting that the mean bandwidth of multi-peaked FRAs appears to be more nearly constant as a function of BF than excitatory FRA width, although there is considerable variability in the data (Loftus and Sutter 2001; Sutter and Loftus 2003; unpublished findings). This may mean that the addition of flanking inhibition (and possibly excitation) increases overall FRA width to a greater extent at high frequencies than low, perhaps serving to normalize FRA size with respect to BF.

Several other techniques have recently become popular for estimating the structure of auditory RFs. One of these employs ripple or spectral sine-profile-modulated stimuli (Green 1986), analogous to the sine-wave gratings used in vision research. Studies using static ripple stimuli have shown that the best ripple frequencies for cortical neurons range between ~1–4 cycles/octave (Schreiner and Calhoun 1994) and 0.3–3 cycles/octave (Shamma et al. 1994) with best-mean frequencies of 1.1 and 1.0 cycles/octave, respectively. These results correspond well to the Gabor, sine-profile frequencies fit to the spectra in our study: these range between 0.2 and 3 cycles/octave with a mean of 1.17. In the Shamma et al. study, the distribution of best sine-profile phase sensitivities was sharply peaked at about $0^\circ$ with very few neurons responding maximally near $180^\circ$. This implies that most neurons had symmetric or nearly symmetric RFs with an excitatory center, a finding in agreement with two-tone studies (see preceding text). These results stand in contrast to our own, in which inhibitory-center RFs were found to be about as common as excitatory-center RFs (Fig. 8D).

The reverse correlation technique (or closely related spike-triggered average) is another approach for assessing a neuron’s response to complex acoustic stimulation. It employs a spatially complex, temporally varying stimulus (such as white noise) to determine the average stimulus preceding a neuron’s spike. This technique generates a neuron’s time varying, frequency-sensitive RF, or spectro-temporal RF (STRF), which provides the best linear model of a neuron’s response to a dynamic stimulus. In agreement with the RF estimates made using two-tone, tone-plus-noise, and static ripple stimuli, reverse correlation STRFs obtained from AI neurons tend to show a strong excitatory region (usually in response to stimulus onset) flanked by one or more weaker inhibitory and excitatory bands (Blake and Merzenich 2002; deCharms et al. 1998; Depireux et al. 2001; Miller et al. 2002; Rutkowski et al. 2002). In the reverse correlation STRFs, inhibitory regions of longer duration, but lesser magnitude, often follow excitatory ones.

Estimates of spectral RF width from reverse correlation vary quite a bit from study to study and seem to depend on the precise technique used. deCharms et al. (1998), using random-chord stimuli, found a median STRF width (for the total response area) of 1.8 octaves. This is larger than our median bandwidth of 1.29 octaves measured at twice the one-half power width of the spectral Gaussian Gabor envelope. Blake and Merzenich (2002), also using random-chord stimuli, found STRF width to depend on spectrotemporal tone density, with both mean excitatory and inhibitory bandwidth increasing (by 0.8–2 and 2–2.5 octaves, respectively) as tone density decreased. Miller et al. (2002), using dynamic ripple stimuli, found the distribution of best ripple frequencies to be very skewed toward low values (mean = 0.46; median = 0.25 cycles/octave), implying a preponderance of RFs with very broad bandwidths (50% >4 octaves). The results from Miller et al. are in contrast to those from static ripple experiments, as well as from our own study, showing that the AI preferred spectra average about one-half of this width. Most bandwidth estimates from reverse correlation, then, give larger values than our study and the more conventional techniques, though it is clear that the spectrotemporal structure of the stimulus is important in bandwidth determination.
A recent technique of estimating a static rather than temporally dynamic spectral RF is termed the “random spectral stimulus” (RSS) method, which employs stimuli similar to those in our study (Barbour and Wang 2003; Yu and Young 2000). Barbour and Wang used a linear RSS technique on AC neurons and revealed linear weighting functions with excitatory centers and surrounding inhibitory sidebands, the sort of RF type that predominates in other studies. They also obtained AC RFs at different mean intensities for individual neurons and found RF width to be relatively level tolerant or invariant. Our results extend this finding of structural invariance to large changes in spectral width and translation.

The weighting functions of Barbour and Wang bear some similarity to the preferred spectra in our study (though they are expressed in different units). The examples they show, however, seem to indicate a preponderance of excitatory-center/inhibitory-surround RF types, but because the authors did not present a quantitative summary of these data (the proportion of off-center neurons), comparison with our own results is difficult. The predominance of this structure would be congruent with the STRF results described above. It seems likely that adaptive search is better able to reveal inhibition because it does not rely on the elicitation of high response rates to do so, which is the case with the stimulus averaging techniques.

The similarity and scale invariance of the preferred spectra in our study are striking given the range of intensities used across experiments. It is possible that in both types of study, across-frequency interactions normalized cortical responses, producing at least two different sorts of invariance: Invariance with respect to intensity and invariance with respect to scale. Whether the preferred spectra obtained from individual neurons using adaptive stimulus optimization will display level invariant size and structure must await further work.

In summary, our findings show that average, relative AI preferred spectrum size remains constant as function of frequency and basic preferred spectrum structure has a relatively simple quantitative description with inhibition about as prominent as excitation. When considering the differences between our results and those of previous studies, it is important to bear in mind that most of these studies used cats although several different species were studied, and that most, but not all, used anesthetized subjects.

### Masking, the critical band and AI preferred spectra

The “critical band”—the frequency range within which sound energy is effective in perceptually masking a tone—has been extensively studied psychophysically, and investigators have looked for its neural correlate, the neural critical bandwidth (CBW). The CBWs of central IC (ICC) neurons parallel psychophysically determined CBWs when both are plotted against frequency although they are generally broader (this similarity included a leveling off of the functions for frequencies below ~0.5 kHz) (Ehret and Merzenich 1985, 1988). The relationship between CBW and frequency appeared linear on log–log coordinates (above ~0.5 kHz), and best-fitting regression lines (>0.5 kHz) had slopes ranging from 0.63 to 0.97 [mean = 0.72 ± 0.038 (SE)], indicating that bandwidth increased at a slower rate than frequency. A similar relationship between CBW and BF (slope = 0.63) was obtained for neurons in ventral and central cat AI when CBW was determined using narrowband noise maskers but not when broadband noise and the critical ratio (the noise intensity required to mask a tone divided by the tone intensity) were used (Ehret and Schreiner 1997). The CBW measurements from ICC and AI neurons imply that their RFs tend to be relatively narrower at high frequencies when characterized using combinations of tones and noise, a result similar to that obtained from single-tone excitatory FRAs.

From their results, Ehret and his colleagues have argued that the ICC is a likely site for the neural mechanisms underlying critical-band related phenomena—more so than the cochlea. This claim is based on neural tuning properties such as bandwidth, level tolerance, and linearity, and is substantiated by the similarity between the neural and psychophysical CBW-frequency plots for the cat, which yield log-log slopes centering on 0.7–0.8 (Ehret and Merzenich 1985, 1988; Pickles 1979; Pickles and Comis 1976). This is evidence that neural and psychophysical CBWs are narrower at high frequencies in this species (Felis catus). In our study, the log–log plot of spectrum width and center frequency shows a best-fit slope of 0.99 (Fig. 7A), indicating that width remains relatively constant as a function of frequency. An important question is how this relationship compares with that of the behaviorally measured critical band and frequency in macaques. When plotted against frequency on log-log axes, critical-band measures in pig-tailed macaques (Macaca nemestrina) reveal a slope of 1.01 (Gourevitch 1970).2 (Such data are not available for rhesus macaques.) The pig-tailed macaque CBWs are comparable to those of the human within the frequency range tested (Fay 1988; Zwicker et al. 1957), suggesting that psychophysical CBWs may be similar across primates. The macaque and human psychophysical CBWs are plotted as a function of frequency in Fig. 10, along with the preferred spectrum widths found in our study. The neural bandwidths tend to be broader with the behavioral CBWs arrayed along the lower bound of

---

2 The slope was calculated using values given in Fay (1988).
the scatterplot, results that are similar to those found with ICC CBWs in the cat. We do not mean to imply with these data that AI neurons are solely or principally responsible for critical-band phenomena in macaques or humans. The results do, however, parallel the relationship between estimated RF size and behavioral and neural masking in the cat, and so are also consistent with the structure of the preferred spectra found in our study relating to frequency selectivity or analysis.

Complex RF organization and model fits

As noted in the preceding text, studies using single- and two-tone stimuli have demonstrated that some frequency-response areas in AI have multiple excitatory bands separated by an inhibitory band (Abeles and Goldstein 1972; Sutter and Schreiner 1991). Our adaptive optimization results also revealed this type of spectral sensitivity in neurons with ~180° sine-phase, i.e., those having a center trough with surrounding peaks (Fig. 6D). Studies using two-tone stimuli to reveal inhibitory bands also show some very complex multi-lobed AI FRAs (Sutter and Loftus 2003; Sutter et al. 1999), a form that is not obvious in our results. It may be that the central two or three lobes of the RF are those that overwhelmingly affect a neuron’s response during the adaptive search process. It is also possible that the increased number and/or range of possible frequency interactions available in our broadband stimuli yield a simpler spectral form; i.e., that across-frequency interactions are critical in creating this form.

Although both the Gabor and DoG models described the basic structure of the preferred spectra, they did not always capture their finer details (it should be noted that the DoG fit can, at most, capture 1 central and 2 surround bands). For instance, the spectrum peaks and troughs sometimes under- or overshoot the fitted functions (e.g., Fig. 4, G, I, and J). Also, the functions sometimes failed to account for variations at the fringes of the preferred spectra (e.g., Fig. 4, B, D, and F). The alternating pattern of these outlying variations suggests that they are genuine low-amplitude skirts of the preferred spectra. These patterns are consistent with two-tone studies demonstrating more complex multi-banded antagonistic regions in AI FRAs (Sutter et al. 1999). Despite these limitations, both functions appear to provide reasonable simplifying formal descriptions of the largest components of AI preferred spectra.

Response variability and stimulus optimization

Gradient optimization methods might seem unsuitable for single units in cortex because of response variability. Our positive results, and the success of a similar procedure for VI neurons (Foldiak 2001), counter this argument. We were able to obtain two independent, convergent optimization runs for one-half of the 72 single units we tested.

Other adaptive search techniques have been used in the auditory system with some success. Nelken and colleagues (1994b) applied a stimulus optimization technique using the simplex algorithm, maximizing the (filtered and summed) activity of clusters of units in auditory cortex. The parameters to be optimized in their study were the frequencies, but not amplitudes, of pure and complex tones, where the complex tones were the sum of two, four, or nine pure tones. In the successful optimization cases, they obtained stimuli the tone elements of which (from different optimization runs) were close in frequency and found that optimization improved with the more complex stimuli. They concluded that the location of a spectral peak in a sound, or the locations of pairs of peaks, is an important parameter for cortical sound analysis. This finding agrees with our single-unit preferred spectrum estimates where we find strong single-peaked spectra (Fig. 4, A–C and G–J) and weaker two-peaked spectra associated with a central inhibitory region (Fig. 4, E and F).

Recently, another group (Bleeck et al. 2003) used a “genetic algorithm” to search a limited parameter space for the most effective stimuli for single units in the cochlear nucleus and IC. In their study, parameters controlling amplitude-modulated (AM) pure tones (e.g., carrier frequency and AM frequency) were manipulated by the technique, which was successful in many cases in estimating the most effective modulation. The success of these studies, and the results reported on here, support the general feasibility of adaptive search methods.

The nonstationarity of cortical single-unit responses, as well as response habituation, might make estimation of response strength over the course of an optimization experiment difficult. However these possible sources of response variability do not appear to hinder estimation of the preferred spectrum. There are several possible reasons for this. One is that response variance increases at a rate larger than the mean response. As we describe in the preceding text (see Stimulus optimization and neural responses), the ratio of response variance to the mean (the Fano factor) tends to be greater than one, indicating that the variability is higher than one would predict from a Poisson process, consistent with measures in other regions of cortex (e.g., Baddeley et al. 1997; Tolhurst et al. 1983). As such, the preference for particular stimuli, in terms of a signal-to-noise ratio, will not increase with response strength. This analysis shows that the ability of a neuron to signal stimulus preferences does not crucially depend on response magnitude.

Another reason that variation in activity level should not seriously impede optimization is due to response normalization within an iteration. The normalization term $\Delta r_{\text{max}}^{(i)}$ in Eq. 1 normalizes response level within a stimulus set, which should attenuate the effects of changes in response strength over the course of an experiment. If response variation did seriously impair optimization, the likely result would be random drift in the form of the base vector and a failure of the two optimization sessions to converge on a preferred stimulus form. Finally, because within each iteration the stimuli were presented in random order on each presentation, the more gradual changes in time occurring at much slower rates can be, to at least some extent, overridden. It seems quite improbable, then, that optimization depended crucially on changes in response strength over a session. Nor does it appear that response variability or habituation pose a severe obstacle to an adaptive search procedure.

Optimization and methodological concerns

Estimates of auditory RF size and structure vary considerably over studies. Much of this variation is no doubt species specific and due to differences in methodology. Blake and Merzenich’s results, where changes in tone density affected bandwidth estimates by a factor of two, illustrate just how
sensitive RF properties are to changes in just a single stimulus variable. We chose our tone densities (12 or 16 tones/octave) with the aim of sampling frequency with a resolution high enough for adequate structural detail but low enough for the number of parameters to be tractable for our optimization program. It is possible that by changing tone density or by altering spectral content by some other means (such as controlling levels of filtered bands of noise), the results might differ.

Temporal factors such as AM and FM, or sequential effects such tonal suppression and facilitation, are other variables that might be expected to influence results from stimulus optimization experiments. We sought to optimize our stimuli by modifying both the phase and amplitude parameters of the stimuli with the anticipation that we might find systematic changes in the structure of the waveform that would converge on an optimal form. These changes might, for example, take the form of particular AM or FM variations for which many AC neurons are known to be sensitive (AM: Barbour and Wang 2002; Bieser and Muller-Preuss 1996; Eggermont 1994, 1998; Ehret and Schreiner 2000; Gaese and Ostwald 1995; Liang et al. 2002; Lu et al. 2001; Schreiner and Urbas 1988; FM: Barbour and Wang 2002; Eggermont 1994; Gaese and Ostwald 1995; Heil and Irvine 1998; Heil et al. 1992; Liang et al. 2002; Mendelson et al. 1993; Nelken and Versnel 2000; Tian and Rauschecker 1994, 1998, 2004).

We found no clear evidence of phase convergence, however. There are several possible reasons for this failure. First, no global optima may have existed. Or it may have been the case that the phase parameter space was too large or too complex. It is also possible that the relatively short stimulus duration we used did not provide a time base long enough for significant temporal effects to manifest themselves. Results may be more positive if future stimulus optimization experiments were to explicitly modify sprectrotemporal structure (e.g., AM and FM) than to attempt to do so via the phase parameters—and to examine these variables as a function of stimulus duration.

Complicating this picture, reverse correlation experiments have shown that modulations in time and frequency may be manifested in quantitatively different ways. For example, some STRFs are temporally and spectrally independent or separable, that is, they can be generated by the product of separate temporal and spectral profiles, and appear symmetric when plotted (Depireux et al. 2001; Miller et al. 2002; Qiu et al. 2003; Sen et al. 2001). Nonseparable STRFs are dynamically more complex; they are not independent and appear asymmetric. Because we obtained static spectrum estimates in our study, we can draw no conclusions about RF changes over time. It is clear that to obtain the nonlinear counterpart to the STRF, it will be necessary to optimize stimuli dynamically. On the other hand, with the adaptive stimulus optimization technique, unlike reverse correlation, an estimate of the preferred spectrum can be obtained even when spikes fail to lock to stimulus events, i.e., when there is a great deal of temporal variability in the evoked spike train.

It is clear, then, that methodology—including stimulus design and presentation—has been an important factor in estimating cortical RF size and shape using previous techniques, and this may hold, as well, for adaptive stimulus optimization methods. For this reason, it is important to view our results in light of their experimental context in which we attempted to find the preferred spectra of AI neurons when presented with tone complexes having a large but limited set of tone components—and with the particular filtering properties of the whole system including the effects of the subject’s body, head, and pinnae. We’ve already discussed the possibility that the optimal spectrum may not be static. It is also possible that more than one optimal spectrum may exist for some AI cells, similar to the case of VI complex cells, which respond robustly to a grating at the preferred orientation and spatial frequency, regardless of the spatial phase—as Foldiak demonstrated in his experiments on VI. However, we did not find evidence of analogous neurons in our own data, i.e., we did not find large shifts in the position of the spectra on the two optimization runs. It is also important to point out that, because a putative “optimal” stimulus would have to be tested against a countless number of variations to prove its optimality, there is no way of being sure that a better stimulus has not been tested, i.e., it is impossible, in principle, to prove a stimulus is optimal.

Within this context, we believe we have found the short-term preferred spectra of this sample of neurons. However, far from viewing these results as the last word on the subject, we consider them to be a successful first effort using a methodology that will no doubt evolve and refine our notions of RF structure and single-unit models of auditory cortical function.

ACKNOWLEDGMENTS

We thank B. A. Olshausen, P. Sallee, M. L. Spezio, and E. D. Young for critical comments on the manuscript. We also thank P. Barruel and P. Yin for assistance in performing the experiments and T. Fickett for assistance in data analysis.

GRANTS

This work was supported by National Institute on Deafness and Other Communication Disorders Grant DC-02514.

REFERENCES


