

Spectral Receptive Field Properties Explain Shape Selectivity in Area V4

Stephen V. David,¹ Benjamin Y. Hayden,² and Jack L. Gallant^{3,4}

¹Graduate Group in Bioengineering, ²Department of Molecular and Cellular Biology, ³Department of Psychology, and ⁴Program in Neuroscience, University of California, Berkeley

Submitted 31 May 2006; accepted in final form 12 September 2006

David, Stephen V., Benjamin Y. Hayden, and Jack L. Gallant. Spectral receptive field properties explain shape selectivity in area V4. *J Neurophysiol* 96: 3492–3505, 2006. First published September 20, 2006; doi:10.1152/jn.00575.2006. Neurons in cortical area V4 respond selectively to complex visual patterns such as curved contours and non-Cartesian gratings. Most previous experiments in V4 have measured responses to small, idiosyncratic stimulus sets and no single functional model yet accounts for all of the disparate results. We propose that one model, the spectral receptive field (SRF), can explain many observations of selectivity in V4. The SRF describes tuning in terms of the orientation and spatial frequency spectrum and can, in principle, predict the response to any visual stimulus. We estimated SRFs for neurons in V4 of awake primates by linearized reverse correlation of responses to a large set of natural images. We find that V4 neurons have large orientation and spatial frequency bandwidth and often bimodal orientation tuning. For comparison, we estimated SRFs for neurons in primary visual cortex (V1). Consistent with previous observations, we find that V1 neurons have narrower bandwidth than that of V4. To determine whether estimated SRFs can account for previous observations of selectivity, we used them to predict responses to Cartesian gratings, non-Cartesian gratings, natural images, and curved contours. Based on these predictions, we find that the majority of neurons in V1 are selective for Cartesian gratings, whereas the majority of V4 neurons are selective for non-Cartesian gratings or natural images. The SRF describes visual tuning properties with a second-order nonlinear model. These results support the hypothesis that a second-order model is sufficient to describe the general mechanisms mediating shape selectivity in area V4.

INTRODUCTION

Cortical area V4 lies near the middle of a hierarchical sequence of visual areas that mediate shape perception (Felleman and Van Essen 1991; Ungerleider and Mishkin 1982; Van Essen et al. 1994). Early in this pathway, in primary visual cortex (V1), neurons are selective for a small number of simple stimulus features such as position, orientation, and spatial frequency (De Valois et al. 1982a; Hubel and Wiesel 1968). At more central stages of processing, in the inferior temporal cortex (IT), neurons are selective for more complex patterns. Tuning in central areas is often related to object identity and invariant to stimulus position and size (Desimone et al. 1984; Kobatake and Tanaka 1994). V4 plays a crucial role in transforming simple physical stimulus features to the abstract form representation in IT; damage to V4 interferes with shape perception, color perception, and attention (De Weerd et al. 1996; Gallant et al. 2000; Merigan 1996; Merigan and Pham 1998; Schiller 1995; Schiller and Lee 1991).

Neurophysiological studies have not produced consistent descriptions of shape coding in V4. One early experiment

reported that V4 neurons are tuned for size and invariant to stimulus position, properties not found in more peripheral areas (Desimone and Schein 1987). A later series of experiments compared selectivity for Cartesian gratings and for polar and hyperbolic (non-Cartesian) gratings in V4 (Gallant et al. 1993, 1996); V4 neurons are most selective for non-Cartesian gratings containing multiple orientations. A separate study reported that the optimal stimulus for single V4 neurons varied widely, but that most cells respond strongly to stimuli containing multiple orientations (Kobatake and Tanaka 1994). A more recent study used a parameterized set of contour features varying in angularity, curvature, and orientation (Pasupathy and Connor 1999, 2002). Among these stimuli, a large fraction of V4 neurons are tuned for angled or curved contour features.

These previous studies agree that single V4 neurons are tuned for multiple orientations and show position invariance, but they differ in their specific conclusions about shape selectivity in V4: Are V4 neurons tuned for non-Cartesian gratings, simple objects, or curved and angled contour elements? The most likely explanation is that, to some extent, V4 neurons are tuned for *all* of these patterns. Different studies have used different, limited stimulus sets to test specific hypotheses about shape coding in V4, and most have not systematically compared responses between classes of stimuli (but see Gallant et al. 1996). An experiment that uses a limited stimulus set can maximize statistical power for testing a specific hypothesis about tuning, but the conclusions that can be drawn about underlying mechanisms are ambiguous. The observed tuning might actually reflect tuning along untested dimensions correlated with those tested, and the observed tuning reveals nothing about tuning along dimensions orthogonal to those tested. This uncertainty can be resolved only with a general model whose scope is not restricted to a limited stimulus set (Wu et al. 2006).

We hypothesized that a single functional model, the spectral receptive field (SRF), can explain previous observations of shape tuning in V4. The SRF accounts for second-order nonlinear response properties, describing tuning in terms of the orientation and spatial frequency power spectrum, independent of spatial phase (Bredfeldt and Ringach 2002; David and Gallant 2005; Mazer et al. 2002). The power spectrum is a basic feature of all visual stimuli; thus the scope of the SRF is not limited to a particular stimulus set. Independence from spatial phase introduces a nonlinearity that enables the SRF to describe spectral tuning, even for neurons with position-invariant responses.

Address for reprint requests and other correspondence: J. L. Gallant, Program in Neuroscience, 3210 Tolman Hall #1650, University of California, Berkeley, Berkeley, CA 94720 (E-mail: gallant@berkeley.edu).

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

To determine whether the SRF provides an effective general description of shape selectivity in V4, we recorded the responses of single V4 neurons to a large set of natural images and estimated the SRF of each neuron using linearized reverse correlation (David and Gallant 2005; Theunissen et al. 2001; Wu et al. 2006). We then used the SRFs to predict how each neuron would respond to stimuli used in the studies described above (i.e., Cartesian and non-Cartesian gratings and curved-contour elements). Across the entire set of V4 SRFs, we observed a pattern of selectivity for non-Cartesian gratings and curved contours consistent with the conclusions of experiments using synthetic stimulus sets. Inspection of estimated SRFs revealed mechanisms that may underlie this selectivity for complex features. We performed a similar analysis using primary visual cortex (V1) neurons. Neurons in V1 have consistently simpler spectral tuning and are selective for Cartesian gratings rather than for the other stimulus classes. Therefore the tuning we observe in area V4 is an emergent property of the extrastriate cortical network.

METHODS

Neurophysiological procedures and data acquisition

SUBJECTS AND PHYSIOLOGICAL PROCEDURES. Data were collected from four adult male macaques (*Macaca mulatta*; two animals used in V4 recordings and two in V1 recordings). All procedures were in accordance with National Institutes of Health and U.S. Department of Agriculture guidelines and were approved by University oversight committees. Details of neurophysiological procedures were previously published (V4: Hayden and Gallant 2005; V1: Vinje and Gallant 2002). During recording, V4 and V1 neurons were identified on the basis of both stereotaxic coordinates and receptive field properties (e.g., size/eccentricity ratios and latencies; Gallant et al. 1996; Gattass et al. 1988; Mazer and Gallant 2003; Vinje and Gallant 2002).

RECEPTIVE FIELD ESTIMATION. The boundaries of each classical receptive field (CRF; specifically, the minimum response field) were measured while each animal performed a passive fixation task. Bars, Cartesian gratings, and non-Cartesian gratings were presented under manual control to determine basic receptive field properties (Mazer and Gallant 2003; Vinje and Gallant 2002). Receptive field size, shape, and location were confirmed by reverse correlation using a dynamic sequence of small white, black, and textured squares flashed randomly in and around the CRF (V1: 72 Hz; V4: 10 Hz; Hayden and Gallant 2005; Vinje and Gallant 2002). CRF diameter was defined to be the diameter of a circle circumscribing the minimum response field. In the few cases where manual and automated estimates disagreed, the estimate from the automated procedure was used. V4 CRFs were centered 3–8° from the fovea (median 5.6°) and ranged from 5 to 10° in diameter (median 10.2°). V1 CRFs were centered 0.9–12° from the fovea (median 2.2°) and ranged from 0.3 to 3.0° in diameter (median 0.65°).

STIMULI. Stimuli were circular natural image patches cut out of black and white photos (Corel). Images were chosen at random by an automated algorithm that favored images with broad spatial frequency spectra. For V4 data, the size of each image patch corresponded to the measured CRF size. For V1 data, the size of each image patch ranged from two to four times the CRF diameter. In both cases, the outer 10% of each image was blended linearly into the mean-luminance gray background.

BEHAVIORAL TASK. Neuronal activity was recorded from single V4 neurons of two animals while they performed a delayed match-to-sample task (Hayden and Gallant 2005). Each trial was initiated when

the animal grabbed a capacitive touch bar. A fixation spot then appeared at the center of the display. The animal was required to acquire and maintain fixation for the duration of the trial (fixation window radius, 0.5°). A feature cue and a spatial cue then appeared simultaneously for 150–600 ms. The feature cue was the target for that trial, a natural image the size of the CRF, centered at the fixation point. The spatial cue was a small red line (<1°) superimposed on the edge of the feature cue nearest the stream to be attended. After an 850-ms blank delay, two stimulus streams appeared simultaneously: one in the CRF and the other in the opposite hemifield at the same distance from the fovea, 180° away from the first. Images appeared at a constant rate (3.5–4.5 Hz, varying across cells), and there was no blank interval period between successive images. The target image appeared 4–10 s after the onset of the image stream. To receive reward animals had to release the touch bar within 1 s after the onset of the target in the attended stream. Incorrect trials were aborted immediately after broken fixation or early bar release. Only data from correct trials (95%) were included in the analysis. Four attention conditions were constructed by crossing two spatial conditions (attend in and attend out) with two feature conditions (search for target A, search for target B). The data presented in this report were obtained by averaging across all four attention conditions. Responses to target stimuli were excluded from the data because of their behavioral relevance. This ensured that receptive field estimates reflected only visual tuning and were not influenced by attention.

Neuronal activity was recorded from area V1 of two different animals while they performed a fixation task (fixation duration, 5 s; fixation window radius, 0.35°), with no explicit manipulation of attention. While the animal fixated, a sequence of natural image patches was presented at 60 Hz in the receptive field of an isolated neuron. Only data from periods when fixation was successfully maintained were included in the analysis.

Based on previous studies of V1, the difference in stimulus presentation rates for recordings in V4 (3.5–4.5 Hz) and V1 (60 Hz) should be irrelevant for the current study. Our analysis focused on the spectral tuning of excitatory responses. In area V1, temporal stimulus dynamics do not affect the spectral tuning of excitatory responses (although tuning of inhibitory responses in V1 can depend on stimulus dynamics; see David et al. 2004).

By averaging V4 responses across attention conditions, we intended to remove the effects of attention and to preserve only the visual response. We assume that this averaging controls for differences in behavior between the V4 and V1 experiments. However, we cannot exclude the remote possibility that some of the differences in tuning between V4 and V1 neurons that we report here might be caused by differences in behavioral state.

DATA ACQUISITION. Behavioral control, stimulus presentation, and data collection were performed on a Linux workstation using custom software. For V4 data, eye movements were recorded with an infrared eye tracker (RK-801 at 120 Hz, ISCAN, Burlington, MA; or Eyelink II at 500 Hz, SR Research, Toronto, Canada). Eye tracker latency was corrected during subsequent analysis (Gawne and Martin 2000). For V1 data, eye movements were measured using a scleral search coil (Riverbend Instruments; Judge et al. 1980).

Single-neuron responses were recorded using high-impedance epoxy-coated tungsten microelectrodes (nominal impedance 10–25 M Ω , 125- μ m diameter, 20–25° taper; FHC, New Brunswick, ME). For V4 data, neuronal signals were acquired using an integrated multichannel recording system (amplification, filtering, and spike detection; MAP, Plexon, Dallas, TX). For V1 data, signals were amplified (AM Systems, Everett, WA), band-pass filtered, and isolated with a hardware window discriminator. Only clearly isolated single units were included in the data set. Spike times were recorded with 0.1-ms resolution and synchronized with the behavioral task and eye recordings.

Spectral receptive field model and estimation procedure

THE FOURIER POWER MODEL. Simple cells in peripheral visual areas can be characterized by a linear spatial receptive field model (DeAngelis et al. 1993; Jones and Palmer 1987). According to the linear model, the response of a neuron is a weighted sum of stimulus luminance over space and time. However, the linear model cannot be used to characterize V4 neurons because these cells show nonlinear position invariance and visual selectivity does not depend on the precise position of the stimulus in the receptive field (Desimone and Schein 1987; Gallant et al. 1996). To account for position invariance in V4 we used a nonlinear Fourier power model. According to this model the response of a neuron is a weighted sum of the spatial Fourier power of the stimulus. The map of weightings is called the *spectral receptive field* (SRF; David and Gallant 2005; Theunissen et al. 2001).

A visual stimulus, $s(x, y, t)$, can be described in terms of luminance sampled at $N \times N$ spatial positions (x, y) and at times $t = 1 \dots T$. The Fourier power transform of the stimulus $\hat{s}(\omega_x, \omega_y, t)$ is

$$\hat{s}(\omega_x, \omega_y, t) = \left| \sum_{x=1, y=1}^N e^{i(\omega_x x + \omega_y y)} s(x, y, t) \right|^2 \quad (1)$$

The value of \hat{s} at each two-dimensional spatial frequency channel, (ω_x, ω_y) , indicates how much power is present at a particular orientation and spatial frequency in a single stimulus frame (see Eq. 7, below, for interpretation of spatial frequency channels).

According to the Fourier power model, the response is the inner product of the Fourier power transform of the stimulus and the SRF (Bredfeldt and Ringach 2002; David and Gallant 2005; Mazer et al. 2002)

$$r(t) = \sum_{\substack{\omega_x = -N/2 \\ \omega_y = -N/2 \\ \omega_x = N/2}}^{N/2} h(\omega_x, \omega_y) \hat{s}(\omega_x, \omega_y, t) + r_0 + \varepsilon(t) \quad (2)$$

The response $r(t)$ is the average firing rate during time bin t . The SRF, $h(\omega_x, \omega_y)$, describes the weight that should be applied to each Fourier power channel to produce the minimum mean-squared error estimate of the response. The baseline r_0 represents the response expected when no stimulus is present. The residual $\varepsilon(t)$ represents observed deviations from Fourier power model predictions (i.e., unexplained variance). These deviations reflect both unmodeled nonlinear response properties and neuronal noise.

The Fourier power transform *linearizes* the relationship between stimulus and response. That is, the stimulus is nonlinearly transformed so that a linear model more accurately describes the functional relationship between the transformed stimulus and the response (Aertsen and Johannesma 1981; David and Gallant 2005; Wu et al. 2006). The Fourier power model discards spatial phase but preserves information about stimulus orientation and spatial frequency. It is therefore related to the energy model used to describe complex cells in area V1 (Adelson and Bergen 1985). However, the Fourier power model is more general than the energy model because it can account for excitation and inhibition across any number of spatial frequency and orientation channels.

Some receptive field models include an additional nonlinear output term to account for spiking threshold and saturation (Albrecht and Geisler 1991; David et al. 2004). A sigmoidal output nonlinearity does lead to a modest improvement in the predictive power of the SRFs estimated in this study (data not shown). However, fitting the output nonlinearity has no effect on measurements of orientation and spatial-frequency tuning. Because this study focuses on spectral tuning, the SRFs reported here do not include an output nonlinearity.

FITTING THE FOURIER POWER MODEL BY LINEARIZED REVERSE CORRELATION. We estimated SRFs by linearized reverse correlation of neuronal responses and natural image stimuli. This procedure finds the

minimum mean-squared error, linear mapping between the Fourier power transform of the stimulus $\hat{s}(\omega_x, \omega_y, t)$ and the observed response $r(t)$ (David and Gallant 2005; Theunissen et al. 2001). According to this solution, the SRF is the weighted average of the stimulus and response, normalized by the inverse of the stimulus autocorrelation function C_{ss}

$$h(\omega_x, \omega_y) = \sum_{\substack{\omega'_x = -N/2 \\ \omega'_y = -N/2 \\ \omega'_x = N/2}}^{N/2} \left[C_{ss}^{-1}(\omega_x, \omega_y, \omega'_x, \omega'_y) \frac{1}{T} \sum_{t=1}^T \hat{s}(\omega'_x, \omega'_y, t) r(t) \right] \quad (3)$$

The stimulus autocorrelation function measures the correlation between each pair of spectral channels in the stimulus

$$C_{ss}(\omega_x, \omega_y, \omega'_x, \omega'_y) = \frac{1}{T} \sum_{t=1}^T \hat{s}(\omega_x, \omega_y, t) \hat{s}(\omega'_x, \omega'_y, t) \quad (4)$$

The autocorrelation function can be represented as a matrix with rows corresponding to spectral channels (ω_x, ω_y) and columns corresponding to channels (ω'_x, ω'_y) . The inverse autocorrelation function is equivalent to the inverse of this matrix (Theunissen et al. 2001).

Normalization by the stimulus autocorrelation in Eq. 3 removes bias arising from the autocorrelation inherent in natural scenes (Field 1987; Zetzsche and Barth 1990). Although necessary for achieving a minimum mean-squared error estimate of the SRF, normalization can amplify noise at high spatial frequencies, overfitting the SRF to noise in the estimation data. To minimize this effect, we used singular-value decomposition (SVD) to estimate a pseudoinverse of the stimulus autocorrelation function (Theunissen et al. 2001). The pseudoinverse forces tuning on spectral dimensions to be zero if the stimulus variance along that dimension is not large enough to reliably estimate its effect on responses. This procedure requires selecting a parameter that determines the noise threshold, which was determined simultaneously with the shrinkage parameter (see following text).

A shrinkage filter was used to further reduce noise in the SRF estimate (Brillinger 1996; David and Gallant 2005). The shrinkage filter applies a soft threshold to each SRF parameter, based on its signal-to-noise level. Signal-to-noise was defined as the ratio of mean to standard error and was measured using a jackknife procedure: jackknife SRFs, $h_i(\omega_x, \omega_y)$, $i = 1 \dots N = 20$, were estimated from subsets of the estimation data set, each excluding a different 5% of the available samples. The mean SRF was computed by averaging over the jackknife estimates, $\bar{h}(\omega_x, \omega_y) = 1/N \sum_{i=1}^N h_i(\omega_x, \omega_y)$, and the standard error of each parameter was measured according to the jackknife theorem (Efron and Tibshirani 1986)

$$\sigma(\omega_x, \omega_y) = \sqrt{\frac{N-1}{N} \sum_{i=1}^N (h_i(\omega_x, \omega_y) - \bar{h}(\omega_x, \omega_y))^2} \quad (5)$$

The shrinkage filter was applied to the mean SRF to produce the final SRF estimate (Brillinger 1996)

$$h(\omega_x, \omega_y) = \bar{h}(\omega_x, \omega_y) \sqrt{1 - \gamma(\sigma(\omega_x, \omega_y)/\bar{h}(\omega_x, \omega_y))^2}^+ \quad (6)$$

Applying the filter requires selecting a parameter γ , that determines the filter threshold. Optimal pseudo-inverse and shrinkage parameters were chosen simultaneously by cross validation (David and Gallant 2005). This entire procedure (including cross validation) was completed using only the estimation data set. The validation data set (see below) was reserved only for testing SRF prediction accuracy.

DATA PREPROCESSING. For SRF estimation, each stimulus frame was cropped to an area equivalent to one classical receptive field diameter. Each frame was then smoothed, downsampled to 20×20 pixel resolution, and multiplied by a Hanning window (ramped from 1 to 0) to reduce edge artifacts in the Fourier transform. This downsampling procedure preserves spatial frequencies ≤ 10 cycles

per receptive field diameter (cyc/RF). In theory, a more accurate model might be obtained by including higher spatial frequencies. However, natural scenes have relatively low power at high spatial frequencies (Field 1987), which makes it difficult to obtain data sets large enough to characterize tuning at high frequencies. For V4 data, the response $r(t)$, evoked by each 3.5- to 4.5-Hz stimulus frame $s(x, y, t)$, was defined as the mean spike rate (spikes/s) from 50 to 250 ms after the onset of the frame.

The stimuli used during V1 recordings had the same spatial statistics as those used for V4, but they were shown much more rapidly (60 Hz). We therefore used a slightly different procedure to estimate SRFs for V1 neurons. First, we estimated a complete spectro-temporal receptive field (STRF) for each neuron by repeating the SRF estimation procedure described above at 13 temporal delays (0–192 ms), with the same 20×20 pixel/CRF downsampling as for the V4 data. Separable spectral and temporal receptive fields were then extracted from each STRF by SVD (David et al. 2004; Mazer et al. 2002). The resulting SRF describes orientation and spatial frequency tuning in the same Fourier power parameter space as that used for V4.

EXCLUSION OF NEURONS WHOSE RECEPTIVE FIELDS COULD NOT BE CHARACTERIZED. The goal of this study was to determine whether the SRF can account for shape selectivity in V4. Therefore we used a cross-validation procedure to exclude neurons whose SRF failed to provide any information about visual response properties. For each neuron, a subset of the stimulus–response data (5%) was reserved before SRF estimation (validation data set). The SRF, including regularization parameters, was estimated using only the remaining 95% of the data (estimation data set). Predicted responses to stimuli in the validation data set were then generated from the SRF using Eq. 2. This procedure was repeated 20 times; each time a different 5% subset of the data were reserved for validation. The 20 predicted responses were concatenated into a single prediction of the entire response. Prediction accuracy was quantified in terms of the correlation (Pearson's r) between predicted and observed responses. Because we strictly separated the estimation and validation data sets, measurements of prediction accuracy were not biased by overfitting to noise in the data. A neuron was included in further analyses only if its SRF predicted the observed responses in the validation data with greater accuracy than would be expected by chance ($P < 0.05$) (David et al. 2004).

Of the 103 V4 neurons in our original sample, 87 had SRFs that significantly predicted responses in the reserved cross-validation data set. The mean prediction correlation was 0.29 for the entire sample of V4 neurons and 0.32 for the 87 significant cells. [Note that this measurement was not corrected to reflect the noise ceiling on predictions (David and Gallant 2005; Wu et al. 2006). Thus this value is smaller than the theoretical maximum for the Fourier power model in the absence of noise.] Of 56 V1 neurons in the sample, 45 had SRFs that significantly predicted responses in the cross-validation data set. The mean prediction correlation was 0.33 for the entire sample of V1 neurons and 0.37 for the 45 significant cells. (These figures were also not corrected to reflect the noise ceiling.) Excluding neurons whose SRFs did not predict with significant accuracy did not change any trends in the data reported here, but it did slightly increase the magnitude and significance of some effects.

The correlation coefficient indicates the portion of the response in the validation data set explained by the Fourier power model (David and Gallant 2005). The remaining, unexplained portion the response results from two factors: visual tuning properties not described by the Fourier power model and nonvisual influences on the response. The latter category includes noise in the neuronal response and changes in attention state. The effect of nonvisual influences is reduced by averaging across stimulus presentations, but it is unlikely to be removed completely.

Analysis of tuning and selectivity

ORIENTATION AND SPATIAL FREQUENCY TUNING CURVES. To facilitate visualization of neuronal tuning and selectivity, each SRF was transformed from the Fourier power domain to an explicit representation of orientation and spatial frequency. This was accomplished by applying a polar-to-Cartesian transformation to the SRF

$$H(\theta, \omega) = h(\omega \cos \theta, \omega \sin \theta) \quad (7)$$

Figure 1 shows several image patches that have been transformed into the orientation spatial frequency representation; transformed SRFs are shown in Figs. 2–4.

Tuning curves were obtained from SRFs transformed according to Eq. 1 by SVD (Mazer et al. 2002). Orientation and spatial frequency tuning curves [$f(\theta)$ and $g(\omega)$, respectively] were defined as the first eigenvectors of each decomposition matrix. According to the definition of the SVD, the product of these two vectors provides the minimum mean-squared-error estimate of the full, two-dimensional SRF

$$[f(\theta), g(\omega)] = \arg \min_{\theta, \omega} \sum (H(\theta, \omega) - f(\theta)g(\omega))^2 \quad (8)$$

In Eq. 8, the sign of the orientation and spatial frequency tuning curves is ambiguous. We fixed the sign so that the orientation tuning curve produced a positive inner product with the mean of the SRF after averaging over all spatial frequencies.

COMPARISON OF TUNING PROPERTIES. Several properties of the orientation and spatial frequency tuning curves for each neuron were used to compare spectral tuning across cells. Two common metrics used to describe orientation tuning curves are the peak and bandwidth (i.e., width at half height; Desimone and Schein 1987; De Valois et al. 1982b). We estimated the peak and bandwidth by fitting a circular Gaussian to the orientation tuning curve obtained for each neuron (Fisher 1993); the tuning peak and bandwidth were taken as the mean and width at half-height of the Gaussian, respectively.

Because many V4 neurons had more than one orientation tuning peak we also computed a bimodal tuning index. First we identified the orientations of the two largest peaks in the orientation tuning curve, p_1 and p_2 , where $f(p_1) > f(p_2)$. Two troughs were then defined as the orientations of the lowest points, t_1 and t_2 , in either direction between the peaks, where $f(t_1) < f(t_2)$. The bimodal tuning index b was taken as the ratio of the difference between the smaller peak and trough, $d_2 = f(p_2) - f(t_2)$, to the difference between the larger peak and trough, $d_1 = f(p_1) - f(t_1)$

$$b = \frac{d_2}{d_1} \quad (9)$$

A neuron with two orientation tuning peaks and troughs of equal size will have a bimodal tuning index value of 1. As the relative size of one peak grows larger, index values grow smaller. Orientation tuning curves with only one peak have an index value of 0.

Spatial frequency peak and bandwidth were measured by fitting a Gaussian function to the spatial frequency tuning curve on a logarithmic scale, $g[\log(\omega)]$. Peak spatial frequency was taken as the peak of the Gaussian fit. Spatial frequency bandwidth was taken as the width of the Gaussian at half-height, divided by peak spatial frequency (De Valois et al. 1982a).

SELECTIVITY FOR COMPLEX FEATURES. If the SRF accurately describes response characteristics of V4 neurons then it should predict responses to any stimulus. Previous work showed that V4 neurons are selective for non-Cartesian (polar and hyperbolic) gratings over Cartesian gratings (Gallant et al. 1996). To test the SRF model we therefore used estimated SRFs to predict responses to both Cartesian and non-Cartesian gratings.

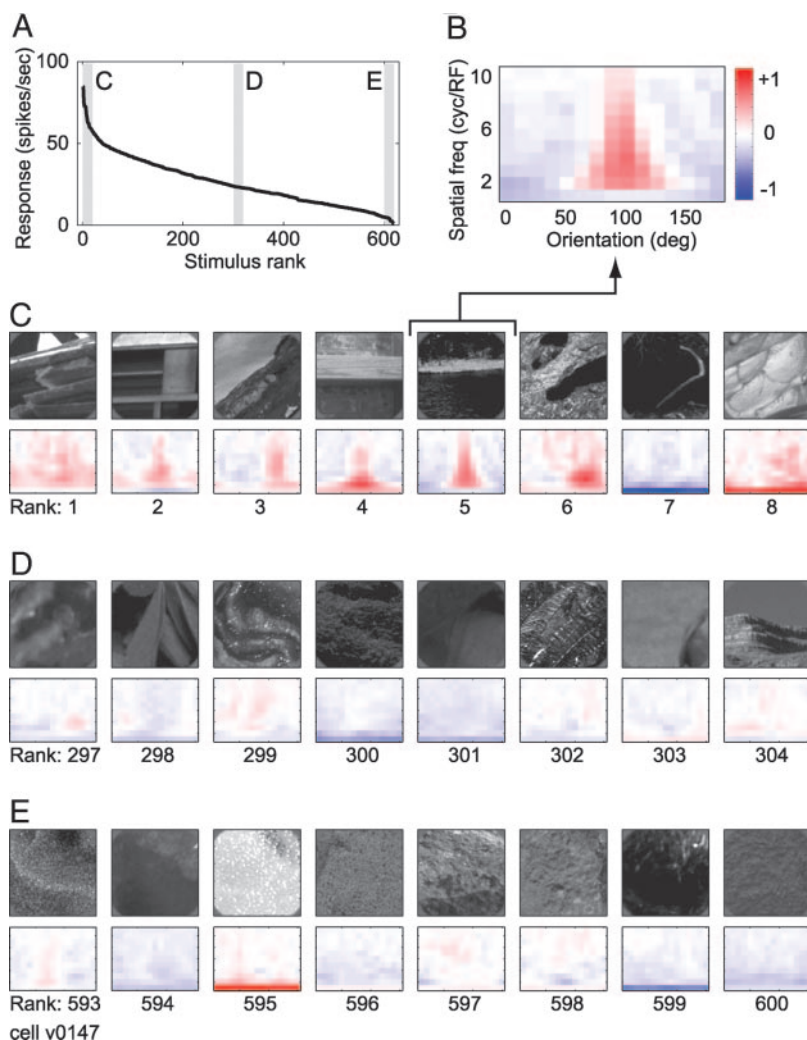


FIG. 1. Responses of one V4 neuron to a set of natural images. **A**: 600 different natural images were presented in the receptive field in random order (4 Hz, 4 repeats). Response strength was calculated by averaging firing rate 50–250 ms after stimulus onset over the 4 repeats. Here, responses are sorted by magnitude and vary from 0 to nearly 100 spikes/s. **B**: Fourier power transform of a typical natural image, enlarged from **C**. To highlight distinctive spectral features, the mean power spectrum (averaged over the entire set of natural images) was subtracted from the Fourier-transformed images. Without this subtraction images would be dominated by the $1/f^2$ spectrum typical of natural images. Red regions indicate power greater than the mean at the corresponding orientation (horizontal axis) and spatial frequency (vertical axis); blue indicates power less than the mean. In this example, the red region centered at 90° indicates high power at horizontal orientations. Image has been normalized so that power ranges between -1 and 1 . **C**: 8 natural images that evoked the strongest responses (shaded bar at left of **A**) and their Fourier power spectra (2nd row). Stimuli that evoked strong responses tend to have power at orientations of 90 or 150° . **D**: 8 natural images that evoked moderate responses (bar in middle of **A**) and their Fourier power spectra. **E**: 8 natural images that evoked the weakest responses (bar at right of **A**) and their Fourier power spectra. These stimuli tend to have low power in most orientation and spatial frequency channels.

Cartesian gratings were generated according to the function (Gallant et al. 1996)

$$s_c(x,y) = L_0 + C_0 \sin[\omega(x \cos \theta - y \sin \theta) + \phi] \quad (10)$$

Each Cartesian grating was described by its orientation θ , spatial frequency ω , and spatial phase ϕ . Mean luminance L_0 and contrast C_0 were normalized to match the root mean-square (RMS) contrast of the natural image set used to fit the SRF. Cartesian gratings were generated at 12 orientations, eight spatial frequencies (1.0 to 9.0 cycles per receptive field diameter), and four spatial phases (0 , 90 , 180 , and 270°).

Polar gratings were generated according to the function (Gallant et al. 1996)

$$s_p(x,y) = L_0 + C_0 \sin[\omega_c \sqrt{x^2 + y^2} + \omega_r \tan^{-1}(y/x) + \phi] \quad (11)$$

Each polar grating was described by its radial spatial frequency ω_r , concentric spatial frequency ω_c , and spatial phase ϕ . Mean luminance L_0 and contrast C_0 were normalized to match the RMS contrast of the natural image set used to fit the SRF. Polar gratings were generated at 12 radial frequencies (-5 to 6 cycles per rotation), eight concentric frequencies (1.0 to 9.0 cycles per receptive field diameter), and four spatial phases (0 , 90 , 180 , and 270°).

Hyperbolic gratings were generated according to the function (Gallant et al. 1996)

$$s_h(x,y) = L_0 + C_0 \sin[\omega \sqrt{(x \cos \theta - y \sin \theta)(x \sin \theta + y \cos \theta)} + \phi] \quad (12)$$

Each hyperbolic grating was described by its orientation θ , spatial

frequency ω , and spatial phase ϕ . Mean luminance L_0 and contrast C_0 were normalized to match the RMS contrast of the natural image set used to fit the SRF. Hyperbolic gratings were generated at eight orientations (0 to 80°), 12 spatial frequencies (1.0 to 7.0 cycles per receptive field diameter), and four spatial phases (0 , 90 , 180 , and 270°).

In addition to Cartesian and non-Cartesian gratings, we also used the SRFs to predict responses to a large set of 20,000 natural images. This stimulus set was generated using the same procedure as for the neurophysiological experiments (see above). To compare expected responses to those for gratings, each natural image patch was normalized to have the same mean luminance and RMS contrast as the gratings. (Without normalization, a large fraction of response variance can be attributed to variability in stimulus contrast rather than spatial patterns within the stimulus. Stimulus contrast was not normalized in the neurophysiological experiments. For this reason, the variability of responses in the experimental data was greater than that in the predictions; e.g., compare Figs. 1B and 7A.)

Predicted responses were generated using the same method as in the cross-validation procedure used for measuring the significance of visual tuning. Test stimuli were cropped, downsampled to 20×20 pixels, Hanning windowed, and transformed into the Fourier power domain according to Eq. 1. Predicted responses (spikes/s) were then generated for each SRF according to Eq. 2.

Neurons were grouped according to the stimulus class that evoked the strongest predicted response: Cartesian gratings, non-Cartesian (polar and hyperbolic) gratings, or natural images. The three stimulus

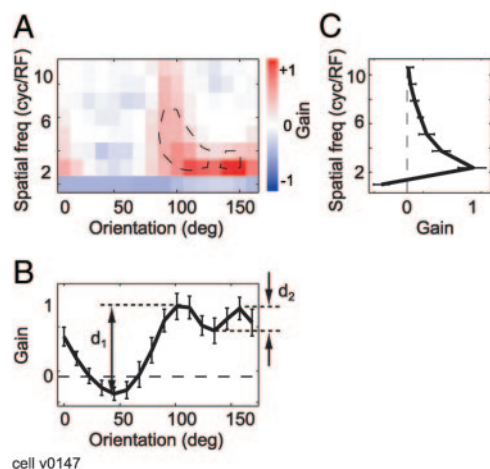


FIG. 2. Spectral receptive field (SRF) of a V4 neuron with bimodal orientation tuning. *A*: SRF describes selectivity in terms of a joint orientation-spatial frequency tuning surface. Red regions indicate orientations (horizontal axis) and spatial frequencies (vertical axis) correlated with increased responses (i.e., positive gain); blue indicates stimulus channels correlated with decreased responses. Contours mark 1 SD above zero after smoothing. This neuron is excited by patterns with horizontal orientation (90°) over a broad range of spatial frequencies and with oblique orientation (150°) at lower spatial frequencies. Stimulus power at low spatial frequencies (1 cycle/RF) tends to decrease responses. SRF was normalized so that gain ranges between -1 and 1 . *B*: an orientation tuning curve is derived by singular value decomposition of the SRF. This measures marginal tuning of the SRF in *A*, collapsed along the spatial frequency axis (Mazer et al. 2002). Each point shows the relative excitatory or inhibitory contribution of stimuli at the respective orientation, collapsed across spatial frequency. Peak orientation tuning is 129° and orientation bandwidth is 73° (measured by fitting a circular Gaussian function). This neuron has 2 orientation tuning peaks (90 and 150°). Relative magnitude of the peaks is quantified by the bimodal tuning index, computed from the ratio of the height of the smaller peak (d_2) to the larger peak (d_1). Bimodal tuning index for this neuron is 0.23 . *C*: a spatial frequency tuning curve is derived by the same singular value decomposition, effectively collapsing the SRF in *A* along the orientation axis. Peak spatial frequency tuning for this neuron is 2.5 cycles/RF and spatial frequency bandwidth is 1.1 octaves.

classes contained different numbers of exemplars. To ensure that this difference in sampling did not bias estimates of maximum expected response, we normalized responses according to the number of exemplars in each class. The smallest stimulus set was Cartesian gratings, containing 384 distinct patterns; the maximum Cartesian response was defined as the expected response to the single best Cartesian grating. The non-Cartesian grating class contained twice as many patterns (768); the maximum non-Cartesian response was defined as the average of expected responses to the *two* best non-Cartesian gratings. The natural image class contained 20,000 distinct images; the maximum natural image response was defined as the median of expected response to the 52 best images (0.26% , equivalent to $1/384$).

CONTOUR SELECTIVITY. We also used the SRFs to generate predicted responses to a set of curved contours that were used in a previous study of V4 (Pasupathy and Connor 1999). Contours were composed of two oriented segments (see Fig. 10*A*). The length of each segment was fixed to be one half the diameter of the classical receptive field. Segments were joined at one end and separated by an angle of 45 , 90 , 135 , or 180° . The joint between segments was either sharp or smooth. Smooth joints were generated by introducing a spline function between the two segments to produce seven different separation angles (the sharp and smooth 180° contours were the same). Eight absolute orientations were used for each separation angle, giving a total of 42 contour elements.

SIGNIFICANCE TESTING. Unless otherwise specifically mentioned, we used a jackknifed *t*-test to verify the statistical significance of our

findings (Efron and Tibshirani 1986). In many cases, a traditional *t*-test is sufficient to determine whether two mean values are significantly different. However, this test assumes that individual measurements follow a Gaussian distribution, and estimates of SE will be biased if the distributions are not Gaussian. The jackknifed *t*-test uses a bootstrapping procedure that avoids potential bias from non-Gaussian distributions in measurements of SE.

One situation in which a non-Gaussian distribution can be particularly problematic is when the sampled values lie near a hard boundary. We encountered this problem when testing the significance of the bimodal tuning index. If each jackknife estimate of the tuning index is generated independently, the distribution used to compute the SE will be biased toward positive values. This bias leads to artifactually small estimates of SE and can cause some neurons to appear to have significant bimodal tuning when they do not. To avoid this problem, we fixed the position of the peaks (p_1 and p_2) and troughs (t_1 and t_2) according to the orientation tuning curve averaged across jackknife estimates. Index values measured from the individual jackknifed tuning curves could then fluctuate below zero, leading to unbiased SE estimates.

RESULTS

Diversity of spectral tuning properties among V4 neurons

We characterized the spectral tuning properties of 103 V4 neurons in two animals while they performed a delayed match-to-sample task. The stimuli were sequences of natural image patches selected at random from a large image database (see examples in Fig. 1) and flashed in the receptive field at a rate of 3.5 – 4.5 Hz.

Figure 1*A* shows the responses of one V4 neuron to 600 distinct natural images, sorted by response magnitude. The visual response is defined as the firing rate 50 – 250 ms after stimulus onset (averaged over four presentations). For this neuron, responses range from 0 to nearly 100 spikes/s. The eight natural images that evoke the strongest responses are shown in the *top row* of Fig. 1*C*. Most of these images contain contours with either horizontal or oblique orientations (90 – 150°). The images that evoke average or weak responses (Fig. 1, *D* and *E*, respectively) have little in common with each other, although the least-preferred stimuli tend to have very low contrast.

This neuron responds most strongly to images with salient horizontal or oblique contours, but the precise spatial position of the contour does not appear to be important (Fig. 1*C*, *top row*). This is consistent with previous studies reporting that the responses of V4 neurons are often position and phase invariant (Desimone and Schein 1987; Gallant et al. 1996). Therefore the patterns that evoke large responses from this neuron might be clearer if we discard information about the precise spatial position of image features while preserving information about orientation and spatial frequency. One efficient way to do this is to compute the Fourier power spectrum of each image patch, as illustrated in Fig. 1*B*. After transformation into the Fourier power domain, each stimulus channel indicates the relative energy at a single orientation and spatial frequency in the original image, regardless of spatial position or phase. The Fourier power spectra of the effective images for this neuron have consistent peaks at orientations between 90 and 150° (Fig. 1*C*, *bottom row*).

It is often difficult to determine the response characteristics of a neuron by simply examining effective and ineffective

image patches. A better way to summarize the response properties of a single cell is to estimate the stimulus–response mapping function (Wu et al. 2006). We used linearized reverse correlation to estimate the spectral receptive field (SRF), a function that describes the mapping from the Fourier power transformation of the stimulus to the neural response (David and Gallant 2005; Theunissen et al. 2001). The SRF describes concisely which orientations and spatial frequencies tend to evoke responses. Figure 2A gives the SRF computed for the data in Fig. 1. Spectral domains shown in red indicate orientations and spatial frequencies that evoke strong responses (i.e., excitatory spectral channels); blue domains indicate orientations and spatial frequencies that suppress responses (i.e., inhibitory spectral channels). Consistent with the data in Fig. 1, the SRF reveals that this neuron is excited both by horizontal orientations and by oblique orientations near 150° . Furthermore, the SRF reveals that this neuron is sensitive to a higher and broader range of spatial frequencies at 90° than at 150° .

To visualize spectral tuning properties more clearly we extracted orientation and spatial frequency tuning curves from the SRF (Fig. 2, B and C, respectively; see Mazer et al. 2002). We measured three properties of orientation tuning: orientation peak, bandwidth, and bimodal tuning. The orientation tuning peak of the neuron illustrated in Fig. 2 is 129° and its orientation bandwidth is 73° . This neuron (and many others in our sample) has two distinct peaks in its orientation tuning curve. To measure bimodal orientation tuning we used a bimodal tuning index (Eq. 8); index values near 1.0 indicate that the secondary peak has the same height (measured between the shorter peak and shallower trough) to the primary peak and values near 0 indicate just a single peak in the orientation tuning curve. For this neuron the bimodal tuning index is 0.23, indicating that the secondary peak is 23% of the height of the primary peak. We also measured two properties of spatial-frequency tuning: peak and bandwidth. For this neuron peak spatial frequency tuning is 2.5 cycles per receptive field diameter (cyc/RF; or 0.31 cycles per degree, cyc/deg) and bandwidth is 1.1 octaves.

Some V4 neurons have simpler spectral tuning properties. Figure 3 shows the SRF of one V4 neuron whose orientation tuning profile resembles that typically encountered in area V1 (De Valois et al. 1982b). The orientation tuning of this neuron is unimodal (bimodal tuning index, 0.01), its orientation peak is 143° , and its orientation bandwidth is 29° . However, the same is not true for spatial frequency tuning. This neuron has a spatial frequency bandwidth of 1.7 octaves, substantially higher than that typically reported for V1 (De Valois et al. 1982a).

Comparison of V4 and V1 spectral tuning

We compared the tuning properties of our sample of V4 neurons to those of 45 neurons in primary visual cortex (V1), where spectral tuning properties are better understood (David et al. 2004). Neurons in V1 generally have much narrower and simpler spectral tuning than V4 neurons. One V1 SRF is shown in Fig. 4. The orientation tuning peak is 97° , orientation bandwidth is 29° , and tuning is nearly unimodal (bimodal tuning index, 0.02). The spatial frequency tuning peak is 2.5 cyc/RF and the spatial frequency bandwidth is 0.9 octaves.

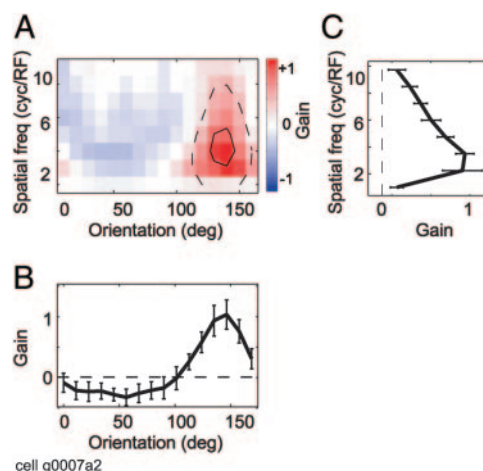


FIG. 3. SRF for a V4 neuron with narrow orientation tuning. Axes are as in Fig. 2. A: this neuron is excited by a relatively narrow range of orientations but a broad range of spatial frequencies. B: orientation tuning curve (peak, 143°) has narrow bandwidth (29°) and nearly unimodal tuning (bimodal tuning index, 0.01). C: spatial frequency tuning curve has a peak at 3.1 cycles/RF and a bandwidth of 1.7 octaves.

Across our sample of 103 V4 neurons, 87 (84%) had significant spectral tuning, and only this subset was used for comparison. Neurons without significant tuning either gave visual responses that could not be described by the Fourier power model or gave responses dominated by noise or other nonvisual inputs (see METHODS for selection criteria). Excluding these neurons increased the significance of some effects across the population but did not affect any trends.

Figure 5 compares orientation tuning properties in V4 and V1. The orientation bandwidth of V4 neurons varies widely and the median is 74.4° (Fig. 5A). A few V4 neurons (6%, 5/87) do not have measurable orientation tuning and instead respond equally to all orientations (white bar in Fig. 5A). In contrast, the median bandwidth across the sample of V1 neurons is just 43.7° (Fig. 5B), significantly lower than that in V4 ($P < 0.01$, Fig. 5C). Only a small number (5/45) of V1 neurons have orientation bandwidths $>90^\circ$. These values are comparable to those reported in previous studies of V4 (Desimone and Schein 1987) and V1 (De Valois et al. 1982b; Ringach et al. 2002) that used sinusoidal gratings.

As noted above (see Fig. 2), many V4 neurons in our sample have bimodal orientation tuning. Across the sample, the median bimodal tuning index for V4 neurons is 0.09 (Fig. 5D). Of these neurons, 28% (24/87) have a bimodal tuning index significantly greater than zero ($P < 0.05$; black bars in Fig. 5D). In contrast, the median bimodal tuning index in V1 is only 0.01 and only 11% (5/45) of V1 neurons have significant bimodal tuning ($P < 0.05$; Fig. 5E). The median bimodal tuning index for V1 neurons is significantly lower than that for V4 ($P < 0.01$; Fig. 5F).

Figure 6 compares spatial frequency tuning properties of V4 and V1 neurons. The median peak spatial frequency in V4 and V1 is not significantly different when measured in cycles per receptive field (cyc/RF), although they are likely to differ when measured in cycles per degree (see following text). The median is 2.6 cyc/RF in V4 (Fig. 6A) and 2.5 cyc/RF in V1 (Fig. 6B; $P > 0.25$, see Fig. 6C). Despite being similar to V1 on average, peak spatial frequency tuning varies more widely in V4, from <1 cyc/RF to over 6 cyc/RF. In contrast, the tuning of most V1

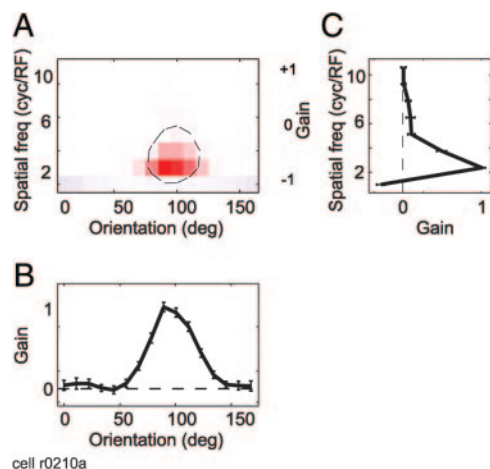


FIG. 4. SRF for a V1 neuron. Axes are as in Fig. 2. *A*: this neuron has simple spectral tuning and narrow orientation and spatial frequency bandwidth. *B*: orientation tuning curve (peak, 97°) has narrow bandwidth (29°) and is nearly unimodal (bimodal tuning index, 0.02). *C*: spatial frequency tuning curve has a peak at 2.5 cycles/RF and a bandwidth of just 0.9 octaves.

neurons falls between 2.0 and 3.5 cyc/RF. These spatial frequency tuning properties are similar to those reported in previous studies of V4 (Desimone and Schein 1987) and V1 (De Valois et al. 1982a) that used sinusoidal gratings.

We also observed substantial differences in spatial frequency bandwidth between V4 and V1 neurons. The median spatial frequency bandwidth in V4 is 1.2 octaves (Fig. 6D), which is significantly greater than the median of 0.9 octaves in V1 (Fig. 6E; $P < 0.01$, see Fig. 6F). In fact, nearly half of the V4 neurons in our sample (41/87, 47%) have spatial frequency tuning curves that extend outside the range of our analysis, compared with only about one fifth of V1 neurons (10/45, 22%; white bars in Fig. 6; SRFs were estimated over 1–10 cyc/RF). For neurons whose spatial frequency tuning extends beyond the tested range, bandwidth could be substantially broader than measured. Because these neurons are more common in V4, the true difference in bandwidth between areas is likely to be even larger than our data suggest.

In this report spatial frequency tuning was measured in cycles per receptive field rather than cycles per degree. Because the spatial extent of receptive fields is much larger in V4 than in V1 (Gattass et al. 1988), the median peak spatial frequency data suggest that the V4 neurons in our sample have a substantially lower peak spatial frequency than the V1 neurons when measured in cycles per degree. However, V4 and V1 neurons were sampled at different eccentricities with different cortical magnification factors, so a direct comparison is not possible. In any case, the possibility that V4 neurons may have lower peak spatial frequency tuning does not imply that high spatial frequency information is absent from their responses. Instead, high spatial frequency information appears to be integrated into the responses of neurons with large bandwidth that spans both high and low spatial frequencies. The increased bandwidth of V4 neurons enables a representation of visual features that integrates over a wide range of spatial frequencies, rather than the band-limited representation in V1.

Spectral tuning properties and feature selectivity

Previous studies of shape representation in V4 characterized neuronal tuning using restricted stimulus sets such as non-

Cartesian polar and hyperbolic gratings (Gallant et al. 1993, 1996), curved contours (Pasupathy and Connor 1999), and combinations of simple shape elements (Kobatake and Tanaka 1994). Because each of these studies probed a different part of shape parameter space it is difficult to draw any general conclusions from them about shape representation in V4. The SRF may provide a solution to this problem. Any visual stimulus can be described in terms of its orientation and spatial frequency spectrum, and responses to different spatial patterns can be interpreted in terms of the SRF.

To test the generality of SRFs, we used the SRF estimated for each V4 neuron in our sample to predict responses to both

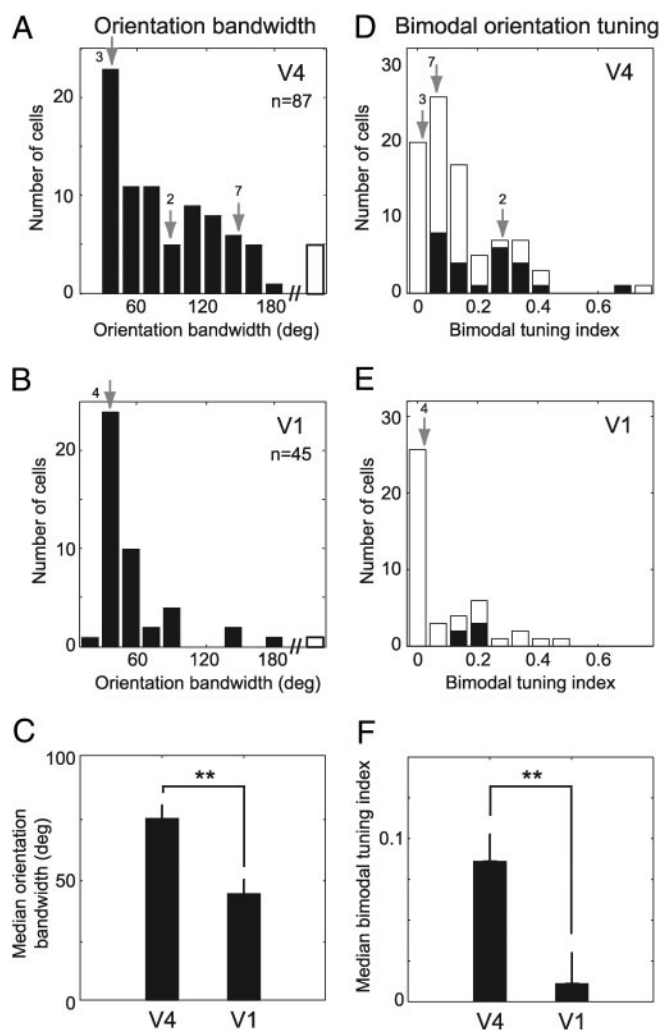


FIG. 5. Comparison of V4 and V1 orientation tuning properties. *A*: histogram of orientation bandwidth for 87 V4 neurons with significant visual tuning (median 74.4°). White bar at right denotes neurons that respond equally well to all orientations and are not orientation tuned. Numbered gray arrows indicate the values of examples shown in Figs. 2, 3, and 7. *B*: histogram of orientation bandwidth for 45 V1 neurons (median 43.7°). Gray arrow indicates the value of the example in Fig. 4. *C*: comparison of median orientation bandwidth between V4 and V1 neurons (error bars computed by jackknifing). Median for V4 neurons is significantly higher than that for V1 ($P < 0.01$). *D*: histogram of bimodal tuning index for V4 neurons (median 0.09). Black bars indicate neurons with index values significantly greater than zero ($P < 0.05$, jackknifed t -test). *E*: histogram of bimodal tuning index for V1 neurons (median 0.01). *F*: comparison of median bimodal orientation tuning between V4 and V1 neurons. Median bimodal tuning index is significantly greater in V4 than in V1 ($P < 0.01$).

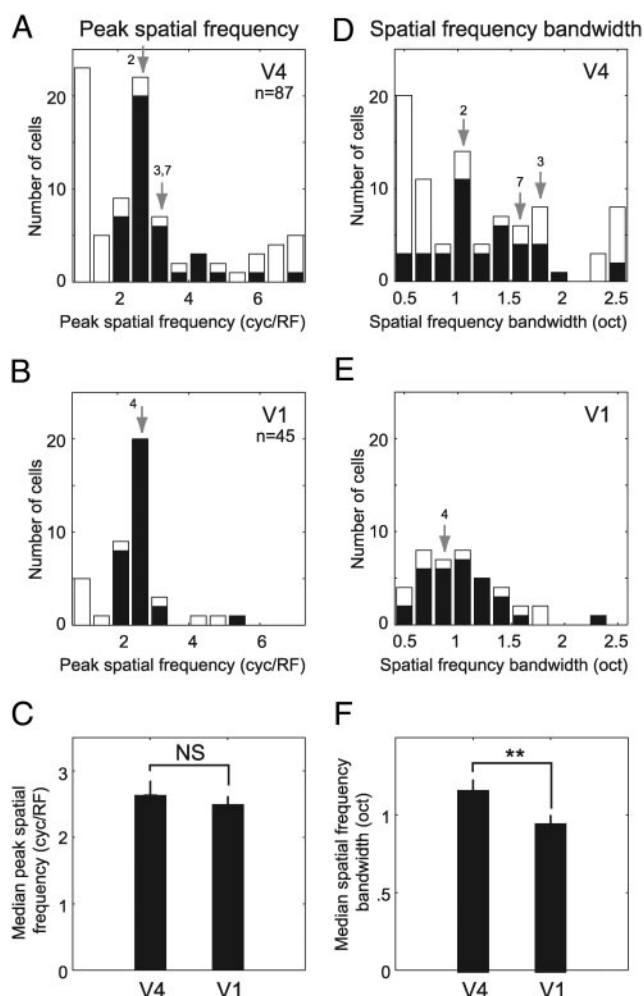


FIG. 6. Comparison of V4 and V1 spatial frequency tuning properties. *A*: histogram of peak spatial frequency tuning for 87 V4 neurons (median 2.6 cycles per receptive field diameter; cyc/RF). White bars indicate neurons (41, 47%) with spatial frequency tuning that extends above or below the range of spatial frequencies tested. Numbered gray arrows indicate the values of examples shown in Figs. 2, 3, and 7. *B*: histogram of peak spatial frequency tuning for 45 V1 neurons (median 2.5 cyc/RF). Gray arrow indicates the value of the example in Fig. 4. *C*: comparison of peak spatial frequency tuning between V4 and V1 neurons. Medians are not significantly different ($P > 0.25$). *D*: histogram of spatial frequency bandwidth for V4 neurons (median 1.2 octaves; shading as in *A*). *E*: histogram of spatial frequency bandwidth for V1 neurons (median 0.9 octaves). *F*: comparison of spatial frequency bandwidth between V4 and V1 neurons. Median bandwidth is significantly higher in V4 than in V1 ($P < 0.01$).

natural images and to synthetic stimuli that had been used in previous studies. We tested a stimulus set that was much larger than could be used in any actual physiology experiment. This consisted of 384 Cartesian gratings, 786 non-Cartesian (polar and hyperbolic) gratings, 20,000 random natural images, and 56 curved contour features. Selectivity for natural images and non-Cartesian gratings is described in this section; results obtained with curved contour features are presented in the following section.

Predictions for a representative V4 neuron are shown in Fig. 7. This neuron has very broad orientation tuning (bandwidth 151° ; SRF shown in Fig. 7A) and is band-pass for spatial frequency (peak, 3.1 cyc/RF; bandwidth 1.6 octaves). In the experimental data, the average response of this neuron was 24

spikes/s. Based on its spectral tuning, this neuron is predicted to respond most strongly to non-Cartesian gratings (best response 35 spikes/s; Fig. 7B). This response is slightly, but not significantly, greater than the best predicted response to natural images (34 spikes/s) and significantly greater than the best predicted response to Cartesian gratings (27 spikes/s, $P < 0.05$). (Best predicted responses are normalized for stimulus class size; see METHODS.) The members of each stimulus class predicted to evoke the five strongest and five weakest responses are shown in Fig. 7, C–E. Stimuli whose spectral power is matched to the excitatory domain of the SRF should evoke the strongest responses (*bottom row* of each panel). The orientation and spatial frequency of the best Cartesian gratings are aligned to the peak excitatory region of the SRF (Fig. 7C), but their orientation bandwidths are much narrower than the SRF bandwidth. The most effective non-Cartesian gratings (Fig. 7D) and natural images (Fig. 7E) have broad orientation bandwidth that more closely matches the excitatory domain of the SRF.

Responses predicted for a different V4 neuron are shown in Fig. 8 (SRF repeated from Fig. 2). In the experimental data, the average response of this neuron was 26 spikes/s. This neuron is predicted to give a significantly stronger response to natural images (best response 50 spikes/s; Fig. 8B) than to either non-Cartesian gratings (46 spikes/s, $P < 0.05$) or Cartesian gratings (38 spikes/s, $P < 0.05$). The stimulus predicted to evoke the strongest response from each class is shown in Fig. 8, C–E. As in the previous example, the spectral energy of the best Cartesian grating is aligned to the excitatory region of the SRF, but the narrow bandwidth does not match the broad, bimodal orientation tuning of the SRF (Fig. 8C). The most effective non-Cartesian, hyperbolic grating has a power spectrum that matches the SRF more closely but spans a much wider range of orientations than the excitatory domain of the SRF (Fig. 8D). The most effective natural image has a power spectrum that matches the bimodal structure of the excitatory SRF even more closely and so should evoke the largest response (Fig. 8E).

We classified each neuron according to the stimulus class predicted to evoke the strongest response and compared the fraction of neurons preferring each stimulus class (Fig. 9A). Cartesian gratings are predicted to be the most effective stimuli for only one quarter of the V4 neurons (21/87, 24%). For only four of these neurons, the best response to Cartesian gratings is significantly greater than that to either other stimulus class ($P < 0.05$). In contrast, non-Cartesian gratings should evoke the largest response from almost half of the V4 neurons (38/87, 44%; 13 significantly greater than either other class, $P < 0.05$). Natural images should evoke the largest response from the rest (28/87, 32%; one significant, $P < 0.05$). We evaluated other measures of selectivity (the difference between maximum and minimum response; sparseness of responses; Vinje and Gallant 2000) and found similar results (data not shown).

We used the same procedure to evaluate shape selectivity in our sample of 45 V1 neurons (Fig. 9A). In this case, we observed a much different pattern of selectivity. Cartesian gratings are predicted to be the most effective stimuli for the majority of V1 neurons (27/45, 60%). For 18 of these neurons, the predicted best response to Cartesian gratings is significantly greater than that to either other stimulus class ($P < 0.05$). Non-Cartesian gratings and natural images should each evoke

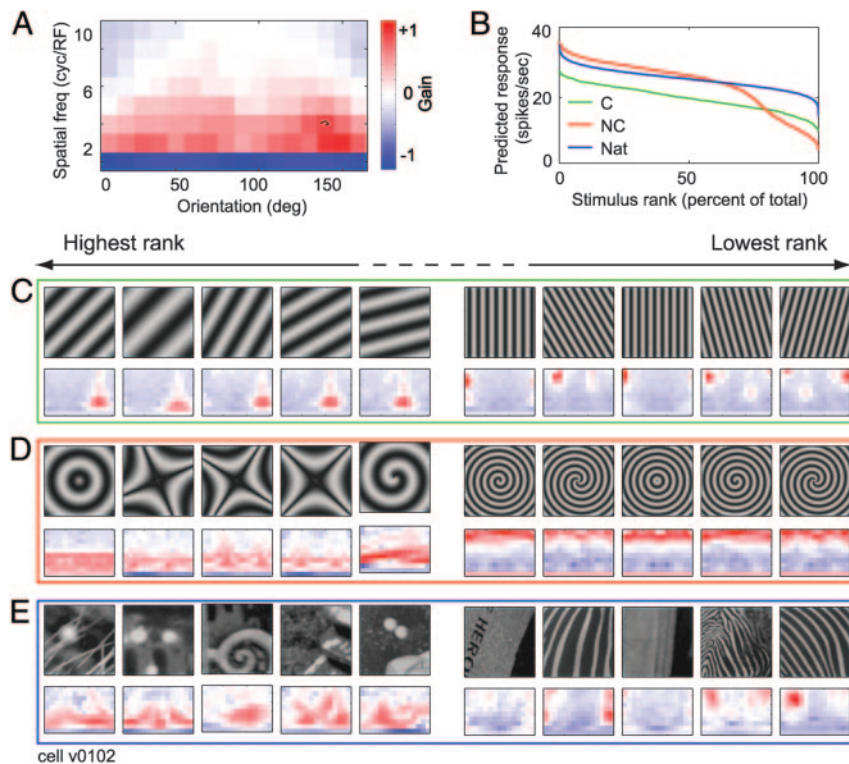


FIG. 7. Selectivity for Cartesian gratings, non-Cartesian gratings, and natural images of a single V4 neuron. *A*: SRF for this neuron indicates broad tuning for orientation (bandwidth 151°) and band-pass tuning for spatial frequency (peak, 3.1 cyc/RF; bandwidth 1.6 octaves). *B*: predicted responses to large sets of Cartesian gratings, non-Cartesian gratings, and natural images are shown sorted from strongest to weakest. Responses are predicted to be strongest for non-Cartesian gratings (NC; best response 35 spikes/s), nearly as strong for natural images (Nat; best response 34 spikes/s), but significantly weaker for Cartesian gratings (C; best response 27 spikes/s, $P < 0.05$). *C*: 5 Cartesian gratings predicted to give the strongest responses (*left*) and 5 predicted to give the weakest responses (*right*). Below each image is its Fourier power transform. Spectral energy of the preferred Cartesian gratings is aligned to the excitatory component of the SRF and nonpreferred gratings have power aligned to the high-frequency inhibitory component of the SRF. *D*: 5 most- and least-preferred non-Cartesian gratings. Strongest predicted responses are to polar and hyperbolic gratings with spectral structure similar to the SRF. *E*: 5 most- and least-preferred natural images. Preferred natural images tend to have localized, round structures also with spectra similar to the SRF.

the largest response from only a minority of V1 neurons (non-Cartesian: 7/45, 16%, two significantly greater than either other class, $P < 0.05$; natural images, 11/45, 24%, three significant, $P < 0.05$). The distribution of preferred stimulus class predicted across the sample of V1 neurons is significantly different from the distribution across V4 neurons ($P < 0.01$, jackknifed Hotelling's t -test).

Our analysis of shape selectivity demonstrates that differences in spectral tuning properties between V4 and V1 neurons are sufficient to explain the selectivity for complex patterns observed only in V4 neurons. To determine which aspects of spectral tuning might influence stimulus selectivity, we compared the tuning properties of V4 neurons classified according to the predicted best stimulus. The orientation bandwidth of neurons in the non-Cartesian class (median 129°) is significantly broader than that of cells in the Cartesian and natural image classes (Cartesian median bandwidth: 39° ; natural image bandwidth: 55° ; $P < 0.01$; Fig. 9B). In contrast, the bimodal tuning index is higher for the neurons in the natural image class (median 0.14) than for those in the Cartesian and non-Cartesian classes (median non-Cartesian index: 0.08; Cartesian index: 0.05; $P < 0.01$; Fig. 9C). There are no significant differences in peak spatial frequency tuning between the three classes of neuron (Fig. 9D). However, spatial frequency bandwidth is significantly greater for neurons in the natural image class (median 1.8 octaves) than that for those in the other two classes (median Cartesian bandwidth: 0.98 octaves; non-Cartesian bandwidth: 1.2 octaves; $P < 0.05$; Fig. 9E). Thus the selectivity for non-Cartesian gratings and natural images observed in V4 (Gallant et al. 1993; Kobatake and Tanaka 1994) can be explained by broad tuning bandwidth and complex orientation tuning profiles, properties that appear in V4 SRFs but not V1 SRFs.

The selectivity analysis presented thus far is based on simulations in which stimuli were centered in the receptive field. In V4, visual selectivity is invariant to changes in stimulus position on the order of one-half receptive field diameter (Gallant et al. 1996). The Fourier power model is invariant to small changes in position and thus should explain this invariance. However, if the spectral structure of a stimulus varies across space, selectivity could be affected by large spatial offsets. To address this issue we repeated the comparison of selectivity for V4 neurons with an expanded stimulus set. In the expanded set, stimuli of all three classes were positioned either in the receptive field center or offset by one-half receptive field diameter (horizontally, vertically, and diagonally). The pattern of selectivity within the expanded stimulus set (neurons preferring Cartesian gratings: 13/87, 15%; non-Cartesian gratings: 44/87, 51%; natural images: 30/87, 34%) is not significantly different from the distribution for the original set, in which stimuli appeared only in the receptive field center ($P > 0.5$, jackknifed Hotelling's t -test). Thus the pattern of selectivity predicted by the Fourier power model does not depend on the position of the stimulus in the receptive field.

Selectivity for curved-contour features in V4

One previous study of shape selectivity in area V4 used stimuli constructed by joining two oriented line segments in a sharp corner or curve (Pasupathy and Connor 1999). That study reported that many V4 neurons are selective for the angle separating the contour components and for the sharpness of the corner. To determine whether V4 SRFs can account for selectivity for these curvature features we used estimated SRFs to predict responses to the same contour configurations (Fig. 10A).

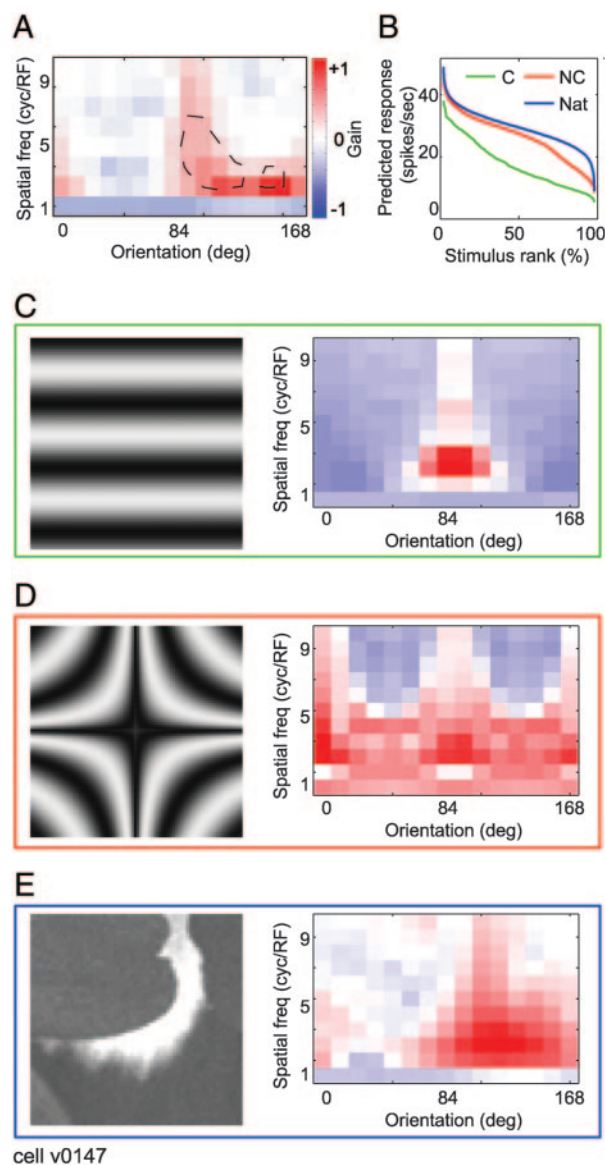


FIG. 8. Selectivity for Cartesian gratings, non-Cartesian gratings, and natural images for a V4 neuron with bimodal orientation tuning. *A*: SRF is repeated from Fig. 2 for reference. *B*: predicted responses to each stimulus class, sorted as in Fig. 7*B*. Strongest response is predicted for natural images (Nat; best response 50 spikes/s), significantly greater than for non-Cartesian gratings (NC; best response 46 spikes/s, $P < 0.05$), and Cartesian gratings (C; best response 38 spikes/s, $P < 0.05$). *C*: preferred Cartesian grating (Fourier power transform at right) has orientation and spatial frequency matched to the peak of the SRF. *D*: preferred non-Cartesian grating is a hyperbolic grating with spectral energy that overlaps and extends beyond the excitatory region of the SRF. *E*: preferred natural image is a curved pattern with strong similarity between its spectral structure and the excitatory tuning of the SRF.

We classified neurons according to which of the seven distinct separation angles and corner shapes are predicted to evoke the strongest response, disregarding absolute orientation. Selectivity varied widely across V4 neurons (neurons preferring 45° separation: 18/87, 21%; 90°: 27/87, 31%; 135°: 7/87, 7%; 180°: 33/87, 38%). Responses predicted for corner shape were equally diverse (neurons preferring sharp corners: 30/87, 34%; smooth corners: 30/87, 34%; straight contours: 27/87, 31%). The wide variability of preferred stimuli, including the

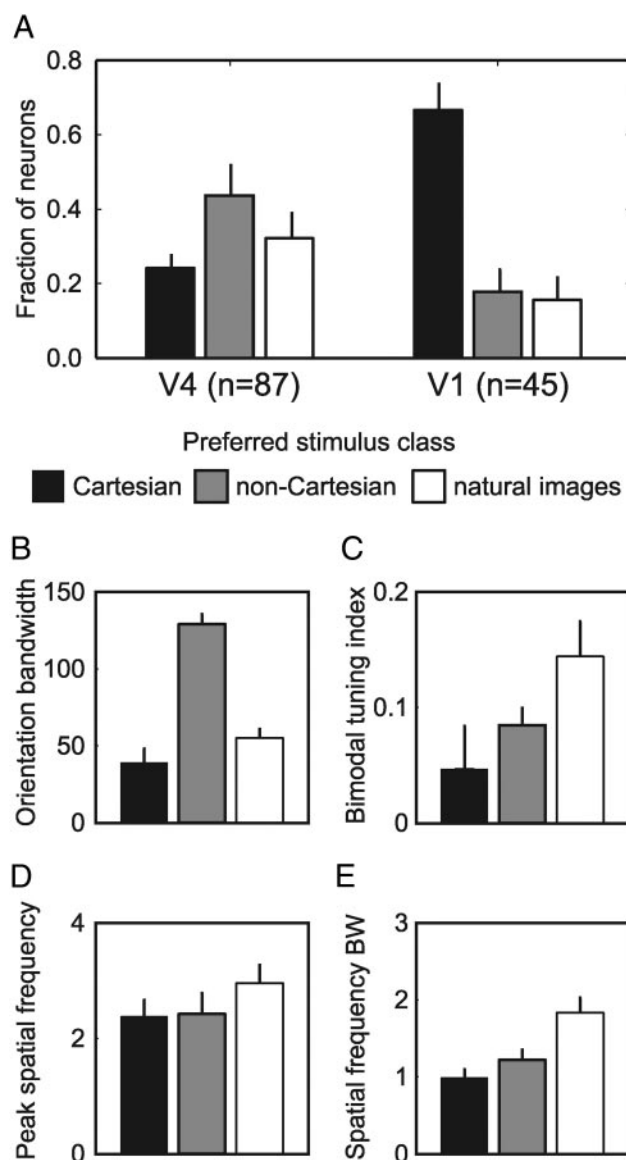


FIG. 9. Comparison of selectivity for Cartesian gratings, non-Cartesian gratings, and natural images in V4 and V1 neurons. *A*: fraction of V4 and V1 neurons selective for each stimulus class. Neurons were categorized according to the class of the stimulus for which their SRF was predicted to give the strongest response. Only 21/87 V4 neurons (24%) are selective for Cartesian gratings, whereas 38 V4 neurons (44%) are selective for non-Cartesian gratings and 28 (32%) are selective for natural images (bars at left). In contrast, the majority of V1 neurons (30/45, 67%) are selective for Cartesian gratings, whereas only 8 neurons (18%) are selective for non-Cartesian gratings and 7 neurons (16%) are selective for natural images (bars at right). Patterns of selectivity observed in V4 and V1 are significantly different ($P < 0.01$, jackknifed Hotelling's t -test). *B*: median orientation bandwidth of V4 neurons selective for each stimulus class. Neurons selective for non-Cartesian gratings have significantly greater bandwidth than neurons selective for either other class ($P < 0.01$). *C*: median bimodal tuning index of V4 neurons selective for each stimulus class. Neurons selective for natural images have significantly greater bandwidth than neurons selective for either other class ($P < 0.05$). *D*: median peak spatial frequency tuning of V4 neurons selective for each stimulus class. There are no significant differences between classes. *E*: median spatial frequency bandwidth of V4 neurons selective for each stimulus class. Neurons selective for natural images have significantly greater bandwidth than neurons selective for the other classes ($P < 0.01$).

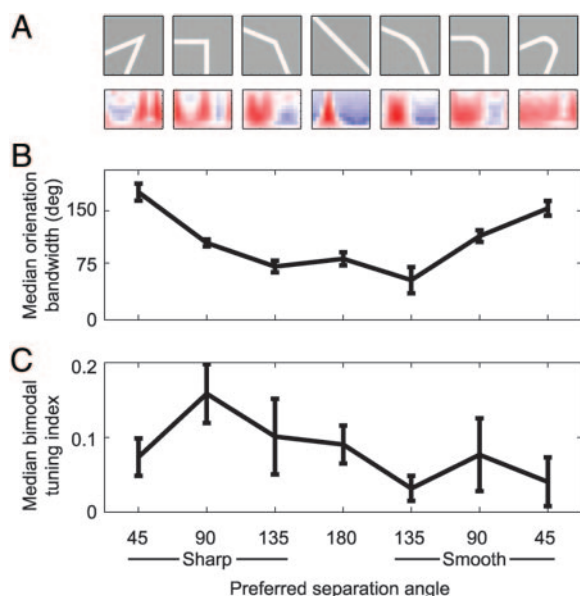


FIG. 10. Spectral tuning properties explain selectivity for curved contours in V4. *A*: contours generated by joining 2 line segments with a sharp corner at different separation angles (*left*) and contours containing the same separation angles but joined by a rounded corner (*right*). Fourier power transform of each stimulus appears in the 2nd row. Entire stimulus set contained these 7 contours at 8 different absolute orientations. Neurons were categorized according to the separation angle (45, 90, 135, or 180°) and corner (sharp or smooth) of the contour that produced the strongest predicted response from the SRF at any absolute orientation. *B*: median orientation bandwidth for neurons in each contour category. Error bars: 1 SE (estimated by jackknifing). Orientation bandwidth is negatively correlated with preferred separation angle ($r = 0.61$, $P < 0.01$). *C*: median bimodal tuning index for neurons in each contour category. Bimodal tuning index is greater for neurons that prefer sharp corners (median 0.15) than neurons that prefer round corners (median 0.05, $P < 0.01$).

bias toward a 180° separation angle, matches the previous neurophysiological observations in V4 (Pasupathy and Connor 1999).

We used the same procedure to evaluate contour selectivity in our sample of 45 V1 neurons. In this case, we observed a much simpler pattern than that for V4 neurons. Most V1 neurons are predicted to respond maximally to straight contours (20/45, 44%) or contours with a smooth corner and 135° separation angle (12/45, 27%).

To determine which aspects of spectral tuning could influence curvature selectivity, we compared the tuning properties of V4 neurons in each preferred contour class. We found a strong negative correlation between orientation bandwidth and the angle of the contour predicted to evoke the largest response: neurons predicted to prefer narrow separation angles (independent of corner shape) tend to have broad orientation bandwidth and vice versa ($r = 0.41$, $P < 0.001$; see Fig. 10*B*). Thus SRF orientation bandwidth can account for V4 selectivity for the separation angle of curved contours.

Our analysis of grating selectivity (see above) suggests that unimodal or bimodal orientation tuning in the SRF might also influence predicted stimulus selectivity. The median bimodal tuning index for the subset of V4 neurons predicted to respond most strongly to contours with sharp corners is 0.15, whereas the index is only 0.05 for those neurons predicted to respond best to contours with rounded corners. This difference is significant ($P < 0.01$; see Fig. 10*C*). Thus bimodal orientation

tuning in the SRF can account for selectivity for curved versus sharp contours.

DISCUSSION

This study tested the hypothesis that selectivity for complex visual features in area V4 neurons can be explained by their second-order spectral tuning properties. Spectral receptive fields (SRFs) estimated from neuronal responses to natural scenes predict most observations of shape selectivity in V4 that have been reported in studies using simpler parametric stimuli (Desimone and Schein 1987; Gallant et al. 1993; Kobatake and Tanaka 1994; Pasupathy and Connor 1999). SRFs can explain why non-Cartesian polar and hyperbolic gratings evoke larger responses than Cartesian gratings in V4 (Gallant et al. 1996). They can also explain important aspects of responses to curved contours (Pasupathy and Connor 1999). This selectivity is not simply implicit to the Fourier power model or the linearized reverse correlation algorithm; SRFs estimated for V1 neurons do not predict the same pattern of stimulus selectivity.

Shape selectivity predicted by V4 SRFs reflects the influence of specialized orientation and spatial frequency tuning properties that are not found in V1. These specializations include both increased orientation and spatial frequency tuning bandwidth and bimodal orientation tuning. The SRF profiles observed in V4 suggest that shape selectivity in this area is constructed by pooling of specific orientation and spatial frequency channels from more peripheral stages of visual processing. Although other mechanisms are likely to contribute to shape selectivity, the observed pooling of spectral channels alone is sufficient to explain the selectivity for complex patterns in V4.

Spectral receptive fields provide a general model of visual processing

The Fourier power model embodied in the SRF uses a second-order nonlinearity to describe spectral tuning properties and, at the same time, to account for phase- and position invariance (David and Gallant 2005; Freiwald et al. 2004). Despite its relatively simple analytical form, the Fourier power model can account for many previous observations of shape selectivity in area V4. It may seem surprising that such a simple model can describe such a wide range of observations, but very little quantitative data yet exist to either support or refute such simple models for V4 (Pollen et al. 2002; Wu et al. 2006).

Although the Fourier power model explains many aspects of shape selectivity in V4, it is not as comprehensive as models for more peripheral areas, such as V1 (Carandini et al. 1997; Daugman 1980; Jones and Palmer 1987). Several response properties previously reported in extrastriate cortex are not well described by second-order nonlinearities. These include tuning to the relative position of features in space (Gallant et al. 1996; Kobatake and Tanaka 1994; Pasupathy and Connor 2002), responses to figure-ground cues (Pasupathy and Connor 1999), and nonlinear spatial summation (Desimone and Schein 1987; Gustavsen et al. 2004).

The relative success of the Fourier power model rests on its incorporation of a specific nonlinear transformation into a general method for systems identification. The choice of nonlinearity was motivated from results of previous studies of

visual cortex that used small, restricted stimulus sets (Gallant et al. 1996; Mazer et al. 2002; Pasupathy and Connor 2002). A similar approach has proven effective for explaining some aspects of pattern selectivity in area MT (Rust et al. 2006). We suspect that a continued effort to expand systems identification approaches with known nonlinear responses will produce even more complete descriptions of neuronal tuning in V4 and other sensory cortical areas (Wu et al. 2006).

Bimodal orientation tuning in V4

More than one quarter of V4 neurons have more than one excitatory orientation tuning peak. Bimodal orientation tuning explains previous observations of selectivity for sharp corners (Pasupathy and Connor 1999). For similar reasons bimodal orientation tuning can also explain selectivity for non-Cartesian gratings (Gallant et al. 1993) and for geometrical patterns containing several oriented features (Kobatake and Tanaka 1994). Bimodal orientation tuning is a property distinct to extrastriate cortex. It occurs only rarely and with smaller magnitude in V1, where the majority of neurons are tuned to a single dominant orientation. Because all input to V4 must first pass through V1 (Felleman and Van Essen 1991), bimodal tuning in V4 must reflect a very precise rule for pooling inputs from V1. Bimodal V4 neurons must receive input from neurons selective for two distinct orientations, while excluding inputs from intermediate orientations. Much of the information that passes from V1 to V4 must pass first through V2 (Felleman and Van Essen 1991). Neurons in V2 are sometimes selective for complex patterns with multiple orientations (Hegde and Van Essen 2000). However, it is an open question whether V2 neurons possess a single dominant orientation peak, like V1, or whether they show bimodal tuning, like V4.

Shape selectivity in V4

One persistent question in visual neuroscience is whether tuning properties in V4 and other more central visual areas can be described along a small number of dimensions (Gallant et al. 1993; Wu et al. 2006). Identifying the relevant dimensions would allow for an efficient description of the tuning of any V4 neuron. An analogy can be made to V1, where numerous studies have concluded that Gabor wavelets provide an efficient description of the tuning space (Daugman 1980; Jones and Palmer 1987).

Tuning properties in V4 have been measured across specific stimulus sets: bars and gratings (Desimone and Schein 1987; Pollen et al. 2002), Cartesian and non-Cartesian gratings (Gallant et al. 1993), shape features (Kobatake and Tanaka 1994), or contour features (Pasupathy and Connor 1999). However, previous studies have not demonstrated the completeness or efficiency of any of these stimulus spaces. Untested stimulus dimensions that are correlated with the tested dimensions could describe tuning more efficiently. For example, within a set of curved contours, a neuron might appear to be tuned to a single curvature. However, changing the curvature of a contour also changes its spatial frequency spectrum (Zetzsche and Barth 1990). The neuron that appears to be tuned to a particular curvature may simply be tuned to the spatial frequency of the stimulus. Determining the stimulus feature for which the neuron is actually tuned requires a more general stimulus that tests both of these possibilities.

The present study used natural images as stimuli to produce a general and behaviorally relevant characterization of neuronal tuning properties. Estimated SRFs explain why non-Cartesian (polar and hyperbolic) gratings are most effective in V4 (Gallant et al. 1993, 1996), whereas Cartesian gratings are most effective in V1 (Jones and Palmer 1987). Non-Cartesian selectivity has also been observed in several human studies using a variety of techniques (Allison et al. 1999; Gallant et al. 2000; James et al. 1999; Wilkinson et al. 1998, 2000; Wilson et al. 1997). These findings suggest that the human homologue of macaque V4 (Hansen et al. 2005) will have SRF properties similar to those found in the macaque.

The SRFs described here motivate more appropriate and efficient parametric stimuli that can be used in future studies of V4. In one subset of V4 neurons in our sample, natural images are predicted to evoke larger responses than either Cartesian or non-Cartesian gratings. These neurons tend to have bimodal orientation tuning and broad spatial frequency bandwidth. Natural images often contain sharp edges and corners, features whose spectral properties are matched to these tuning properties. A more complete parametric stimulus set should include features that probe these tuning properties in addition to the tuning space spanned by Cartesian and non-Cartesian gratings.

In conclusion, the spectral receptive fields of V4 neurons estimated from responses to natural images reveal a diversity of tuning properties that are not observed in primary visual cortex: large orientation and spatial frequency bandwidth and bimodal orientation tuning. These tuning properties are sufficient to explain the emergent selectivity for many complex patterns in extrastriate cortex and suggest how information is pooled from more peripheral areas. Given its explanatory power and its ability to predict responses to arbitrary stimuli, the SRF provides a foundation for a general model of visual processing in area V4.

ACKNOWLEDGMENTS

The authors thank J. Mazer for development of the neurophysiological software suite, K. Gustavsen for valuable discussions regarding feature selectivity, and K. Hansen for comments on the manuscript.

Present addresses: S. V. David, 1103 A.V. Williams Building, University of Maryland, College Park, MD 20742; B. Y. Hayden, Department of Neurobiology, Box 3209, Duke University Medical Center, Durham, NC 27701.

GRANTS

This work was supported by National Institutes of Health grants to J. L. Gallant. S. V. David was partially supported by a National Science Foundation fellowship.

REFERENCES

- Adelson EH and Bergen JR.** Spatiotemporal energy models for the perception of motion. *J Opt Soc Am A* 2: 284–299, 1985.
- Aertsen AMHJ and Johannesma PIM.** A comparison of the spectro-temporal sensitivity of auditory neurons to tonal and natural stimuli. *Biol Cybern* 42: 145–156, 1981.
- Albrecht DG and Geisler WS.** Motion selectivity and the contrast-response function of simple cells in the visual cortex. *Vis Neurosci* 7: 531–546, 1991.
- Allison T, Puce A, Spencer DD, and McCarthy G.** Electrophysiological studies of human face perception: I. Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cereb Cortex* 9: 415–430, 1999.
- Bredfeldt CE and Ringach DL.** Dynamics of spatial frequency tuning in macaque V1. *J Neurosci* 22: 1976–1984, 2002.
- Brillinger DJ.** Some uses of cumulants in wavelet analysis. *J Nonparamet Stat* 6: 93–114, 1996.
- Carandini M, Heeger DJ, and Movshon JA.** Linearity and normalization in simple cells of the macaque primary visual cortex. *J Neurosci* 17: 8621–8644, 1997.

- Daugman JG.** Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Res* 20: 847–856, 1980.
- David SV and Gallant JL.** Predicting neuronal responses during natural vision. *Network* 16: 239–260, 2005.
- David SV, Vinje WE, and Gallant JL.** Natural stimulus statistics alter the receptive field structure of V1 neurons. *J Neurosci* 24: 6991–7006, 2004.
- DeAngelis GC, Ohzawa I, and Freeman RD.** Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. II. Linearity of temporal and spatial summation. *J Neurophysiol* 69: 1118–1135, 1993.
- Desimone R, Albright TD, Gross CG, and Bruce C.** Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci* 4: 2051–2062, 1984.
- Desimone R and Schein SJ.** Visual properties of neurons in area V4 of the macaque: sensitivity to stimulus form. *J Neurophysiol* 57: 835–868, 1987.
- De Valois RL, Albrecht DG, and Thorell LG.** Spatial frequency selectivity of cells in macaque visual cortex. *Vision Res* 22: 545–559, 1982a.
- De Valois RL, Yund EW, and Hepler N.** The orientation and direction selectivity of cells in macaque visual cortex. *Vision Res* 22: 531–544, 1982b.
- De Weerd P, Desimone R, and Ungerleider LG.** Cue-dependent deficits in grating orientation discrimination after V4 lesions in macaques. *Vis Neurosci* 13: 529–538, 1996.
- Efron B and Tibshirani R.** Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat Sci* 1: 54–77, 1986.
- Felleman DJ and Van Essen DC.** Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex* 1: 1–47, 1991.
- Field DJ.** Relations between the statistics of natural images and the response properties of cortical cells. *J Opt Soc Am A* 4: 2379–2394, 1987.
- Fisher NI.** *Statistical Analysis of Circular Data*. Cambridge, UK: Cambridge Univ. Press, 1993.
- Freiwald W, Tsao D, Tootell R, and Livingstone M.** Mechanisms of shape processing in macaque area V4. *Soc Neurosci Abstr* 370.10, 2004.
- Gallant JL, Braun J, and Van Essen DC.** Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science* 259: 100–103, 1993.
- Gallant JL, Connor CE, Rakshit S, Lewis JW, and Van Essen DC.** Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *J Neurophysiol* 76: 2718–2739, 1996.
- Gallant JL, Shoup RE, and Mazer JA.** A human extrastriate cortical area functionally homologous to macaque V4. *Neuron* 27: 227–235, 2000.
- Gattass R, Souza APB, and Gross CG.** Visuotopic organization and extent of V3 and V4 of the macaque. *J Neurosci* 8: 1831–1845, 1988.
- Gawne TJ and Martin JM.** Activity of primate V1 cortical neurons during blinks. *J Neurophysiol* 84: 2691–2694, 2000.
- Gustavsen K, David SV, Mazer JA, and Gallant JL.** Stimulus interactions in V4: a comparison of linear, quadratic and max models. *Soc Neurosci Abstr* 986.15, 2004.
- Hansen KA and Gallant JL.** Human visual area V4 has complementary, contiguous dorsal and ventral portions. *Soc Neurosci Abstr* 114.4, 2006.
- Hayden BY and Gallant JL.** Time course of attention reveals different mechanisms for spatial and feature-based attention in area V4. *Neuron* 47: 637–643, 2005.
- Hegde J and Van Essen DC.** Selectivity for complex shapes in primate visual area V2. *J Neurosci* 20: 1–6, 2000.
- Hubel DH and Wiesel TN.** Receptive fields and functional architecture of monkey striate cortex. *J Physiol* 195: 215–243, 1968.
- James TW, Wilkinson F, Wilson HR, Gati JS, Menon RS, and Goodale MA.** fMRI activation of fusiform face and object areas by concentric and radial gratings. *Invest Ophthalmol Vis Sci Suppl* 819, 1999.
- Jones J and Palmer L.** An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J Neurophysiol* 58: 1233–1258, 1987.
- Judge SJ, Richmond BJ, and Chu FC.** Implantation of magnetic search coils for measurement of eye position: an improved method. *Vision Res* 20: 535–538, 1980.
- Kobatake E and Tanaka K.** Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J Neurophysiol* 71: 856–867, 1994.
- Mazer JA and Gallant JL.** Goal-related activity in V4 during free viewing visual search. Evidence for a ventral stream visual salience map. *Neuron* 40: 1241–1250, 2003.
- Mazer JA, Vinje WE, McDermott J, Schiller PH, and Gallant JL.** Spatial frequency and orientation tuning dynamics in area V1. *Proc Natl Acad Sci USA* 99: 1645–1650, 2002.
- Merigan WH.** Basic visual capacities and shape discrimination after lesions of extrastriate area V4 in macaques. *Vis Neurosci* 13: 51–60, 1996.
- Merigan WH and Pham HA.** V4 lesions in macaques affect both single- and multiple-viewpoint shape discriminations. *Vis Neurosci* 15: 359–367, 1998.
- Pasupathy A and Connor CE.** Responses to contour features in macaque area V4. *J Neurophysiol* 82: 2490–2502, 1999.
- Pasupathy A and Connor CE.** Population coding of shape in area V4. *Nat Neurosci* 5: 1332–1338, 2002.
- Pollen DA, Przybyszewski W, Rubin MA, and Foote W.** Spatial receptive field organization of macaque V4 neurons. *Cereb Cortex* 12: 601–616, 2002.
- Ringach DL, Shapley RM, and Hawken MJ.** Orientation selectivity in macaque V1: diversity and laminar dependence. *J Neurosci* 22: 5639–5651, 2002.
- Rust NC, Mante V, Simoncelli EP, and Movshon JA.** How MT cells analyze the motion of visual patterns. *Nat Neurosci* In press.
- Schiller PH.** Effect of lesions in visual cortical area V4 on the recognition of transformed objects. *Nature* 376: 342–344, 1995.
- Schiller PH and Lee K.** The role of primate extrastriate area V4 in vision. *Science* 251: 1251–1253, 1991.
- Theunissen FE, David SV, Singh NC, Hsu A, Vinje WE, and Gallant JL.** Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Network Comput Neural Syst* 12: 289–316, 2001.
- Ungerleider LG and Mishkin M.** Two cortical visual systems. In: *Analysis of Visual Behavior*, edited by Ingle DG, Goodale MA, and Mansfield RJW. Cambridge, MA: MIT Press, 1982, p. 549–586.
- Van Essen DC, Olshausen B, Gallant JL, Press W, Anderson C, Drury H, Carman G, and Felleman D.** Anatomical, physiological, and computational aspects of hierarchical processing in the macaque visual cortex. In: *Structural and Functional Organization of Neocortex*, edited by Albowitz B, Albus K, Khunt U, Nothdurft H-C, and Wahle P. New York: Springer-Verlag, 1994, p. 317–329.
- Vinje WE and Gallant JL.** Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* 287: 1273–1276, 2000.
- Vinje WE and Gallant JL.** Natural stimulation of the non-classical receptive field increases information transmission efficiency in V1. *J Neurosci* 22: 2904–2915, 2002.
- Wilkinson F, James TW, Wilson HR, Gati JS, Menon RS, and Goodale MA.** An fMRI study of the selective activation of human extrastriate form vision areas by radial and concentric gratings. *Curr Biol* 10: 1455–1458, 2000.
- Wilkinson F, Wilson HR, and Habak C.** Detection and recognition of radial frequency patterns. *Vision Res* 38: 3555–3568, 1998.
- Wilson HR, Wilkinson F, and Asaad W.** Concentric orientation summation in human form vision. *Vision Res* 37: 2325–2330, 1997.
- Wu MC-K, David SV, and Gallant JL.** Complete functional characterization of sensory neurons by system identification. *Annu Rev Neurosci* 29: 477–505, 2006.
- Zetsche C and Barth E.** Fundamental limits of linear filters in the visual processing of two-dimensional signals. *Vision Res* 30: 1111–1117, 1990.