Spatial Structure and Symmetry of Simple-Cell Receptive Fields in Macaque Primary Visual Cortex

DARIO L. RINGACH

Department of Neurobiology, Psychology, and Brain Research Institute, University of California, Los Angeles, California 90095

Ringach, Dario L. Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *J Neurophysiol* 88: 455–463, 2002; 10.1152/jn.00881.2001. I present measurements of the spatial structure of simple-cell receptive fields in macaque primary visual cortex (area V1). Similar to previous findings in cat area 17, the spatial profile of simple-cell receptive fields in the macaque is well described by two-dimensional Gabor functions. A population analysis reveals that the distribution of spatial profiles in primary visual cortex lies approximately on a one-parameter family of filter shapes. Surprisingly, the receptive fields cluster into even- and odd-symmetry classes with a tendency for neurons that are well tuned in orientation and spatial frequency to have odd-symmetric receptive fields. The filter shapes predicted by two recent theories of simple-cell receptive field function, independent component analysis and sparse coding, are compared with the data. Both theories predict receptive fields with a larger number of subfields than observed in the experimental data. In addition, these theories do not generate receptive fields that are broadly tuned in orientation and low-pass in spatial frequency, which are commonly seen in monkey V1. The implications of these results for our understanding of image coding and representation in primary visual cortex are discussed.

INTRODUCTION

Simple cells in V1 behave to a large extent as linear spatio-temporal filters (Carandini et al. 1997; De Valois et al. 1979; Jones and Palmer 1987b; Movshon et al. 1978). A better understanding of simple-cell function may be gained if we had a summary of the distribution of filter shapes in primary visual cortex. In this study, the two-dimensional spatial structure of simple-cell receptive fields (RFs) in primate V1 was measured and analyzed. Available data on the two-dimensional structure of RFs in the monkey have been obtained via indirect methods (Parker and Hawken 1988) or limited to the study of the line-weighting function (De Valois et al. 2000; Hawken and Parker 1987). In the present study, direct measurements of the two-dimensional spatial structure of RFs in macaque V1 were obtained with a subspace reverse-correlation method (Ringach et al. 1997b). The results are compared with those obtained in cat area 17 (DeAngelis et al. 1993a,b; Jones and Palmer 1987a). In agreement with these studies, the profiles of simple-cell RFs in the macaque are well described by a Gabor function: a product of a Gaussian envelope and a sinusoid (Jones and Palmer 1987a; Kulikowski and Bishop 1981; Marcelja

1980). In addition, it was found that the spatial RF profiles cluster into even- and odd-symmetric classes, as conjectured by Movshon et al. (1978). This was a somewhat surprising result because all previous studies in cat area 17 report a uniform distribution of spatial phases (DeAngelis et al. 1993a; Field and Tolhurst 1986; Hamilton et al. 1989; Jones and Palmer 1987b).

In the second part of this study, the measured distribution of RFs in V1 is compared with the predictions of two recent theories of simple-cell function: independent component analysis (ICA) and sparse coding (SC) (Bell and Sejnowski 1997; Olshausen and Field 1996, 1997; van Hateren and Ruderman 1998; van Hateren and van der Schaaf 1998). The basic principle underlying these theories is that shapes of simple-cell RFs are designed to provide an efficient representation of natural scenes (for review, see Simoncelli and Olshausen 2001). These theories have received significant attention as they suggest that a few simple theoretical principles explain the distribution of RFs in V1. However, detailed comparisons between the predictions of such theories to the experimental data have been scarce and limited to one-dimensional measurements (based on the line-weighting function) of the RF (van Hateren and van der Schaaf 1998). Here, I compare the two-dimensional shape of the filters in the theoretical predictions and the experimental data.

METHODS

Preparation and recording

Acute experiments were performed on adult Old World monkeys (*Macaca fascicularis*) weighing between 2.5 and 5.1 kg. The methods of preparation and single-cell recording are essentially the same as those described in Ringach et al. (1997a). Animals were tranquilized with acepromazine (50 μ/kg im), then anesthetized with ketamine (30 mg/kg im), and maintained on opioid anesthetic (sufentanil citrate, 6 μg $\cdot$ kg^{-1} $\cdot$ h^{-1} iv) for the surgery. For recording, anesthesia was continued with sufentanil (6 μg $\cdot$ kg^{-1} $\cdot$ h^{-1}), and paralysis was induced with pancuronium bromide (0.1–0.2 mg $\cdot$ kg^{-1} $\cdot$ h^{-1}). Electrocardiogram (EKG), electroenchephalogram (EEG), and end-tidal CO_2 were continuously monitored. Blood pressure was measured noninvasively at 5-min intervals. Body temperature was maintained at 37°C. Extracellular action potentials were recorded with glass-coated tungsten microelectrodes, exposed tips 5–15 μm (Merrill and Ains-

Address for reprint requests: Dept of Neurobiology and Psychology, Franz Hall, Rm 8441B, University of California, Los Angeles, CA 90095-1563 (E-mail: dario@ucla.edu).

worth 1972). Electrical signals were amplified in the conventional manner, and spikes were discriminated using a two-channel window sorter, which generated transistor-transistor logic (TTL) pulses that were accumulated as event times by the computer (with 1-ms accuracy). Strict criteria for single-unit recording included: fixed nerve impulse height and waveform and absence of impulse intervals shorter than an absolute refractory period. In most of the experiments described here, data were collected by a CED 1401+ laboratory interface connected to a PC. Stimuli were generated on a Silicon Graphics O2 and displayed on monitor at refresh rate of 100 Hz. For all displays, the mean luminance was between 55 and 65 cd/m^2. The displays were calibrated and linearized by lookup tables. A Photo Research Model 703-PC spectro-radiometer was used to calibrate the display screens.

RF measurement

Each cell was stimulated monocularly via the dominant eye and characterized by measuring its steady-state response to conventional drifting sinusoidal gratings (the nondominant eye was occluded). With this method, we measured basic attributes of the cell, including spatial and temporal frequency tuning, orientation tuning, contrast response function, and color sensitivity, as well as area, length, and width tuning curves. To classify neurons as simple or complex, we determined the ratio between the modulated response (1st harmonic or F_1) and the mean response (DC component or F_0) for a drifting sine grating of optimal spatial frequency, temporal frequency, and orientation. A cell was classified as simple if $F_1/F_0 > 1$ for the optimal stimulus condition; otherwise it was classified as complex (Skottun et al. 1991).

The spatiotemporal RF of simple cells was measured using subspace reverse correlation (Ringach et al. 1997b). The measurement technique is a variant of the standard reverse-correlation method where, instead of white noise or sparse dots, the input is a sequence composed of a finite number of orthonormal sinusoidal gratings at various spatial frequencies, orientations, and spatial phases. In the standard reverse-correlation paradigm, one correlates the response of the neurons with the luminance values at each pixel on a grid. This provides the coefficients with which each pixel should be summed to generate a RF estimate. When the input is a sequence of random gratings at various spatial frequencies, orientations, and phases, the correlations between the response and the appearance of the gratings give the coefficients with which such grating (or basis function) should be summed to generate the RF estimate. In essence, the basic idea of the technique is to estimate the RF by estimating Fourier coefficients of the RF rather than estimating the RF in the space domain. A detailed account of the method can be found in Ringach et al. (1997b). As with the standard method, the technique provides an estimate of the RF up to a scaling factor.

It is important during the reverse correlation mapping experiments to have no eye movements. Even in the paralyzed animal (*Pancuronium bromide*), eye movements are sometimes observed. Eye movements are revealed by changes in the modulation of simple-cell responses stimulated with drifting sinusoidal gratings. Specifically, repeats of the same stimulation trial can yield responses that appear to be translations of one another in time. Eye movements are responsible for such changes in temporal phase of the response. Relatively small eye movements can have important consequences for the measurement of the spatial profile of the RF. In the macaque parafovea, for example, it is typical for cells to respond optimally to sinusoidal gratings of ≈4 cycles/° (De Valois et al. 1982). The distance between excitatory and inhibitory subfields in such a cell would be ~7.5 min of arc. Thus movements as small as 2 min of arc would be enough to seriously corrupt the measurements as such a shift represents a change of 48° in spatial phase. To minimize the effect of eye movements, only data for which the eyes appeared stable when stimulated with drifting sinusoidal gratings, before and after the reverse correlation experiment, were considered. This was done by calculating the variance of the phase of the first-harmonic response on a cycle-by-cycle

basis. Furthermore, in cases where sufficient data were collected, the RF of the neurons was estimated from the first and second halves of the reverse correlation data. If the center of the fitted Gabor envelope or the phase of the fit was significantly different, the data were discarded. A total of 37 neurons of 107 measurements (34%) were discarded based on these considerations, resulting in a data set of $n = 70$.

Typical measurements of spatiotemporal RFs using subspace reverse correlation are shown in Fig. 1. Because the method recovers the linear kernel of the system up to multiplication by a scalar, one can normalize the kernels so that their maximum absolute value over space and time is one. Figure 1*A* illustrates the RF of an orientation-tuned cell. Each panel in Fig. 1*A* is a slice of the impulse response of the filter at different delay times, which are indicated at the *inset*. Red areas are those where light increments induce the cell to fire more than its mean rate. Areas in blue are those where light increments induce the cell to fire less than its mean rate. The RF in Fig. 1*A* has two elongated subfields, one excitatory and one inhibitory. The delay time at which the
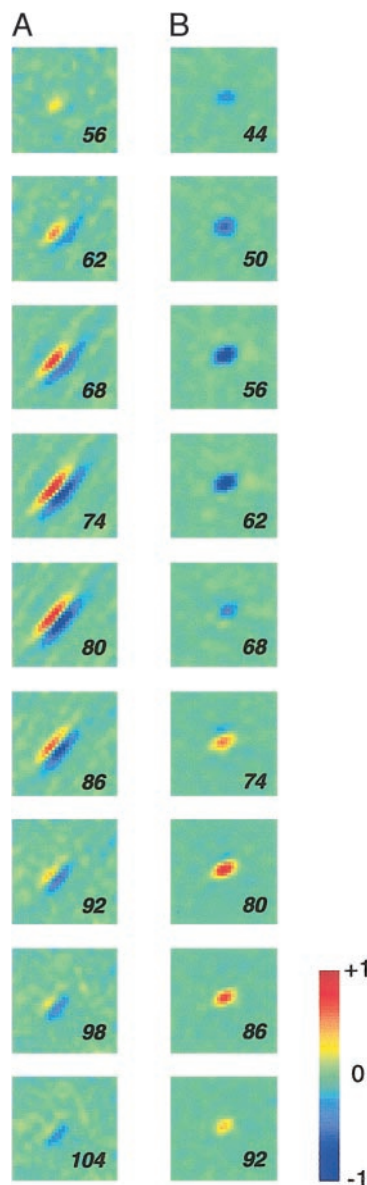

FIG. 1. Spatiotemporal receptive fields of simple cells in primate V1 estimated via subspace reverse correlation. *A*: example of an oriented receptive field with 2 subfields. *B*: example of a cell that is broadly tuned in orientation and low-pass in spatial frequency. The value of τ (in milliseconds) at which each frame has been computed is shown at the *inset*.
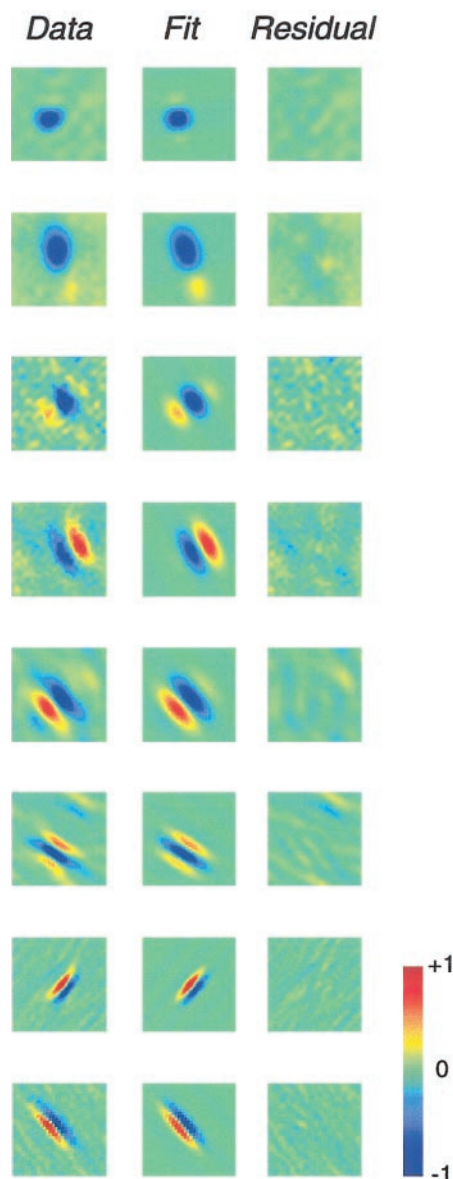
Data Fit Residual

FIG. 2. Two-dimensional Gabor fits to the data. Left: examples of the measured receptive fields. *Middle*: the best Gabor fit in the least squares sense. *Right*: the residual images. In general, 2-dimensional Gabor functions provide a good representation of the shapes of receptive fields in V1.

spatial filter achieves maximal energy (or variance) is defined as the optimal delay time. In this case, the optimal delay time is 74 ms. A second example of a cell that is broadly tuned for orientation and has a biphasic temporal response is shown in Fig. 1*B*. At the optimal delay time of 56 ms, the spatial profile of this cell resemble a circularly symmetric blob. In this study, the spatial profile of V1 RFs were analyzed at their optimal delay time (DeAngelis et al. 1993b).

To analyze the spatial profile of RFs, a two-dimensional Gabor function was fit to the data (Jones and Palmer 1987a)

$$h(x', y') = A \exp(-(x'/\sqrt{2}\sigma_{x'})^2 - (y'/\sqrt{2}\sigma_{y'})^2) \cos(2\pi f x' + \phi) \quad (1)$$

where (x', y') is obtained by translating the original coordinate system by (x_0, y_0) and rotating it by θ

$$x' = (x - x_0) \cos \theta + (y - y_0) \sin \theta$$

$$y' = -(x - x_0) \sin \theta + (y - y_0) \cos \theta \quad (2)$$

In this coordinate system, the cosine function in the Gabor varies only along the x' axis. Notice that one of the axes of the Gaussian envelope aligns with the x' axis and the other with the y' axis. The parameter A is the amplitude, $\sigma_{x'}$ and $\sigma_{y'}$ represent the width of the Gaussian envelope along the x' and y' axes, respectively, f is the spatial frequency of the sinusoidal grating in cycles/degree, and ϕ is the spatial phase of the grating. A spatial phase of $\phi = 0$ results in an even symmetric kernel, while a spatial phase of $\phi = \pi/2$ gives an odd-symmetric kernel. I did not find it necessary to add a parameter to vary the relative orientation of the envelope with respect to the orientation of the grating as done by Jones and Palmer (1987a). Such a parameter helped only in a small number of cases.

Figure 2 illustrates examples of the spatial profiles measured at their optimal delay time together with the best fitting Gabor function (in the least squares sense) and the corresponding residual images. It can be seen that, similar to previous reports in cat area 17 (Jones and Palmer 1987a), the two-dimensional Gabor function provides a reasonable summary of the shape of spatial RFs profiles in macaque primary visual cortex. This is also evidenced in the distribution of the fraction of unaccounted variance over the population (Fig. 3). Assuming independent and additive noise, the fraction of unaccounted variance is defined as $\sigma_{err}^2/(\sigma_{data}^2 - \sigma_{noise}^2)$. Here, σ_{err}^2 represents the variance of the residual image, σ_{data}^2 is the variance of the estimated RF at the optimal delay time, and σ_{noise}^2 is the variance of the noise estimated as the variance of the RF map at a delay of 0 ms.

To compare the experimental data to the predictions of existing theories, the shape of RFs predicted by ICA and SC were analyzed in a similar fashion. The ICA data have been provided by Dr. Hans van Hateren and colleagues and are available on the web from http://hlab.phys.rug.nl/demos/ica/comp_filt.html. I report results for the data set with log intensity transformation and dimension reduction. Similar results were obtained for the linear intensity data set. Dr. Bruno Olshausen provided the RF predictions of SC. These data correspond to the implementation reported in Olshausen (2001).

RESULTS

To analyze the structure of RFs in V1 over the population, a scatter-plot of $n_x = \sigma_x f$ versus $n_y = \sigma_y f$ based on the fitted parameters (Fig. 4) was first constructed. One can think of these values as the number of sinusoidal cycles of the Gabor carrier fitting in a segment of length σ_x and σ_y, respectively. In other words, the size of the Gaussian envelope is measured in units of the period of the sinusoidal grating, $T = 1/f$. This visualization is invariant to translations, rotations, isotropic scaling, and the symmetry (or spatial phase) of the RF. Invariance to isotropic scaling of the RF results because, for any α, we have $n_x = (\alpha\sigma_x)(f/\alpha) = \sigma_x f$ (and the same holds for n_y).
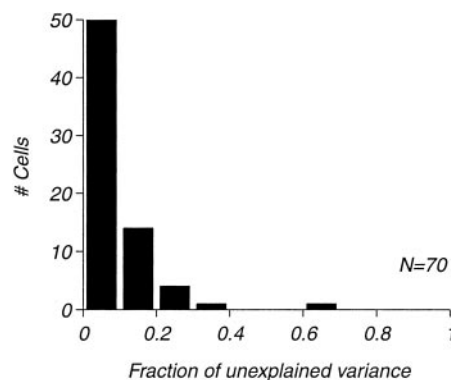

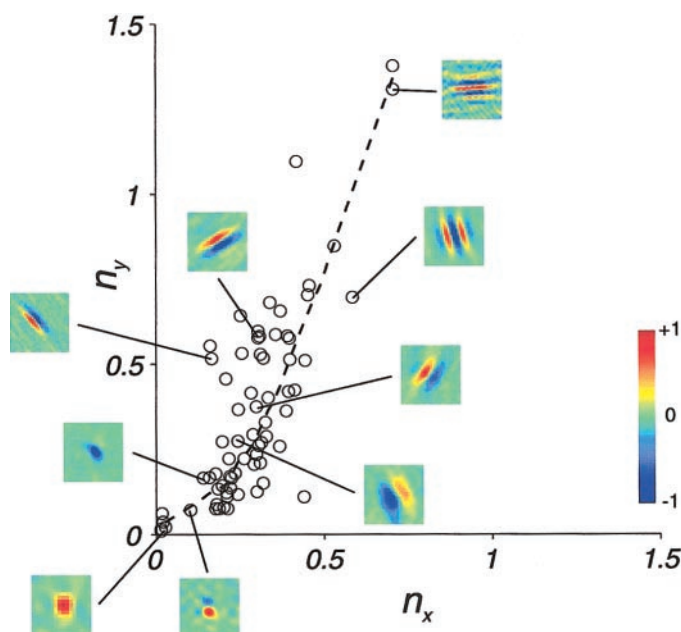FIG. 3. Distribution of the amount of variance unaccounted for in the Gabor fits.

FIG. 4. Distribution of receptive field shapes in the (n_x, n_y) plane. A number of receptive fields are shown along the distribution. - - -, a smooth version of the scatterplot. Blob-like receptive fields are mapped to points near the origin. Neurons with several subfields are mapped to points away from the origin.

Invariance with respect to translation, rotation, and spatial phase is obtained simply because (n_x, n_y) do not depend on the values of x_0, y_0, θ, and ϕ. The distribution of the spatial phase variable is analyzed separately in the following text.

Figure 4 shows that the distribution of (n_x, n_y) in macaque V1. The shape of some RFs at different locations along the distribution are also shown. Blob-like RFs are mapped to points near the origin. RFs with a number of elongated subfields are mapped to points away from the origin. Interestingly, the distribution of (n_x, n_y) appears to lie, approximately, on a one-dimensional curve. This implies a constraint between the variables: (n_x, n_y) are correlated. a smoothed version (estimated by local robust linear regression) of the scatter plot (Cleveland and Devlin 1988) is provided (- - -). For comparison with previously published results in cat, the data in Table 1 of Jones and Palmer (1987a) are re-plotted here using the same analysis (Fig. 5, ×). Overall, the data in cat area 17 and macaque V1 are comparable. The cat data appear to be shifted slightly to the left of the monkey data, suggesting a smaller number of subfields. However, I discuss in the following text a methodological difference between these studies that might explain this discrepancy.

To analyze the distribution of the spatial phase variable the following should be noted. A consequence of *Eq. 1* is that if two RFs are the same except for their spatial phase, a number of simple relationships hold. First, if $\phi_2 = \phi_1 + \pi$, the RFs are identical up to a change in sign, $h_2(x', y') = -h_1(x', y')$. Second, if $\phi_2 = k(\pi/2) + \beta$ and $\phi_1 = k(\pi/2) - \beta$, where k is even, and β an arbitrary angle, the RFs are mirror symmetric around the x axis and $h_2(x', y') = h_1(-x', y')$. Third, if $\phi_2 = k(\pi/2) + \beta$ and $\phi_1 = k(\pi/2) - \beta$, where k is odd, the RFs are related by mirror symmetry and a change in sign, $h_2(x', y') = -h_1(-x', y')$. As mirror symmetry and flips in sign do not change the basic shape of the filter, we discard these transfor-
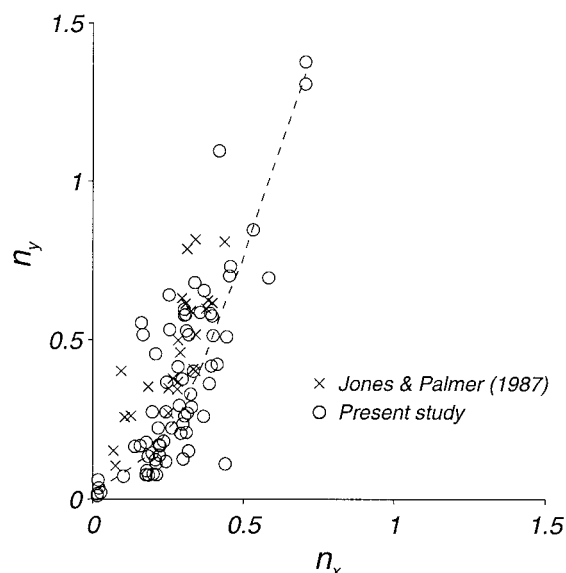

FIG. 5. Comparison between monkey and cat simple-cell receptive fields. ○, data obtained in macaque (present study); ×, the data in Table 1 of Jones and Palmer (1982) re-plotted in the (n_x, n_y) plane.

mations and define the effective range of the spatial phase parameter to be $0 \leq \phi \leq \pi/2$. Mapping an arbitrary spatial phase angle to this range can be achieved by defining $\hat{\phi} = \arg(|\cos \phi| + i|\sin \phi|)$. Even symmetry is obtained when $\hat{\phi} = 0$ and odd symmetry when $\hat{\phi} = \pi/2$. These relationships are summarized graphically in Fig. 6*A* (see also, Fig. 1 in Field and Tolhurst 1986 and the accompanying discussion).
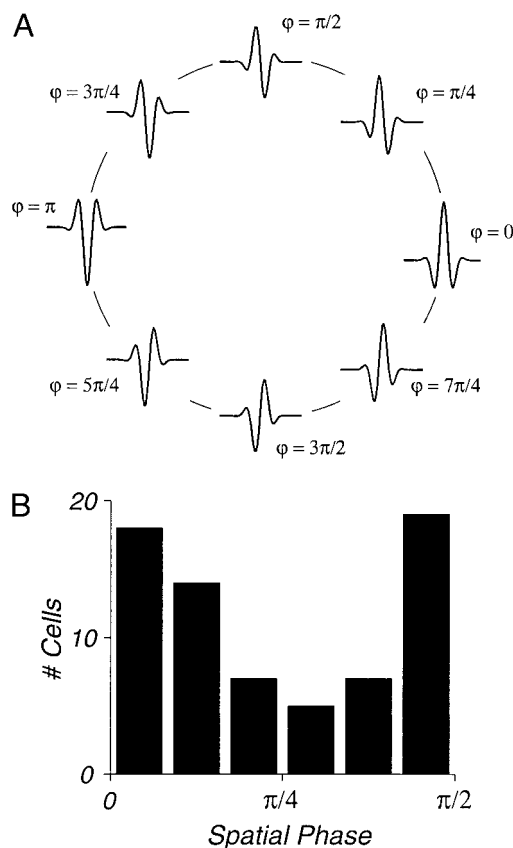

FIG. 6. *A*: summary of symmetry relationships and the spatial phase of a Gabor function. *B*: distribution of ϕ in macaque V1.

In a minority of neurons, usually selective for the direction of motion of the stimulus, the spatial phase of the Gabor fit was observed to change in an approximate linear fashion over the time course of the response (DeAngelis et al. 1993a; Reid et al. 1987). To avoid the complication of defining the spatial phase of the RFs in these neurons, such data were not included in our data set. Only RFs where the spatial phase of the Gabor fits remained stable within 20° during the central 15 ms of the response were included in the analysis. This criterion was preferred over previous methods such as estimating the degree of inseparability via SVD or a by fitting a quadrature model (DeAngelis et al. 1999) and discarding cells with high direction selectivity indices (DeAngelis et al. 1993a). The former method was not employed because in a number of cases the kernels had a well-defined symmetry but were inseparable in space and time. The RF of an LGN neuron, where the surround is delayed in time with respect to the center, is an example of a spatiotemporal inseparable RF that is even symmetric at any point in time. The direction selectivity index was not used because, in the primate, we have observed simple cells that appear to be spatiotemporal separable but are selective to the direction of motion (M. Hawken, unpublished observations).

Figure 6*B* shows the distribution of the spatial phase variable, $\hat{\phi}$, in our population. In contrast to previous findings (DeAngelis et al. 1993a; Field and Tolhurst 1986; Hamilton et al. 1989; Jones and Palmer 1987a), the macaque data are significantly bimodal [$P < 0.02$, Hartigan's dip test, (Hartigan and Hartigan 1985)], with cells clustering near 0 (even symmetry) and $\pi/2$ (odd symmetry). In addition, there is a tendency for cells that are well tuned for orientation and spatial frequency [located away from the origin of the (n_x, n_y) plane] to be odd symmetric, and a tendency for blob-like cells (located near the origin) to be even symmetric. This can be seen by first calculating the distance of each point in Fig. 4 from the origin, *d*. Figure 7, shows the histograms of $\hat{\phi}$ for two groups of cells: a group of neurons with distances less than the median *d* (broadly tuned) and the group of neurons with distances larger than the median of *d* (well tuned). The tendency for even symmetry in cells near the origin is explained by the fact that

many of these RFs are blob-like, which necessarily implies even symmetry (Jones and Palmer 1987a).

The estimates of $\hat{\phi}$ were reasonably accurate. For each cell, we calculated the confidence interval of the estimated value $\hat{\phi}$. The size of the confidence interval represents a measure of how well the spatial phase can be determined given the noise in the data. With the amount of data collected, the first-quartile, median, and third-quartile of the distribution for the size of the confidence interval in our data set were 0.08 rad (4.58°), 0.16 rad (9.17°), and 0.28 rad (16.04°), respectively.

A natural question is why the cortex evolved one particular family of spatial filters (Fig. 4) as well as observed distribution of spatial phases in V1 (Figs. 6 and 7). Recently, a number of related theories have been put forward to explain the shape of simple cell RFs in V1 (Bell and Sejnowski 1997; Olshausen and Field 1996, 1997; van Hateren and Ruderman 1998; van Hateren and van der Schaaf 1998); for a review, see Simoncelli and Olshausen 2001). I refer the reader to this literature for a detailed discussion of the concepts, mathematics, and implementation issues. Briefly, the basic idea behind these theories is that the cortex evolved RFs needed to "represent" natural images in an "efficient" manner. To make a theory explicit, one has to define what is meant by "representation" and in what way the representations are meant to be efficient. Most theories assume a linear representation of the image. If $I(x, y)$ is a two-dimensional image, one seeks a linear representation (or linear additive model) in terms of basis functions $\Phi_1(x, y)$

$$I(x, y) = \sum_i \alpha_i \Phi_i(x, y) + \varepsilon(x, y) \qquad (3)$$

Here, α_i are the coefficients corresponding to each basis function, and $\epsilon(x, y)$ is the error in the representation. The theories state that during evolution or development the basis set $\{\Phi(x, y)\}$ is optimized so that the coefficients α_i have some specific properties when the classes of images represented are drawn from the natural environment. Independent component analysis defines an efficient representation as one in which the coefficients are statistically *independent* one from the other. Normally the number of basis functions equals the dimension of the input and the error in the representation is zero. Sparse coding defines an efficient representation as one in which the coefficients are statistically *independent* and *sparse*. Sparseness is defined as a statistic on the distribution of a coefficient. There are various ways of defining sparseness (Olshausen and Field 1996, 1997; Willmore et al. 2000), but the basic principle is that the measure should give large values when the distribution of the coefficient has long tails and a large mass concentration around zero. The number of basis functions in sparse coding is normally (but not necessarily) larger than the dimension of the input. In this case, the representation is called *overcomplete* as there are many different ways (different sets of coefficients) of representing the image with some error. SC posits that from all these alternatives the brain computes the set of coefficients that is maximally sparse. Algorithms were developed that, when presented with natural images, will calculate an optimal set of basis functions according to the preceding criteria.

It should be emphasized that RFs predicted by these theories are *not* the basis functions (Φ_i) themselves. Both ICA and SC postulate that the goal of V1 is to compute the coefficients α_i.
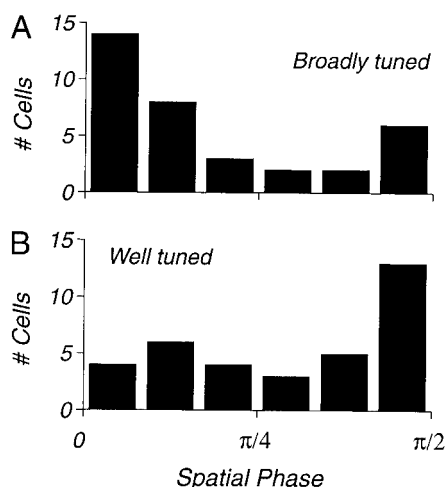

FIG. 7. *A*: distribution of $\hat{\phi}$ for neurons near the origin, $d <$ median (*d*). *B*: distribution of $\hat{\phi}$ for neurons away from the origin, $d \geq$ median (*d*). The median value of *d* was 0.43. There is a tendency for cells near the origin to have even-symmetric receptive fields and for cells away from the origin to be odd symmetric.

Because the basis functions are not necessarily orthogonal one to the other, the RFs (or the filters) required to compute the coefficients are not equal to the basis functions. In ICA, the set of filters required to calculate the coefficients is obtained simply by inverting the basis function matrix $\Phi = [\Phi_1 \Phi_2 \ldots \Phi_n]$ (Bell and Sejnowski 1997; van Hateren and van der Schaaf 1998). Thus the shapes of ICA *filters* and the V1 receptive fields can be compared directly. For convenience, Fig. 8A shows a few examples of the resulting ICA filter shapes. The shape of receptive fields predicted by SC can be obtained by simulating an "reverse correlation" mapping experiment (Olshausen 2001; Olshausen and Field 1997). In the APPENDIX, a fast method for estimating these receptive fields without the need of lengthy mapping simulations is suggested. However, as the data are already available in this case, the original receptive field maps as reported in Olshausen (2001) were analyzed. Figure 8B provides a few examples of the receptive fields resulting from reverse-correlation mapping of the SC nonlinear network.

As with the experimental data, both ICA and SC RFs can be approximately described by two-dimensional Gabor functions. The distribution of (n_x, n_y) obtained from these theories are shown in Fig. 9. Only RFs that were well within the boundaries of the image patch were included in this analysis. To facilitate a comparison, the experimental data are re-plotted here as well.

It can be seen that ICA generates filters that tend to be quite

FIG. 9. Comparison between the receptive field shapes predicted by ICA (□), SC (△), and the experimental data (○). Both ICA and SC generate RFs that have larger number of subfields than the experimental data. Neither theory predicts a significant number of receptive fields near the origin.

similar one to the other up to translation, rotation and scaling (Fig. 9, □). This is inferred from the fact that all RFs appear to map to a very similar location in the (n_x, n_y) plane. The ICA data points are shifted to the right the experimental data. This means that ICA filters have more subfields than actually observed. The distributions of filter shapes predicted by SC are similar and depicted by △. While these data exhibit a larger scatter, the vast majority of the points are clearly to the right of the experimental data as well. In addition, note that neither theory generates a significant population of units with RFs that are broadly tuned for orientation. The experimental data, on the other hand, show plenty of such RFs in all layers of macaque V1 (Hawken et al. 2000). These differences between the predictions of ICA/SC and the data might also be apparent by visual inspection of the kernels in Fig. 8 with those in Figs. 2B and 4. Finally, Fig. 10 shows the distributions of spatial phases predicted by SC. Interestingly, there is a tendency for RFs to be odd symmetric, similar to what is observed in the group of well-tuned cells in Fig. 6B.
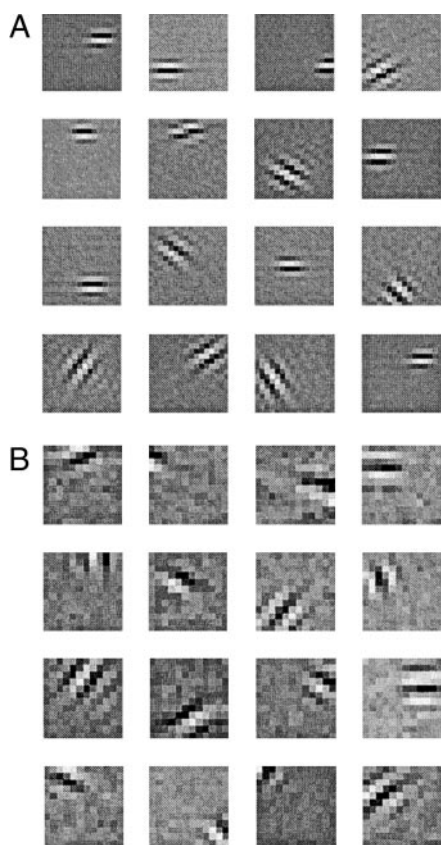

FIG. 8. A: examples of independent component analysis (ICA) filters obtained with dimensionality reduction and log intensity transformation (van Hateren and Ruderman 1998; van Hateren and van der Schaaf 1998) These data are available from: http://hlab.phys.rug.nl/demos/ica/comp_filt.html. B: examples of sparse coding (SC) receptive fields obtained by simulating a reverse-correlation mapping experiment (Olshausen 2001).
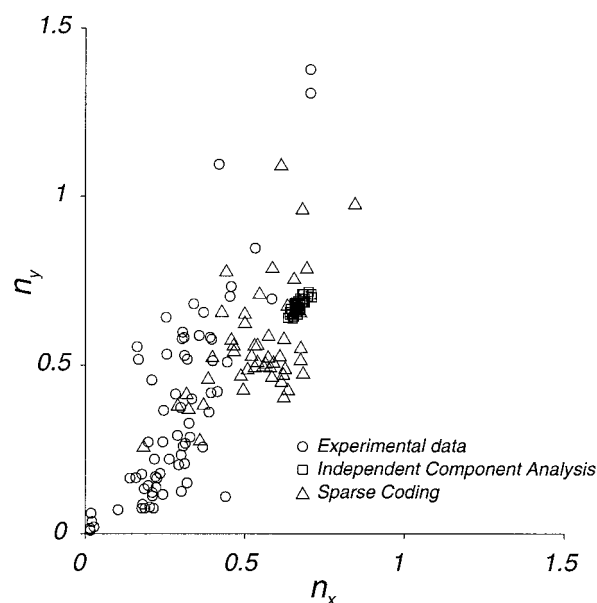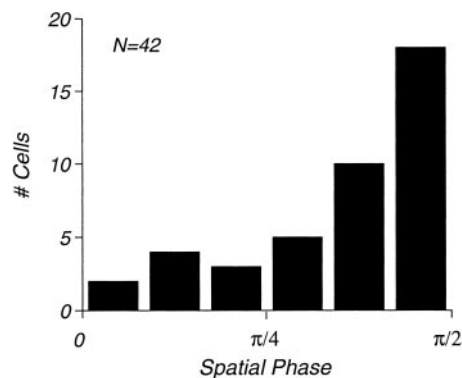

FIG. 10. Distribution of spatial phases of SC. There is a tendency toward odd-symmetric receptive field profiles.

DISCUSSION

The two-dimensional shape of simple-cell RFs at their optimal delay time was measured using subspace reverse correlation. The estimated filters are well approximated by two-dimensional Gabor functions as has been previously reported in cat area 17 (Jones and Palmer 1987a). Furthermore, the distribution of the filter shapes in monkey and cat are comparable (Fig. 5). One difference is that the cat data appear to be shifted slightly to the left with respect to the macaque data, which would suggest smaller number of effective subfields. It should be noted, however, the study of Jones and Palmer (1987a) used small spots of light as stimuli, while the present study employed extended grating stimuli. It is could possible for these data to be reconciled if one assumes that localized stimuli tend to underestimate the size of the envelope due to thresholding effects. A more detailed comparison between simple-cell RFs in cat and monkey will require the use of identical methods in both species.

The distribution of spatial phases in macaque V1 is bimodal: RFs cluster into even-and odd-symmetric classes. There is also a tendency for RFs that are well tuned in orientation and spatial frequency to be odd symmetric. This finding was unexpected as all previous studies in cat area 17 report a uniform distribution of spatial phases (DeAngelis et al. 1993a; Field and Tolhurst 1986; Hamilton et al. 1989; Jones and Palmer 1987b). I verified that neither the measurement technique nor the Gabor fitting procedure exhibit a bias toward even/odd symmetry. This was done by simulating RFs as a with random spatial phases followed by rectification with various threshold levels, estimating the RF via reverse-correlation (Ringach et al. 1997b), fitting them with two-dimensional Gabor functions, and generating a scatter-plot of the simulated versus estimated spatial phase. No biases were observed (data not shown). There could be other reasons for the discrepancy between cat and monkey data. First, earlier studies in cat, except for the work of DeAngelis and co-workers, did not take into account the existence of spatiotemporal inseparable RFs. This might have added noise to the data and wash away possible effects. Second, all previous reports, except for the work of Jones and Palmer, studied the phase of a 1D projection of the RF (or the line-weighting function). It can be checked via numerical simulations that fitting a two-dimensional Gabor function to a two-dimensional estimate of the RF provides more precise estimates of the spatial phase than projecting the RF on one axis and then fitting a one-dimensional Gabor function, as done by DeAngelis et al. (1993a). This is particularly true if one also allows for some error in the projection axis of the RF. Thus estimates of the spatial phase from line-weighting functions (or a 1D analysis) are inherently noisier than those from a two-dimensional analysis. Third, it should be clear that in measuring the receptive-field spatial phase, it is very important to have the eyes steady in their orbit as changes in eye position may contaminate the measurements. Obviously, this is more important in monkey than in cat as the RFs in the macaque are much smaller. Nevertheless, it is unclear to what extent previous studies were concerned about eye movements and how they verified that the eye was stable during the measurements. Fourth, we must also consider the possibility that cat and monkey are different in this respect.

An important step in understanding the function of simple cortical cells will be to explain the observed distribution of filter shapes (Fig. 4) and spatial phases (Figs. 6B and 7). The predictions of two recent theories, independent component analysis and sparse coding, do not match the data (Fig. 9). One similarity is that the SC RFs tend to be odd symmetric as seen in the population of well-tuned neurons (Fig. 7 and 10). The odd symmetry of the SC RFs might not be surprising as taking differences of nearby locations in space is clearly one way of generating a sparse distribution of responses (Ruderman 1994).

There are many possible reasons behind the failure of ICA and SC in explaining the distribution of filter shapes in V1. One possibility is that the function of simple cells is not to generate a full representation of the image, as suggested by these theories. It could also be that mean square error is not the measure being optimized. It is well known that the mean square error is not an appropriate psychophysical measure of image similarity or fidelity (see Heeger and Teo 1995 and references therein). One wonders if minimization of other measures (such as the L_1 norm of the error) would produce RFs that match the experimental data better. Another caveat about these comparisons is that the RF predictions of these theories are dependent on the preprocessing of the images used to train the algorithms, the statistics of the images selected for training, the actual algorithm used to optimize the basis set (van Hateren and Ruderman 1998), and whether the analysis is done on static images or on image sequences (van Hateren and van der Schaaf 1998). The dependence of the shapes of the predicted RFs on these implementation details has not been explored thoroughly yet. Thus it appears premature to argue based on the present comparisons that the basic principles put forward by these theories are unsuitable for explaining simple-cell function in V1. It might be possible that future realizations of these ideas will be able to account for the experimental data. Our dataset provides an appropriate benchmark against which theories of simple-cell RF could be tested. To make this possible, the data in Figs. 4, 6, and 7 are made available at http://manuelita.psych.ucla.edu/~dario.

APPENDIX

Here, I show that the RF maps of the sparse coding network obtained via reverse correlation can be approximated by the expression $\Psi \approx (\Phi^T \Phi + \beta 1)^{-1} \Phi^T$. The reason for this simplification is that the type of stimulus used in reverse-correlation experiments effectively "linearizes" the network around its operating point. A perturbation analysis around the operating point results in the preceding relationship.

In what follows, the following notation is used. The column vector $\mathbf{I}$ (of size $M \times 1$) represents the image, where $M = n^2$ is the size of the image. The matrix Φ ($M \times N$) encodes in each column one of the N basis functions elements. The column vector α (of size $N \times 1$) represents the coefficients of the representation. Given the image, the sparse coding network finds the optimal coefficients by relaxation of the nonlinear evolution equation (Olshausen and Field 1997)

$$\dot{\alpha} = \Phi^T(\mathbf{I} - \Phi\alpha) - \lambda S'(\alpha) \qquad (A1)$$

Let us assume that the stimulus consists of a small perturbation around a mean, $\mathbf{I} = \bar{\mathbf{I}} + \delta\mathbf{I}$, and that in response to this stimulus the
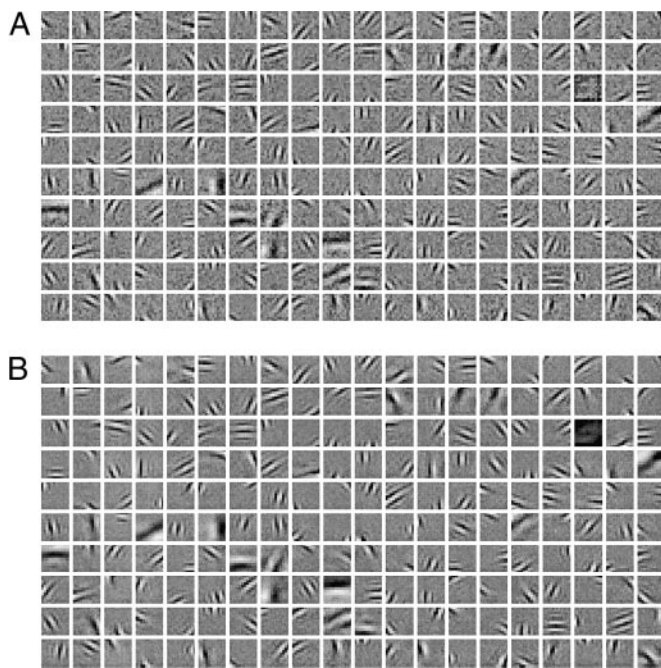
FIG. 11. A linear approximation to the receptive field mapping of the sparse network. *A*: receptive fields obtained by simulating a receptive-field mapping experiment on the nonlinear sparse network (Olshausen 2001). *B*: the approximate receptive fields obtained by the linear approximation technique.

coefficients are perturbed around a mean, $\boldsymbol{\alpha} = \bar{\boldsymbol{\alpha}} + \delta\boldsymbol{\alpha}$. The coordinate $(\bar{\mathbf{I}}, \bar{\boldsymbol{\alpha}})$ is the "operating point" of the network. Linearization of the evolution equation around the operating point yields

$$\frac{d\delta\boldsymbol{\alpha}}{dt} = \boldsymbol{\Phi}^T(\delta\mathbf{I} - \boldsymbol{\Phi}\delta\boldsymbol{\alpha}) - \lambda \, \text{diag} \, (S''(\bar{\boldsymbol{\alpha}}))\delta\boldsymbol{\alpha} \quad (A2)$$

Here, I used $S'(\bar{\boldsymbol{\alpha}} + \delta\boldsymbol{\alpha}) \approx S'(\bar{\boldsymbol{\alpha}}) + \text{diag}[S''(\bar{\boldsymbol{\alpha}})]\delta\boldsymbol{\alpha}$. In steady state, we have $d\delta\boldsymbol{\alpha}/dt = 0$ and one gets

$$\delta\boldsymbol{\alpha} = [\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \lambda \, \text{diag} \, (S''(\bar{\boldsymbol{\alpha}}))]^{-1}\boldsymbol{\Phi}^T\delta\mathbf{I} \quad (A3)$$

If the stimulus is symmetric and centered around zero (such as Gaussian white noise), it is easy to see that the coefficients will also be centered around zero; i.e., $\bar{\boldsymbol{\alpha}} = \mathbf{0}$. In this case

$$\delta\boldsymbol{\alpha} = [\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \beta\mathbf{1}]^{-1}\boldsymbol{\Phi}^T\delta\mathbf{I} \quad (A4)$$

where $\beta = \lambda S''(0)$ and the matrix $\mathbf{1}$ represents the identity matrix. For any reasonable definition of sparseness, we have $S''(0) > 0$, as $S(x)$ is convex at the origin. Thus there is no change of sign, and we can define β as a positive number. Because *Eq. A4* is a linear system ($\delta\boldsymbol{\alpha}$ and $\delta\mathbf{I}$ are linearly related), cross-correlation between the input and the output will recover the constant of proportionality, which equals

$$\boldsymbol{\Psi} = [\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \beta\mathbf{1}]^{-1}\boldsymbol{\Phi}^T \quad (A5)$$

To verify this approximation, the set of RFs obtained via reverse-correlation with small spots reported by Olshausen (2001) (Fig. 11*A*) and the set of RFs obtained by *Eq. A5* (Fig. 11*B*) were compared. The approximation is quite reasonable: the overall correlation between the data sets is 0.89 for a value of $\beta = 0.1$.

REFERENCES

BELL AJ AND SEJNOWSKI TJ. The "independent components" of natural scenes are edge filters. *Vision Res* 37: 3327–3338, 1997.

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.

CLEVELAND WS AND DEVLIN SJ. Locally-weighted regression: an approach to regression analysis by local fitting. *J Am Statist Assoc* 83: 596–610, 1988.

DEANGELIS GC, GHOSE GM, OHZAWA I, AND FREEMAN RD. Functional micro-organization of primary visual cortex: receptive field analysis of nearby neurons. *J Neurosci* 19: 4046–4064, 1999.

DEANGELIS GC, OHZAWA I, AND FREEMAN RD. Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development. *J Neurophysiol* 69: 1091–1117, 1993a.

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, 1993b.

DE VALOIS KK, DE VALOIS RL, AND YUND EW. Responses of striate cortex cells to grating and checkerboard patterns. *J Physiol (Lond)* 291: 483–505, 1979.

DE VALOIS RL, ALBRECHT DG, AND THORELL LG. Spatial frequency selectivity of cells in macaque visual cortex. *Vision Res* 22: 545–559, 1982.

DE VALOIS RL, COTTARIS NP, MAHON LE, ELFAR SD, AND WILSON JA. Spatial and temporal receptive fields of geniculate and cortical cells and directional selectivity. *Vision Res* 40: 3685–3702, 2000.

FIELD DJ AND TOLHURST DJ. The structure and symmetry of simple-cell receptive-field profiles in the cat's visual cortex. *Proc R Soc Lond B Biol Sci* 228: 379–400, 1986.

HAMILTON DB, ALBRECHT DG, AND GEISLER WS. Visual cortical receptive fields in monkey and cat: spatial and temporal phase transfer function. *Vision Res* 29: 1285–1308, 1989.

HARTIGAN JA AND HARTIGAN PM. The dip test of unimodality. *Ann Stat* 13: 70–84, 1985.

HAWKEN MJ AND PARKER AJ. Spatial properties of neurons in the monkey striate cortex. *Proc R Soc Lond B Biol Sci* 231: 251–288, 1987.

HAWKEN MJ, SHAPLEY R, RINGACH DL, SCENIAK MP, JOHNSON EN, AND MARESCHAL I. Laminar distribution of orientation selectivity of macaque V1 neurons. *Invest Ophthal Vis Sci* 41: 276B, 2000.

HEEGER DJ AND TEO PC. A model of perceptual image fidelity. *Int Conf Image Process* 2: 95, 1995.

JONES JP AND PALMER LA. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J Neurophysiol* 58: 1233–1258, 1987a.

JONES JP AND PALMER LA. The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *J Neurophysiol* 58: 1187–1211, 1987b.

KULIKOWSKI JJ AND BISHOP PO. Fourier analysis and spatial representation in the visual cortex. *Experientia* 37: 160–163, 1981.

MARCELJA S. Mathematical description of the responses of simple cortical cells. *J Opt Soc Am* 70: 1297–1300, 1980.

MERRILL EG AND AINSWORTH A. Glass-coated platinum-plated tungsten microelectrodes. *Med Biol Eng* 10: 662–672, 1972.

MOVSHON JA, THOMPSON ID, AND TOLHURST DJ. Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *J Physiol (Lond)* 283: 53–77, 1978.

OLSHAUSEN BA. *Sparse Codes and Spikes. Probabilistic Models of Perception and Brain Function.* Cambridge, MA: MIT Press, 2001.

OLSHAUSEN BA AND FIELD DJ. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381: 607–609, 1996.

OLSHAUSEN BA AND FIELD DJ. Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vision Res* 37: 3311–3325, 1997.

PARKER AJ AND HAWKEN MJ. Two-dimensional spatial structure of receptive fields in monkey striate cortex. *J Opt Soc Am A* 5: 598–605, 1988.

REID RC, SOODAK RE, AND SHAPLEY RM. Linear mechanisms of directional selectivity in simple cells of cat striate cortex. *Proc Natl Acad Sci USA* 84: 8740–9744, 1987.

RINGACH DL, HAWKEN MJ, AND SHAPLEY R. Dynamics of orientation tuning in macaque primary visual cortex. *Nature* 387: 281–284, 1997a.

RINGACH DL, SAPIRO G, AND SHAPLEY R. A subpsace reverse correlation method for the study of visual neurons. *Vision Res* 37: 2455–2464, 1997b.

RUDERMAN DL. The statistics of natural images. *Network Comput Neural Syst* 5: 517–548, 1994.

SIMONCELLI EP AND OLSHAUSEN BA. Natural image statistics and neural representation. *Annu Rev Neurosci* 24: 1193–1216, 2001.

SKOTTUN BC, DE VALOIS RL, GROSOF DH, MOVSHON JA, ALBRECHT DG, AND BONDS AB. Classifying simple and complex cells on the basis of response modulation. *Vision Res* 31: 1079–1086, 1991.

VAN HATEREN JH AND RUDERMAN DL. Independent component analysis of natural image sequences yields spatiotemporal filters similar to simple cells in primary visual cortex. *Proc R Soc Lond B Biol Sci* 265: 2315–2320, 1998.

VAN HATEREN JH AND VAN DER SCHAAF A. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc R Soc Lond B Biol Sci* 265: 359–366, 1998.

WILLMORE B, WATTERS PA, AND TOLHURST DJ. A comparison of natural-image-based models of simple-cell coding. *Perception* 29: 1017–1040, 2000.