Behavioral/Systems/Cognitive

Natural Stimulus Statistics Alter the Receptive Field Structure of V1 Neurons

Stephen V. David,[1] William E. Vinje,[2,4] and Jack L. Gallant[3,4]

[1]Program in Bioengineering, Departments of [2]Molecular and Cellular Biology and [3]Psychology, and [4]Program in Neuroscience, University of California Berkeley, Berkeley, California 94720-1650

Studies of the primary visual cortex (V1) have produced models that account for neuronal responses to synthetic stimuli such as sinusoidal gratings. Little is known about how these models generalize to activity during natural vision. We recorded neural responses in area V1 of awake macaques to a stimulus with natural spatiotemporal statistics and to a dynamic grating sequence stimulus. We fit nonlinear receptive field models using each of these data sets and compared how well they predicted time-varying responses to a novel natural visual stimulus. On average, the model fit using the natural stimulus predicted natural visual responses more than twice as accurately as the model fit to the synthetic stimulus. The natural vision model produced better predictions in >75% of the neurons studied. This large difference in predictive power suggests that natural spatiotemporal stimulus statistics activate nonlinear response properties in a different manner than the grating stimulus. To characterize this modulation, we compared the temporal and spatial response properties of the model fits. During natural stimulation, temporal responses often showed a stronger late inhibitory component, indicating an effect of nonlinear temporal summation during natural vision. In addition, spatial tuning underwent complex shifts, primarily in the inhibitory, rather than excitatory, elements of the response profile. These differences in late and spatially tuned inhibition accounted fully for the difference in predictive power between the two models. Both the spatial and temporal statistics of the natural stimulus contributed to the modulatory effects.

Key words: **striate cortex; natural vision; receptive field; nonlinear model; reverse correlation; inhibition**

Introduction

Many computational theories have been proposed to explain the functional properties of neurons in the primary visual cortex (V1). Models based on these theories account for properties ranging from basic linearity (Hubel and Wiesel, 1959; Movshon et al., 1978; Jones et al., 1987) to nonlinear properties such as phase invariance in complex cells (Pollen and Ronner, 1983; Adelson and Bergen, 1985) and nonlinear temporal summation (Tolhurst et al., 1980). Neurophysiological data supporting a given model generally consist of responses to stimuli synthesized specifically to modulate the property of interest. These stimuli differ qualitatively from natural visual stimuli, and it is not known how models developed with synthetic stimuli generalize to natural vision.

Natural scenes have complex high-order statistics that reflect the structure of the visual world. Phenomena such as the three-dimensional geometry of objects, the projection of the world onto the retina, and the dynamics of eye movements give rise to complex but systematic spatiotemporal patterns in visual inputs (Field, 1987; Zetzsche et al., 1993; Woods et al., 2001). It has been proposed that the visual system takes advantage of these regularities to recognize and react to the natural environment (Barlow, 1961; Olshausen and Field, 1997). The fact that visual neurons respond in a nonlinear manner suggests that natural stimuli could evoke substantially different responses from what would be predicted by current models estimated using synthetic stimuli. Understanding the behavior of neurons during natural stimulation is critical for developing accurate models of information transmission and coding, yet few studies have investigated this issue directly, particularly in the visual cortex (Vinje and Gallant, 1998; Ringach et al., 2002; Smyth et al., 2003; Weliky et al., 2003).

In this study, we asked two questions related to this problem. First, in the framework of current models, do natural spatiotemporal statistics affect response properties in primate V1? To address this question, we recorded the responses of neurons to two classes of stimuli: one that closely approximated natural visual stimulation in primates and the other composed of synthetic sinusoidal gratings. We used the response data to estimate spatiotemporal receptive fields (STRFs) for each stimulus class (DeBoer and Kuyper, 1968; Theunissen et al., 2001). To study both simple and complex cells in the same framework, we developed a nonlinear STRF model that accounted for response properties of both types of neurons (David et al., 1999). We compared STRFs estimated using the two stimulus classes according to how well

Received March 5, 2004; revised June 2, 2004; accepted June 3, 2004.

This work was supported by grants from the Sloan and Whitehall Foundations and from the National Eye Institute and National Institute of Mental Health (J.L.G.). S.V.D. was supported by a National Science Foundation fellowship. We thank Frederic Theunissen for valuable input regarding kernel estimation and for the original inspiration to conduct reverse correlation in the frequency domain; Garrett Stanley for additional help regarding reverse correlation on natural stimuli; James Mazer for assistance in the development of these analyses and on data acquisition; and Kate Gustavsen, Kathleen Hansen, Benjamin Hayden, and Ben Willmore for helpful comments on this manuscript.

Correspondence should be addressed to Jack L. Gallant, University of California Berkeley, 3210 Tolman Hall, 1650, Berkeley, CA 94720-1650. E-mail: gallant@socrates.berkeley.edu.

DOI:10.1523/JNEUROSCI.1422-04.2004

 0270-6474/04/246991-16$15.00/0

they predicted natural visual responses (Theunissen et al., 2000). If one STRF predicted responses more accurately than the other, then we could infer that it provided a more accurate description of V1 response properties during natural vision (Gallant, 2003). In more than half the neurons we studied, STRFs estimated using the natural stimulus predicted natural visual responses significantly better than STRFs estimated using the synthetic stimulus.

Second, we asked what aspects of STRFs estimated using the natural stimulus contribute to their improved predictions. We compared the spatial and temporal structure of STRFs estimated using each stimulus class. Natural stimulation increased late temporal inhibition and induced complex shifts in inhibitory spatial tuning. After controlling for these changes, differences in predictive power were eliminated. This suggests that nonlinear modulation of inhibition is the primary source of differences between stimulus conditions. Models that account explicitly for this modulation should show improved performance for both stimulus classes. They may also be useful for deriving synthetic stimuli that drive neurons in visual cortex in the same manner as natural stimuli.

Materials and Methods

Data collection

We recorded spiking activity from 74 well isolated neurons in parafoveal area V1 of two awake, behaving male macaques (*Macaca mulatta*). Extracellular activity was recorded using tungsten electrodes (FHC, Bowdoinham, ME) and amplified (AM Systems, Everett, WA); a custom hardware window discriminator was used to identify action potentials (temporal resolution, 8 kHz). During recording, the animals performed a fixation task for a liquid reward. Eye position was monitored with a scleral search coil, and trials were aborted if eye position deviated >0.35° from fixation. All procedures were performed under a protocol approved by the Animal Care and Use Committee at the University of California and conformed to National Institutes of Health standards. Surgical procedures were conducted under appropriate anesthesia using standard sterile techniques (Vinje and Gallant, 2002).

After isolating a neuron, we estimated its receptive field size and location manually using bars and gratings. For 32 neurons, we used an automatic procedure that confirmed these estimates by reverse correlation of responses to a dynamic sparse noise stimulus consisting of black and white squares positioned randomly on a gray background (Jones et al., 1987; DeAngelis et al., 1993; Vinje and Gallant, 2002). Squares were scaled so that six to eight squares spanned the manually estimated receptive field (0.1–0.5°/square). The classical receptive field (CRF) was designated as the circle around the region where sparse noise stimulation elicited spiking responses. Our manual and automatic estimation procedures were generally in good agreement.

Stimuli

We used three types of stimuli to probe neural responses (Fig. 1): natural vision movies, grating sequences, and natural image sequences. All stimuli had the same mean luminance and root mean squared (RMS) contrast. They were presented on a cathode ray tube (CRT) display with a gray background matched to the mean stimulus luminance. Stimuli were centered on the receptive field of the neuron while the animal fixated. Each stimulus sequence was divided into several 5 sec segments. Different segments were presented on successive fixation trials in random order; trials from different stimulus classes were not interleaved. To avoid transient trial onset effects, the first 196 msec of data acquired on each trial was discarded before analysis.

Natural vision movies. Natural vision movies mimicked the stimulation occurring in and around the CRF during free inspection of a natural scene with voluntary eye movements. A Monte Carlo model of eye movements was used to extract an appropriate series of image patches from a natural scene (Vinje and Gallant, 2000, 2002). Fixation durations were chosen randomly from a Gaussian distribution (mean, 350 msec; SD, 50 msec). Saccade directions were selected randomly from a uniform distribution. Saccade velocities and lengths were chosen randomly from a B-spline fit to the distribution of eye movements recorded during free viewing of stationary photographs of natural scenes. Natural scenes were 1280×1024 pixel images obtained from a high-resolution commercial photo compact disk library (Corel). Images included landscapes, man-made objects, animals, and humans. Color images were converted to gray scale before extraction. These images were not calibrated to match natural luminance and contrast levels exactly, but they did contain the higher-order spatial structure of natural scenes. Circular image patches were extracted from the scene along the simulated scan path, clipped to two to four times the CRF diameter, and blended into the display background to avoid edge effects (outer 10% of the patch radius). To reduce temporal aliasing artifacts that might result from using a display with a 72 Hz refresh rate, each 14 msec frame was constructed by averaging 14 images representing the position of the CRF at intervals of 1 msec. A segment taken from a natural vision movie showing each 14 msec frame during the transition between two simulated fixations appears in Figure 1*A*. The frames appearing in a series of simulated fixations appear in Figure 1*B*, along with the response recorded from a single neuron averaged over 10 repeated trials.

Figure 1*E* shows temporal and spatial statistics of a natural vision movie. The temporal autocorrelation function (left) decreases at a constant rate until it reaches a value near zero at time lags of ~500 msec. This pattern is attributable to the low temporal frequency bias created by the dynamics of the simulated saccades used to construct the movies. Log spatial power (Fig. 1*E*, right) was averaged over the entire natural vision movie and plotted in the phase-separated Fourier domain (for details on the phase-separated Fourier transformation, see below). Each subpanel shows the two-dimensional power spectrum of the natural vision movie at a different spatial phase; brighter pixels indicate greater power. Power falls off linearly from low frequencies at the center of each subpanel, reflecting the $1/f^2$ power spectrum typically observed in natural images (Field, 1987). Horizontal and vertical orientations have slightly higher power because of their predominance in natural scenes. The empty points at the center of all but the top right subpanel reflect the fact that mean (DC) luminance is always positive and real in the images, so that its power lies entirely in the $\phi = 0$ phase channel in the Fourier domain. Attenuation at low frequencies in the top left ($\phi = 180$) subpanel is an effect of the Hanning window applied before the Fourier transform to reduce edge artifacts.

For each neuron, the total duration of natural vision movies ranged from 10 to 200 sec. Two different procedures were used to select a segment for display on each trial. For 17 neurons, each trial contained a unique segment of a natural vision movie, and no segment was repeated. In these cases STRFs were estimated from the peristimulus time histogram (PSTH) approximated from a single stimulus presentation. For the remaining 57 neurons, fewer natural vision movie segments were used, and each segment was repeated 10–30 times. In these cases, reverse correlation was performed on the PSTH obtained by averaging responses over repeated presentations of each movie segment. For all neurons, an additional data set was acquired in which a 5–10 sec natural vision movie was repeated 20–40 times. These data were not used in the reverse correlation analysis but were reserved for evaluating responses predicted by the STRFs (see below).

Grating sequences. Grating sequences consisted of a series of sinusoidal gratings that varied randomly in orientation, spatial frequency, and spatial phase (Ringach et al., 1997; Mazer et al., 2002). Orientation was sampled uniformly from 0 to 180°, spatial frequency from 0.5 to 6 cycles per CRF diameter, and phase from 0 to 360°. All gratings were two to four times the CRF diameter, and their outer edges (10% of the radius) were blended into the gray background of the display. Mean luminance and RMS contrast were normalized to match natural vision movies. For 33 neurons, stimuli were updated on each video refresh (72 Hz). For an additional 15 neurons, stimuli were shown at 24 Hz. The total duration of grating sequence stimulation ranged from 50 to 150 sec, depending on the neuron. A brief segment taken from a grating sequence appears in Figure 1*C*, and a typical response appears in Figure 1*D*.

Figure 1*F* shows the temporal and spatial statistics of a grating sequence, plotted using the conventions of Figure 1*E*. Because grating

parameters are varied randomly in each 14 msec frame, the temporal autocorrelation is zero for all nonzero time lags. The log spatial power spectrum is nearly flat, with a slight bias toward low spatial frequencies. This reflects the fact that gratings were sampled uniformly in orientation and spatial frequency, and fewer orientation bins exist at low spatial frequencies in the Fourier domain. Despite the slight bias, the grating sequence has less power at low frequencies and more power at high frequencies than the natural vision movie. Total power (integrated over the entire spatial power spectrum) was the same for both stimulus classes.

Natural image sequences. Natural image sequences were used as a control stimulus to dissociate the effects of natural spatial and temporal statistics on response properties. They were constructed with the spatial statistics of natural vision movies and temporal statistics of grating sequences. Each 14 msec frame of the natural image sequence contained a random image patch taken from the same library of images used to generate natural vision movies. All patches were two to four times larger than the CRF, and their outer edges (10% of the radius) were blended linearly into the gray background of the display. Natural image sequences were updated on each video refresh (72 Hz). The total duration of natural image sequence stimulation ranged from 100 to 150 sec, and each sequence was shown only once. Natural image sequences were used to acquire data from 21 neurons. The temporal autocorrelation of natural image sequences is the same as for grating sequences (Fig. 1*F*), whereas their spatial power spectrum is the same as for natural vision movies (Fig. 1*E*).

Linearized spatiotemporal receptive field model

Sensory neurons using a rate code can be modeled in terms of a linear STRF (DeBoer and Kuyper, 1968; Marmarelis and Marmarelis, 1978; Theunissen et al., 2001). Given an arbitrary stimulus, $s(x_i,t)$, varying in space and time, the instantaneous firing rate response, $r(t)$, is:

$$r(t) = \left| \sum_{i=1}^{N} \sum_{u=0}^{U} h(x_i,u)s(x_i,t-u) - \theta \right|^{+} + \epsilon(t). \qquad (1)$$

The value of the linear filter, h, at each point in space, x_i, and time lag, u, describes how a stimulus at time $t - u$ influences the firing rate at time t. Time lags range from 0 to U, so this model assumes that the system is causal and has memory no longer than U. Positive values of h indicate excitatory stimulus channels that increase response for larger values of s, whereas negative values indicate inhibitory channels that decrease response. The spatial coordinates, $x_i \in \{x_1, x_2, \ldots x_N\}$, represent N discrete input channels. We modeled the spiking threshold observed in cortical neurons by half-wave rectification (Albrecht and Geisler, 1991). Rectification is represented by, $|X|^{+} = \max(0, X)$, with threshold specified by the scalar θ. (We tested other output nonlinearities, e.g., expansive nonlinearity and sigmoid, and found that STRF estimates and their predictive power were not substantially different. However, for most neurons, rectification did improve performance over models with no output nonlinearity.) The residual, $\epsilon(t)$, represents deviations from linear behavior attributable to either noise or unmodeled nonlinear response properties. The STRF model is shown schematically in Figure 2*A*.

Phase-separated Fourier domain. Simple cells in the primary visual cortex respond to stimuli having appropriate orientation, spatial frequency, and spatial phase (Hubel and Wiesel, 1959; Daugman, 1980). These neurons obey spatial superposition and so can be modeled as a linear transformation between the luminance at positions (x,y) in space and the mean firing rate, $r(t)$. According to this image domain model, the input channels are simply the luminance values at each point in space, $x_i = (x,y)$ (Jones et al., 1987).

Complex cells have tuning properties similar to those of simple cells, except that they are insensitive to spatial phase (Hubel and Wiesel, 1959; DeValois et al., 1982). These neurons violate spatial superposition because luminance at any point within the receptive field may be either excitatory or inhibitory, depending on the luminance at nearby locations. Because the image domain model requires consistent excitation or inhibition at each spatial position, it cannot be used to estimate STRFs for phase-invariant complex cells (DeAngelis et al., 1995; Theunissen et al., 2001; Touryan et al., 2002).

Complex cells have rarely been studied within the STRF framework because of the limitations of the image domain model. A few studies have estimated complex cell STRFs by removing all spatial phase information from the stimulus, thereby linearizing the stimulus–response relationship (Ringach et al., 1997; Mazer et al., 2002). A similar approach has been used in the temporal dimension for data in the auditory system (Aertsen and Johannesma, 1981; Theunissen et al., 2000; Machens et al., 2004). However, this procedure cannot recover simple cell STRFs because it discards the phase information required to determine which spatial channels are excitatory or inhibitory. We therefore developed a new linearizing procedure that can recover the STRF for both simple and complex cells (David et al., 1999). According to this model, each STRF is a linear filter in the phase-separated Fourier domain. This is accomplished by applying a spatial Fourier transform to each stimulus frame and projecting the resulting complex numbers onto the cardinal real and imaginary axes:

$$\begin{aligned} S_{PS}(\omega_x, \omega_y, 0, t) &= |\mathrm{Re}(S(\omega_x, \omega_y, t))|^{+} \\ S_{PS}(\omega_x, \omega_y, 90, t) &= |\mathrm{Im}(S(\omega_x, \omega_y, t))|^{+} \\ S_{PS}(\omega_x, \omega_y, 180, t) &= |\mathrm{Re}(S(\omega_x, \omega_y, t))|^{-} \\ S_{PS}(\omega_x, \omega_y, 270, t) &= |\mathrm{Im}(S(\omega_x, \omega_y, t))|^{-} . \end{aligned} \qquad (2)$$

Here $S(\omega_x,\omega_y,t)$ is the spatial Fourier transform of the stimulus. Spatial channels are defined over the three-dimensional space, $x_i = (\omega_x,\omega_y,\phi)$. This transformation preserves all of the information in the stimulus but creates an overcomplete representation. In the Fourier domain, spatial phase determines the relative magnitude of real and imaginary components at each spatial frequency. By projecting onto the complex axes, power at different spatial phases is assigned to different spatial channels. The excitatory and inhibitory influence of individual spatial frequency–phase channels is determined by the coefficient associated with each spatial channel in $h(x_i,u)$. In general, both complex and simple cells respond to a narrow range of orientations and spatial frequencies (DeValois et al., 1982) so that only a small number of spatial channels should show excitatory tuning in phase-separated Fourier STRFs. However, our procedure for STRF estimation does not place any constraints on which or how many channels will have either excitatory or inhibitory influence on neural responses.

According to the phase-separated Fourier model, a simple cell will have positive coefficients for one or two adjacent phase channels that excite responses, and negative coefficients at the opposite phase (offset by 180°) (see Fig. 2*B*). In contrast, a complex cell will have positive coefficients for all phases (Fig. 2*C*). Using this model, a neuron can be classified as simple or complex merely by examining the STRF. In practice, many neurons fall between strict classification as simple or complex, and this model can easily account for intermediate spatial phase tuning. Another advantage of the phase-separated Fourier model is that it can reveal spatially tuned inhibition. That is, negative coefficients in $h(x_i,u)$ indicate spatial channels that are correlated with a decrease in response. After the initial nonlinear transformation, the remaining stages of the phase-separated Fourier model are identical to the classical linear STRF model.

Space–time separable model. One simplification of the phase-separated Fourier model is to constrain STRFs to be space–time separable. A space–time separable STRF is the product of a spatial response function, $f(x_i)$, and a temporal response function, $g(u)$:

$$h(x_i, u) = f(x_i)g(u). \qquad (3)$$

The space–time separable model requires fewer parameters than the more general inseparable model and therefore can be estimated more accurately than the inseparable STRF when data are limited. A spatial response function estimated using the phase-separated Fourier model can capture simple and complex cell properties in the same way as an inseparable STRF (Fig. 2*D*,*E*).

There is evidence that some V1 neurons possess a significant space–time inseparable component (Mazer et al., 2002; Shapley et al., 2003). Because the space–time separable model may fail to capture some response properties of such neurons, we compared STRFs estimated using

the inseparable model (Eq. 1) with those estimated using the space–time separable model. Across the sample of 74 neurons, predictions were not significantly different (randomized paired t test). Some neurons had space–time inseparable components, but their contributions were outweighed by the improved signal-to-noise ratio of the separable model. We did not observe any systematic tuning differences other that those we report for the separable model. However, a more exhaustive study of the inseparable model with larger data sets would be required to confirm this. We chose to use the separable model because it lent itself well to quantitative analysis of spatial and temporal response properties.

Hybrid model. A space–time separable STRF can be decomposed easily into spatial and temporal response functions. We took advantage of this feature to construct a hybrid STRF that combined the spatial response function estimated using grating sequences with the temporal response function estimated using natural vision movies:

$$h_{\text{hybrid}}(x_i,u) = f_{\text{syn}}(x_i)\, g_{\text{nat}}(u). \quad (4)$$

The hybrid model allowed us to isolate the effects of natural stimulus statistics on spatial and temporal tuning properties.

For our analysis of natural image sequence data, we estimated two additional types of hybrid STRF, one that combined natural image sequence spatial response functions with temporal response functions estimated using grating sequences and another that combined natural image sequence spatial response functions with temporal response functions estimated using natural vision movies. These hybrid models were formed in a similar manner as Equation 4, but substituting spatial and temporal response functions estimated using the appropriate stimulus class.

Positive space model. The coefficients of a spatial response function can have positive or negative values indicating relative excitatory or inhibitory tuning, respectively. To examine the unique contributions of the excitatory aspects of tuning, we constructed a separable model with only positive coefficients in its spatial response function:

$$f^{+}(x_i) = |f(x_i)|^{+}. \quad (5)$$

This representation includes no negative coefficients in the spatial response function, effectively removing the spatially inhibitory influences. The positive space model still allows STRFs to have negative coefficients but only if the temporal response function contains negative coefficients. Thus, for any single time lag, all coefficients in h must be either excitatory or inhibitory.

STRF estimation theory

Reverse correlation algorithm. We used reverse correlation to estimate STRFs for each stimulus class independently. Reverse correlation experiments are typically based on responses to white noise. White noise has a flat power spectrum and no correlation between spatiotemporal channels (Marmarelis and Marmarelis, 1978; Jones et al., 1987). In this case, the minimum mean-squared error estimate of the STRF is simply the cross-correlation in time of the mean zero response, $r(t)$, and stimulus, $s(x_i,t)$:

$$c_{\text{rs}}(x_i,u) = \frac{1}{T}\sum_{t=1}^{T}(s(x_i,t-u) - \bar{s}(x_i))(r(t) - \bar{r}). \quad (6)$$

Here, $\bar{s}(x_i)$ and $\bar{r}$ indicate the mean over time of the stimulus and response, respectively. When white noise is used to evoke responses, reverse correlation is equivalent to computing the spike-triggered average stimulus.

Correction for natural stimulus bias. Natural vision movies are not well modeled by white noise. Natural scenes have a $1/f^2$ spatial power spectrum (Fig. 1*E*) (Field, 1987) as well as complex higher-order spatial correlations (Field, 1993; Zetzsche et al., 1993; Schwartz and Simoncelli, 2001). Saccadic eye movements bias the temporal power spectrum, concentrating temporal energy in the range of 3–4 Hz (Fig. 1*E*) (Vinje and Gallant, 2000; Woods et al., 2001). These correlations introduce a bias into the spike-triggered average, making it appear that a neuron is tuned to artificially low spatial and temporal frequencies (Theunissen et al., 2001; Willmore and Smyth, 2003; Machens et al., 2004).

To estimate an STRF from responses to natural vision movies accurately, stimulus bias must be removed from the spike-triggered average. Classical methods for bias correction assume the stimulus is stationary; i.e., the correlation between two points depends only on the distance between them rather than their absolute position. In such cases, the cross-correlation can simply be normalized by the power spectrum of the stimulus (Marmarelis and Marmarelis, 1978). In general, the spatial autocorrelation of natural scenes (Field, 1987) and the temporal autocorrelation of a random sequence of fixations are stationary. However, after the nonlinear phase-separated Fourier transformation, the spatial autocorrelation is no longer stationary. A correction for stationary temporal and nonstationary spatial stimulus autocorrelation is required (Theunissen et al., 2001).

The autocorrelation of a stimulus that is nonstationary in space but stationary in time is described by:

$$c_{\text{ss}}(x_i, x_j, u) = \frac{1}{T}\sum_{t=1}^{T}(s(x_i,t) - \bar{s}(x_i))(s(x_j,t+u) - \bar{s}(x_j)). \quad (7)$$

The value of c_{ss} indicates the strength of correlation between two points in space, x_i and x_j, separated by time lag u. The spike-triggered average, c_{rs} (Eq. 6), is related to the STRF by convolution with the stimulus autocorrelation (Theunissen et al., 2001):

$$c_{\text{rs}}(x_i,u) = \sum_{\nu=-U}^{U}\sum_{j=1}^{N} c_{\text{ss}}(x_i, x_j, u+\nu)h(x_j, \nu). \quad (8)$$

We remove stimulus bias by applying the inverse of the autocorrelation, c_{ss}^{-1}, to both sides of the equation:

$$h(x_i, u) = \sum_{\nu=-U}^{U}\sum_{j=1}^{N} c_{\text{ss}}^{-1}(x_i, x_j, u+\nu)c_{\text{rs}}(x_j, u). \quad (9)$$

The autocorrelation function, c_{ss}, can be described by a matrix, and c_{ss}^{-1} is simply its matrix inverse. This procedure has no effect on the estimated STRF when the stimulus has no autocorrelation (e.g., white noise). For correlated stimuli, the magnitude of the bias correction is proportional to the strength of stimulus autocorrelation. Because natural vision movies have much stronger autocorrelation than grating sequences, this procedure introduces a larger correction to STRFs estimated using natural vision movies than to those estimated using grating sequences.

STRF estimation procedure

Estimation and validation data sets. The STRF estimation procedure used here requires fitting many model parameters. In such cases, optimal predictions can only be obtained if care is taken to avoid overfitting to noise. Therefore we divided the data from each neuron into two different data sets: an estimation set that was used to estimate model parameters, and a validation set that was used exclusively to test predictions. The estimation set contained ~90% of the available data (repeated or single trial), and the validation set containing the remaining 10% of the data (5–10 sec/neuron, repeated trials) was reserved exclusively for evaluating predictions. The use of a separate validation data set ensured that our estimates of prediction accuracy would not be artificially inflated by overfitting.

Data preprocessing. We used the estimation data set to measure both stimulus–response cross-correlation and stimulus autocorrelation directly. To estimate STRFs, the stimulus was first cropped with a square window circumscribing twice the CRF diameter. The window was constant, regardless of the true stimulus size, which was two to four times the CRF diameter. To reduce both noise and computational demands, each stimulus frame was smoothed and downsampled to 18×18 pixel resolution before analysis. This low-pass filtering procedure preserved spatial frequencies up to 4.5 cycles per CRF, which was always high enough to reveal the spatial tuning profile from responses to grating sequences. In

theory an analysis using higher spatial resolution would produce a more accurate STRF model. Because of the bias toward low spatial frequencies in natural images, however, larger data sets would be required to achieve sufficient signal-to-noise levels at these high spatial frequencies. Edge artifacts were minimized by applying a Hanning window to each stimulus frame before applying the phase-separated Fourier transform.

The response PSTH, $r(t)$, was defined as the instantaneous spike rate within each time bin (or the mean spike rate when repeated trials were available). Well isolated spikes recorded from each neuron (1 msec resolution) were convolved with a boxcar filter (width 14 msec) and binned at 14 msec, synchronized with the 72 Hz refresh cycle of the CRT. STRFs were calculated across time lags ranging from 0 to 196 msec ($U = 14$ time bins).

Regularization procedure to reduce estimation error. STRF estimation requires fitting a large number of coefficients with a relatively small number of data samples. The accuracy of these estimates is therefore often limited by sampling, and regularization can substantially improve model accuracy. Our regularization procedure combined a jackknife algorithm and a pseudoinverse approximation for stimulus bias correction. First, 20 jackknife data sets were generated by excluding different 5% segments from the complete estimation data set (Efron and Tibshirani, 1986). Each jackknife set was used to obtain an STRF estimate, $h_n(x_i, u)$. The mean, $\bar{h}(x_i, u)$, and SE, $\hat{\sigma}(x_i, u)$, of each spatiotemporal channel were calculated from these jackknifes, and the ratio of mean to SE was taken as the signal-to-noise ratio for each channel. Coefficients were scaled according to a shrinkage filter to produce a final STRF estimate (Brillinger, 1996):

$$h(x_i, u) = \bar{h}(x_i, u)\ \sqrt{\left|1 - \gamma(\hat{\sigma}^2(x_i, u)/\bar{h}^2(x_i, u))\right|^+}. \qquad (10)$$

The brackets, $|\ldots|^+$, indicate half-wave rectification. The optimal shrinkage parameter, γ, varies according to global signal-to-noise level. STRFs were estimated for a range of γ (from 1.0 to 2.0), and the optimal value was chosen in conjunction the pseudoinverse tolerance value (see below).

As noted earlier, bias in the stimulus statistics must be removed from the STRF to obtain an accurate estimate of neuronal tuning properties. Bias removal involves multiplying the spike-triggered average by the inverse of the stimulus autocorrelation matrix (Eq. 9). However, the spatial autocorrelation matrix for natural vision movies is nearly singular, and the true inverse tends to amplify estimation noise. Therefore, we needed to approximate the spatial autocorrelation inverse for estimating inseparable STRFs and spatial response functions. (For temporal response functions, stimulus bias could be corrected without approximation.)

We used a singular value decomposition (SVD) algorithm to construct a pseudoinverse of the autocorrelation matrix, $c^{-1}_{approx}(x_i, x_j, u)$ (Theunissen et al., 2001; Smyth et al., 2003). Pseudoinverse construction via SVD requires selection of a tolerance value that determines the fraction of total stimulus variance preserved in the inverse. The optimal tolerance value is a function of both stimulus statistics and neural noise and cannot be determined a priori for a given neuron. To determine the optimal tolerance for each neuron, we calculated separate STRFs over a range of tolerance values. STRFs generated from each tolerance and shrinkage parameter (γ, from above) were used to predict responses in the entire estimation data set, including segments that were excluded in each jackknife estimate (but not using the reserved validation data). By selecting the tolerance value and shrinkage parameter simultaneously, we avoided overfitting to the estimation data. The STRF with the smallest mean square prediction error was selected as the optimal STRF.

Estimation of space–time separable STRFs. Space-time separable STRFs were obtained through an iterative process. First, a space–time inseparable STRF was estimated using the procedure described above. Because the inseparable STRF contains a large number of coefficients, the regularization procedure tended to select relatively underfit STRFs. Estimating spatial and temporal response functions separately allowed for improved signal-to-noise levels and thus less severe regularization.

Approximate spatial and temporal response functions were derived from the inseparable STRF by singular value decomposition. The first two components of the decomposition satisfy the criterion of best mean-squared error estimate of the full STRF,

$$[f_{approx}(x_i), g_{approx}(u)] = \arg\min_{f,g}(h(x_i, u) - f(x_i)\, g(u))^2. \qquad (11)$$

The sign of the approximate temporal response function, $g_{approx}(u)$, is ambiguous, given only in Equation 11. Its sign was fixed so that it produced a positive inner product with the inseparable STRF averaged over space. This function was used to estimate the spatial response function by convolution with the stimulus:

$$s_{space}(x_i, t) = \sum_{u=1}^{U} (s(x_i, t - u) - \bar{s}(x_i))\, g_{approx}(u). \qquad (12)$$

The result, $s_{space}(x_i, t)$, was the stimulus transformed so that stimulus energy correlated with the neural response was concentrated at a time lag of $u = 0$. The spatial response function, $f(x_i)$, was then estimated using Equations 6 and 9 but substituting $s_{space}(x_i, t)$ for the stimulus and constraining the maximum time lag to be zero, $U = 0$.

The spatial response function was then used to estimate the temporal response function. The stimulus was filtered by the spatial response function:

$$s_{time}(t) = \sum_{i=1}^{N} (s(x_i, t) - \bar{s}(x_i)) f(x_i). \qquad (13)$$

The resulting $s_{time}(t)$ indicated how well the stimulus matched the spatial response function at each point in time. The temporal response function, $g(u)$, was estimated using Equations 6 and 9 for the case with only one dimension of space, $N = 1$. Spatial and temporal response functions estimated in these last two steps were then combined according to Equation 3 to form the space–time separable STRF.

Estimation of hybrid STRFs. A hybrid STRF is a space–time separable STRF with spatial response properties estimated using grating sequence data and with temporal response properties estimated using natural vision movie data. The spatial response function, $f_{syn}(x_i)$, was estimated using grating sequence data as described above. Then $f_{syn}(x_i)$ was used with natural vision movie data to estimate the temporal response function, $g_{nat}(u)$. Other than using different data sets to estimate spatial and temporal components, this procedure was identical to that used to estimate other space–time separable STRFs.

In our control analysis that studied the effects of natural spatial statistics alone, we also estimated hybrid STRFs using natural image sequence data. In this case, the spatial response function was estimated using natural image sequence data, and the temporal response function was estimated using either grating sequence data or natural vision movie data.

Estimation of positive space STRFs. A positive space STRF has no negative coefficients in its spatial response function. This constrains it so that it can account for spatially tuned excitation but not inhibition. Positive space STRFs were estimated using both natural vision movie and grating sequence data. First, spatial response functions were estimated using the appropriate data set. Then their negative coefficients were set to zero according to Equation 5, producing positive spatial response functions $f^+_{nat}(x_i)$ and $f^+_{syn}(x_i)$, respectively. Finally, temporal response functions were estimated for both using natural vision movie data. Combined with the spatial response functions, this produced a positive space natural vision STRF and a positive space hybrid STRF. In general, temporal response functions were quite similar for positive space STRFs and for STRFs estimated using natural vision movies.

Nonlinear threshold estimation. The STRF model also contains a nonlinearity that represents response threshold (θ in Eq. 1). The threshold for each STRF was selected only after applying the jackknife filter and selecting the optimal SVD cutoff. The threshold was chosen to maximize correlation between predicted and observed responses in the estimation data set. Again, validation data were not used at this or any other stage of the STRF estimation procedure.

Generating and evaluating predictions

For each neuron, STRFs obtained using natural vision movies and grating sequences were evaluated in terms of their ability to predict responses using a novel validation data set not used for STRF estimation. Predicted responses were generated according to the procedure outlined in Figure 2*A*. First, the validation stimuli were cropped and downsampled to match the size used for estimation. Second, they were then transformed according to the phase-separated Fourier nonlinearity in Equation 2. Responses were binned at 14 msec in the same manner as the estimation data. Third, the estimated STRF was convolved with the transformed stimulus in time, summed over space and thresholded according to Equation 1. Finally, prediction accuracy was quantified in terms of the correlation coefficient between the predicted and observed response (Pearson's *r*). Note that correlation measurements are strongly influenced by the size of time bins and by temporal smoothing. Comparison of prediction correlation between analyses therefore requires careful consideration of how the response has been binned and smoothed (Theunissen et al., 2000). Correlations reported in this study were all measured against validation responses sampled at 14 msec and smoothed by a 14 msec boxcar filter.

A Matlab implementation of this STRF estimation and validation procedure is available for download at http://strfpak.berkeley.edu.

Comparing response properties

Temporal inhibition index. Temporal response properties observed with the different stimulus classes were compared by means of a temporal inhibition index, which measures the ratio of negative power to total power in the temporal response function:

$$a = \frac{\sum_{u=1}^{U} (|g(u)|^{-})^2}{\sum_{u=1}^{U} (g(u))^2}. \quad (14)$$

Values of *a* are constrained to fall between 0 and 1. Small values correspond to temporal response functions with little negative power, which suggests that the neuron shows little inhibition in the temporal response function.

Normalization of residual spatial bias. An important consideration when comparing spatial response functions is that STRFs estimated using natural vision movies may be residually biased by natural spatial statistics. This is a well known limitation of the pseudoinverse correction method (Theunissen et al., 2001; Smyth et al., 2003; Machens et al., 2004). For natural spatial response functions, the effect of the pseudoinverse correction is to damp power at high spatial frequencies at which the signal-to-noise ratio is worst, effectively preserving the bias in those channels. Spatial response functions estimated using grating sequences do not suffer from residual correlation bias because grating sequences have only a weak autocorrelation (resulting from the presence of a single grating in each frame). Thus, even if the response properties of a neuron are the same under the two stimulus conditions, a residually biased natural spatial response function and an unbiased grating spatial response function may appear different.

Although spatial correlation bias cannot be removed entirely from natural vision spatial response functions, the residual bias can be measured. This bias can be applied to the unbiased grating spatial response function for the same neuron. After bias normalization, any remaining differences between two spatial response functions must reflect differences in the spatial tuning properties of the neuron. If a natural vision movie has spatial autocorrelation, $c_{ss}(x_i,x_j)$, and its inverse is approximated, $c^{-1}_{approx}(x_i,x_j)$, then the grating spatial response function with normalized residual bias is:

$$f_{norm}(x_i) = \sum_{j=1}^{N} \sum_{k=1}^{N} c^{-1}_{approx}(x_i,x_j) c_{ss}(x_j,x_k) f_{syn}(x_k). \quad (15)$$

(Because we are considering spatial response functions only, we can consider the spatial autocorrelation with no time lag.)

In the data reported here, the correlation bias was not as severe in the temporal dimension. Residual bias in the natural vision temporal response functions would drive apparent tuning toward lower temporal frequencies than for gratings, whereas our data showed a strong trend in the opposite direction.

Spatial similarity index. We compared spatial response functions obtained with different stimulus classes using a spatial similarity index. The index was computed by taking the correlation between the natural vision spatial response function and the bias-normalized grating spatial response function:

$$b = \frac{\sum_{i=1}^{N} f_{nat}(x_i) f_{norm}(x_i)}{\sqrt{\sum_{i=1}^{N} f^2_{nat}(x_i) \sum_{i=1}^{N} f^2_{norm}(x_i)}}. \quad (16)$$

Values of *b* can range from −1 to 1. A value near 1 indicates a high degree of similarity between spatial response functions, and a value near 0 indicates no similarity. Values near −1 indicate anticorrelated spatial response functions, but this rarely occurs in practice.

Results

We recorded from 74 neurons in the primary visual cortex of two animals performing a fixation task. Stimuli consisted of natural vision movies that simulated the spatial and temporal pattern of stimulation to a receptive field during free viewing of a natural scene (Fig. 1*A*,*B*; see Materials and Methods) (Field, 1987; Vinje and Gallant, 2000; Woods et al., 2001). Figure 1*B* illustrates the typical time-varying response evoked by a natural vision movie in a single neuron. Approximately 50 msec after the onset of each simulated fixation, the neuron shows a strong, transient increase in firing, followed by a weaker sustained response during the remainder of the fixation. The magnitudes of the transient and sustained responses vary from fixation to fixation. Thus, this neuron appears to code information both about the temporal dynamics of the stimulus as well as its spatial content. Natural vision movies evoke such rich, dynamic responses from most V1 neurons.

We also recorded responses to a second stimulus class, dynamic grating sequences, which consisted of a single sine wave grating whose spatial frequency, orientation, and phase varied randomly in each 72 Hz frame (Fig. 1*C*). A typical response to a grating sequence is shown in Figure 1*D*. This PSTH bears little similarity to the natural vision movie PSTH. Neither the temporal structure associated with fixation onsets nor the variations in response strength associated with different spatial patterns are obvious. These differences could arise from two factors. The tuning properties of the neuron may be the same in both cases, and the differences merely reflect differences in the statistical properties of the stimuli. Alternatively, the tuning properties of the neuron may be modulated by the statistical properties of the stimulus. This would imply that the neuron would transmit different information about the stimulus, depending on stimulus class (Theunissen et al., 2000).

Phase-separated Fourier STRFs

Our goal in this study was to determine whether stimulus statistics influence neuronal response properties and, if they do, to determine what specific mechanisms are affected. Comparison of response properties between stimulus classes required an objective framework independent of any parametric description of a

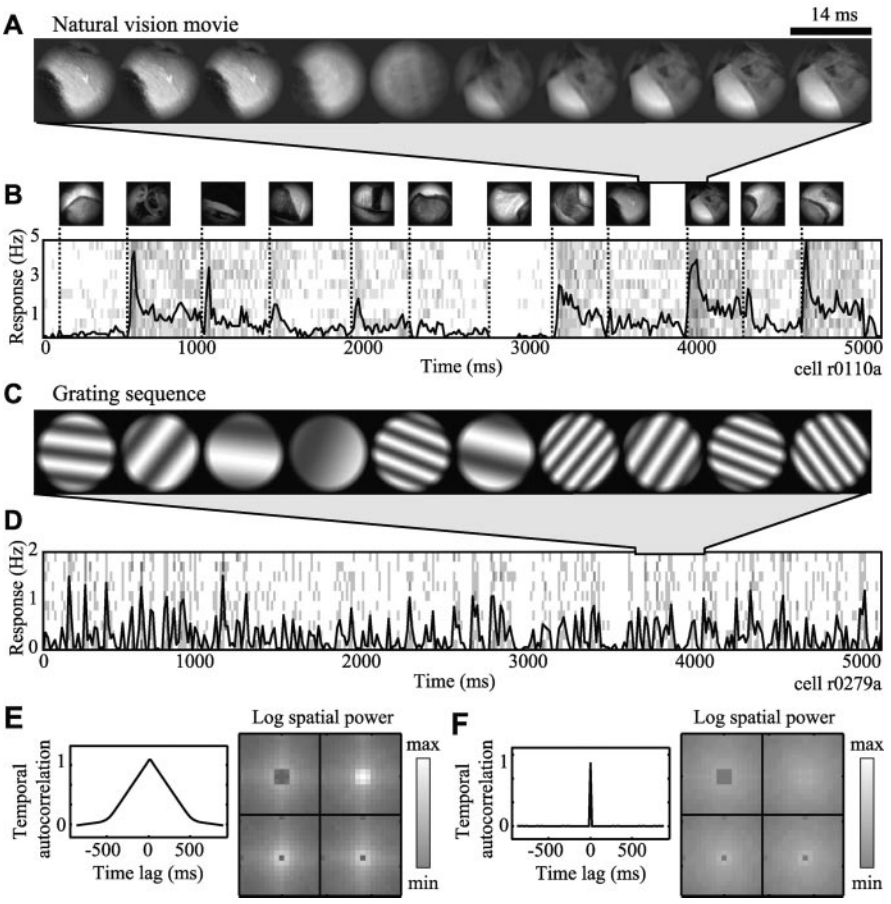

Figure 1. Natural and synthetic stimuli. *A*, Natural vision movies mimic the pattern of stimulation in a parafoveal receptive field during free viewing of a static, monochromatic natural scene. The stimulus remains constant during simulated fixations and changes rapidly during simulated saccades (compare frames 2, 3 with 5, 6). *B*, Five seconds of the same movie are shown schematically in the top row. The pattern appearing during each simulated fixation appears once, aligned on the left edge to the time of fixation onset. Below is the PSTH averaged over 10 repeated movie presentations. After the onset of each new fixation (dotted lines) the neuron responded with a brief burst of activity, followed by a weaker sustained response. *C*, Several frames from a grating sequence. The sine wave grating shown in each 14 msec frame varies randomly in orientation, spatial frequency, and spatial phase. *D*, The plot shows the PSTH averaged over 10 repeated presentations of a 5 sec grating sequence. *E*, Temporal and spatial statistics of a natural vision movie. Temporal autocorrelation (left) decreases linearly to zero at lags of ~500 msec because of the temporal dynamics of simulated saccades. The log spatial power spectrum (right) is plotted in the phase-separated Fourier domain, where each subpanel refers to a different spatial phase (for details, see Materials and Methods, Fig. 2). Power decreases linearly from low frequencies at the center of each subpanel, reflecting the $1/f^2$ power spectrum of natural images. *F*, Temporal and spatial statistics of a grating sequence, plotted in the same manner as *E*. There is no temporal autocorrelation at nonzero time lags because the grating pattern changes randomly in each frame. Log spatial power is nearly uniform, reflecting the sampling of grating parameters.

particular stimulus class. Because the population of neurons in V1 is composed of both simple and complex cells, the analytic framework also had to be applicable to both cell classes. Given these constraints, we chose to characterize neurons by estimating STRFs from the responses to each stimulus class (Theunissen et al., 2001; Smyth et al., 2003). To ensure that STRFs could be estimated for both simple and complex cells, we developed a new, nonlinear phase-separated Fourier model that accounts for response properties of both cell types (Fig. 2*A*; see Materials and Methods) (David et al., 1999). According to this model, each STRF is expressed as a set of coefficients describing how neuronal responses are influenced by stimulus orientation, spatial frequency, spatial phase, and time lag.

To confirm that the phase-separated Fourier model could recover tuning properties of both simple and complex cells, we applied our estimation algorithm to data produced by model neurons. The model simple cell consisted of a temporally modulated Gabor function (Daugman, 1980) followed by a rectifying nonlinearity and Poisson spike generator (Tolhurst et al., 1983). Model parameters were peak orientation 30° counterclockwise from vertical, peak spatial frequency two cycles per receptive field, even spatial phase, and peak latency of 49 msec. Responses were generated by stimulating the model cell with a 9000-frame natural vision movie, comparable with the quantity of data collected for many of our neurons. The phase-separated Fourier STRF estimated from these data fully recovers the tuning properties of the simple cell. Figure 2*B* displays the STRF as a series of panels representing spatial tuning at progressively later time lags. At each time lag, the four subpanels show spatial frequency and orientation tuning at four spatial phases. Orientation and spatial frequency are plotted in the Fourier domain; the radial component gives spatial frequency, and the angular component gives orientation. Light regions indicate positive STRF coefficients (excitatory relative to the mean response), and dark regions indicate negative coefficients (inhibitory). Relative excitatory responses are confined to the top right subpanel ($\phi = 0$), consistent with the fact that the model had even, on-center spatial phase tuning. The top left subpanel reveals weak relative inhibition at a 180° phase offset, reflecting the influence of the threshold nonlinearity. Scattered nonzero values in other phase subpanels reflect estimation noise resulting from finite sampling and noise in the response of the model neuron. Figure 2*D* shows the spatial and temporal response functions that compose the simple-cell STRF. The left panel shows the spatial patterns used to generate responses. Because this was a simple cell, only one subunit was active.

We also applied the STRF estimation algorithm to a model complex cell. This model was constructed by summing the rectified output of four simple cells in the quadrature phase before input to the Poisson spike generator. Orientation, spatial frequency, and temporal tuning were identical to those of the simple-cell model tested above, and responses were generated by stimulating with the same natural vision movie. The resulting STRF is shown in Figure 2*C*, with spatial and temporal response functions in Figure 2*E*. Relative excitatory responses appear at the peak time lag in all four subpanels, reflecting the phase invariance of the complex cell. Other tuning properties are identical to those of the simple cell, as expected.

Neuronal responses during simulated natural vision

Figure 3 compares the STRFs obtained from one V1 neuron using both natural vision movies and grating sequences. Tuning properties can be visualized by decomposing each STRF into spatial

and temporal response functions. The spatial response function estimated using a natural vision movie (Fig. 3*A*) is tuned to horizontal orientations and low spatial frequencies. The temporal response function shows excitation at short time lags, followed by strong inhibition at later time lags. Compared with the natural vision STRF, the STRF estimated using a grating sequence (Fig. 3*B*) preserves tuning for horizontal orientations but is tuned to slightly higher spatial frequencies. The temporal response is also substantially different from the one obtained with natural vision movies. It has a slightly longer latency and shows no inhibition at later time lags.

To determine which of these STRFs better described natural visual responses, we compared how well they could predict responses to a second natural vision movie. Data from this second movie were not used to fit the original STRFs, so any differences in predictive power could not be an artifact of overfitting. The response predicted by the natural vision STRF is shown in Figure 3*C* (solid line) overlaid on the observed response (dashed line). Prediction accuracy was evaluated in terms of the correlation between the predicted and observed responses. A correlation of 1.0 indicates perfect prediction, whereas a value of 0 indicates a prediction that is no better than random. Given that many sources of noise limit the accuracy of STRF estimates (e.g., spiking noise, finite sampling, and unmodeled nonlinear response properties), it is unlikely that an STRF will predict perfectly. However, prediction correlations for two STRFs can be compared as a relative measure of which provides a better model of natural visual response properties. For this neuron, the natural vision STRF predicts the natural vision movie responses with a correlation of $r = 0.55$. The grating STRF performs significantly worse (Fig. 3*D*) ($r = 0.36$; $p < 0.05$, randomized paired t test). We infer from the difference in predictive power that the natural vision movie activated functionally important response properties in a different manner than their activation by the grating stimulus.

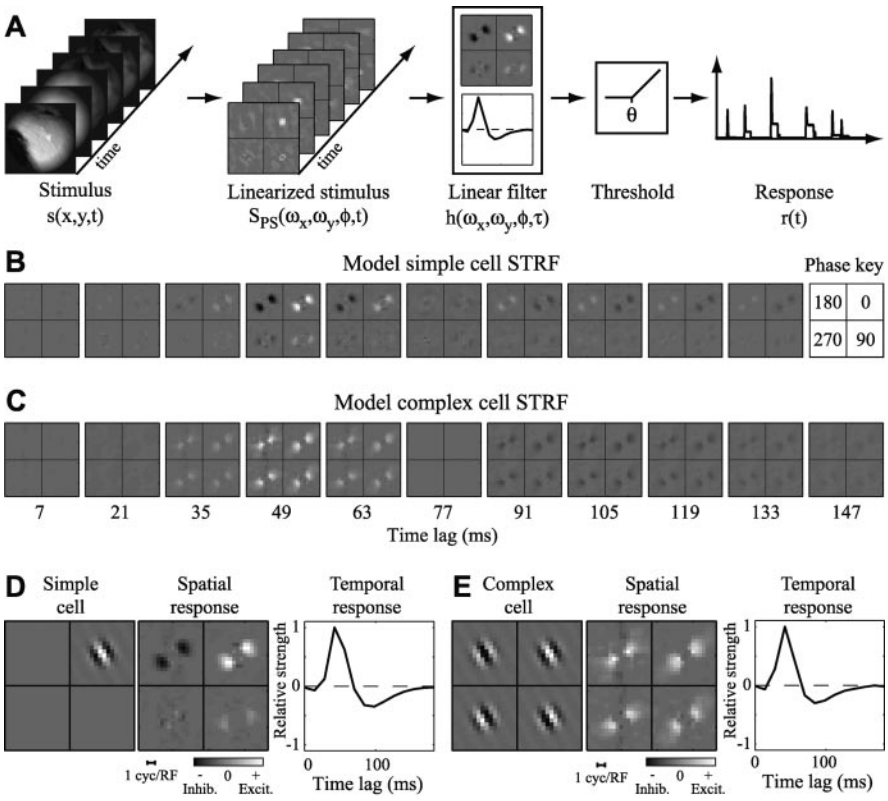

Figure 2. Linearized STRF model. The phase-separated Fourier model incorporates a nonlinear spatial transformation at its input stage to account for response properties of both simple and complex cells. *A*, Visual input is the time-varying sequence of gray scale images at the left. The stimulus is Fourier-transformed and projected onto quadrature spatial phase channels according to Equation 2. The transformed stimulus is convolved with a spatiotemporal filter and thresholded according to Equation 1 to produce the instantaneous firing rate, $r(t)$. *B*, STRF estimated using natural vision movie data for a model simple cell. Brighter regions indicate input channels correlated with an increase in firing (excitation), whereas dark regions indicate channels correlated with a decrease in firing (inhibition). At each time lag, the four subpanels show spatial frequency and orientation tuning at four spatial phases (key at far right). Spatial frequency and orientation are plotted in the Fourier domain. Radial position in the subpanel corresponds to spatial frequency, and angle corresponds to orientation. For this neuron, excitatory responses are confined to the top right subpanel, consistent with the fact that the model simple cell had even, on-center spatial phase tuning. The top left subpanel reveals inhibition at a 180° phase offset, reflecting the linear phase tuning of the simple cell. *C*, STRF estimated for a model complex cell. Spatial and temporal tuning resemble that in *B*, except all four phase channels drive excitatory responses. *D*, Spatial (middle) and temporal (right) response functions for the simple cell STRF in *B*. At left is the Gabor function showing the actual spatial tuning of the model simple cell. The spatial response function shows excitatory tuning at the phase corresponding to the even, on-center Gabor. *E*, Spatial and temporal response functions composing the complex cell STRF in *C*. The model complex cell is excited by spatial patterns matching any of the four Gabor functions at the left. Thus, all four phase channels indicate excitatory tuning at the corresponding orientation and spatial frequency. Inhib., Inhibition; Excit., excitation.

In the first example, temporal response properties were strongly affected by natural stimulus statistics, whereas spatial response properties were mostly unchanged. For other neurons, such as in Figure 4, both temporal and spatial response properties are dependent on stimulus statistics. For this neuron, the natural vision STRF (Fig. 4*A*) is tuned to near-horizontal orientations at low spatial frequencies, with little spatially tuned inhibition. Peak excitatory latency is at 40 msec, followed by an inhibitory component at later time lags. The STRF estimated using a grating sequence (Fig. 4*B*) has similar orientation tuning but prefers higher spatial frequencies. Unlike the natural vision response, the grating spatial response function also has substantial inhibitory tuning. As in the previous example, the grating temporal response function has a much weaker late inhibitory component than the natural vision temporal response. These tuning differences affect STRF predictions of natural vision movie validation responses. The natural vision STRF predicts responses with a correlation of $r = 0.49$ (Fig. 4*C*), whereas the grating STRF predicts with a significantly lower correlation of just $r = 0.22$ (Fig. 4*D*) ($p < 0.05$, randomized paired t test).

Predictive power of STRFs depends on estimation stimulus

To determine the prevalence of changes in tuning properties arising from differences in stimulus statistics, we compared the predictive power of natural vision and grating STRFs across a sample of 44 neurons for which appropriate data were available. For each neuron, predictions were evaluated with natural vision movie data that were not used in STRF estimation (i.e., a validation data set). In 24 neurons (55%), natural vision STRFs predict natural vision movie responses significantly better than grating STRFs ($p < 0.05$, randomized paired t test) (Fig. 5*A*, filled circles). In

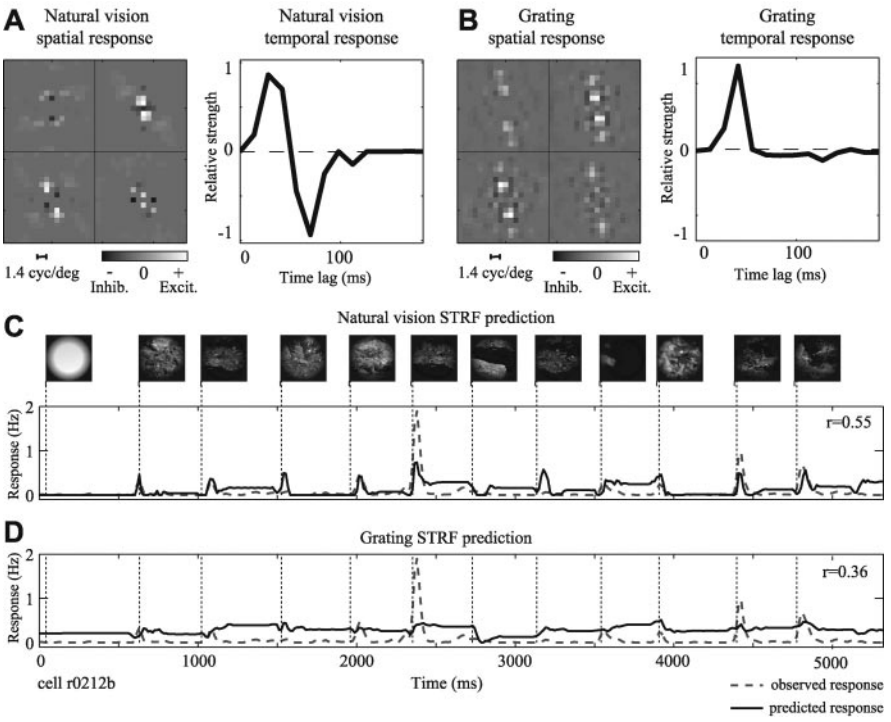

Figure 3. Natural stimulus statistics influence temporal response properties. *A*, STRF estimated using natural vision movie. Axes are as in Figure 2*D*. The spatial response function, left, is tuned to stimuli just counterclockwise of horizontal, with peak spatial frequency tuning of 1.1cycle/°. The temporal response function shows a peak latency of 35 msec followed by a negative, inhibitory component at greater time lags (63–91 msec). *B*, The STRF estimated using a grating sequence shows some differences in tuning. The spatial response function is similar to that in *A*, although with slightly higher excitatory spatial frequency tuning (peak, 1.4 cycle/°). However, the temporal response is quite different, with a peak latency of 49 msec and no late inhibitory component. *C*, STRFs were compared by measuring their ability to predict natural vision movie validation data. The prediction by the natural vision movie STRF (solid line) is overlaid on the observed PSTH (dashed line). This neuron gave a highly transient response during each fixation epoch in the natural vision movie. The predicted PSTH matches these transients well, with a correlation of $r = 0.55$. *D*, The grating sequence STRF fails to predict the transient responses and has a significantly lower prediction correlation ($r = 0.36$; $p < 0.05$, randomized paired *t* test). Inhib., Inhibition; Excit., excitation.

only 3 cases (7%) do natural vision STRFs predict responses significantly less accurately than grating STRFs ($p < 0.05$, randomized paired *t* test) (Fig. 5*A*, shaded circles). Across the entire sample of 44 neurons, natural vision STRFs predict with a mean correlation of 0.42, whereas the correlation for grating STRFs is just 0.19. This difference is significant ($p < 0.001$, randomized paired *t* test). When all 74 neurons with natural vision movie data are included, the mean correlation of natural vision STRF predictions with observed responses is 0.38. This is not significantly different from the mean obtained with the subset of natural vision STRFs analyzed above.

In the above comparisons, the superior predictive power of natural vision STRFs could simply reflect a greater degree of noise in the grating STRF estimates. That is, both models could predict equally well, but the grating STRFs might suffer from noisier data. This is unlikely, given that similar amounts of data were available for both classes. In fact, grating STRFs should be less noisy than those obtained with comparable amounts of natural vision movie data because grating sequences sample a sparse stimulus space with little stimulus autocorrelation. Nevertheless, to ensure that the predictions of grating STRFs were not noise-limited, we evaluated how well natural vision and grating STRFs predicted responses to a reserved validation data set collected with grating sequences (single-trial data). Across the sample of 44 neurons, natural vision STRFs predict grating responses with a mean of only 0.11. In contrast, grating STRFs predict with a mean of 0.31, a significant improvement ($p < 0.001$, randomized paired *t* test). Thus, both natural vision and grating STRFs predict responses best within the stimulus class used for estimation, and they predict poorly across stimulus classes. The consistent advantage of within-class predictions confirms that response properties of V1 neurons are modulated by differences between the stimulus classes. Given that neural response properties are nonlinear, a model fit to one stimulus class should, in principal, predict responses to that stimulus class better than a model fit to another class. Because of the complexity inherent in natural stimuli, however, achieving good fits directly from natural stimuli has proven difficult in practice. This is the first demonstration that more accurate fits can actually be made with natural stimuli in V1.

Temporal response properties during natural visual stimulation

The results presented above demonstrate that natural vision movies and grating sequences evoke different, functionally important tuning properties. The phase-separated Fourier model used for STRF estimation can model both linear responses (which, by definition, do not vary across stimulus classes) and nonlinear phase invariant responses. The different stimulus classes do not drive responses that are more or less linear. Instead, differences between STRFs reflect the influence of additional unmodeled nonlinearities that are differentially activated by natural vision movies and grating sequences.

To obtain a better understanding of the nonlinear mechanisms underlying differences between STRFs estimated using the two stimulus classes, we divided potential nonlinearities into two distinct types: temporal nonlinearities that change the temporal response profile and spatial nonlinearities that change spatial tuning. To investigate these two types of nonlinearity separately, we decomposed the space–time separable STRFs (Eq. 3) into their spatial and temporal response functions.

Characterization of temporal nonlinearities

One simple way to visualize the typical V1 temporal response under different stimulus conditions is to compute the average temporal response across a large sample of neurons. We obtained the average temporal response function across all 44 neurons for which both natural vision movie and grating sequence data were available (Fig. 6*A*,*B*). Individual temporal response functions were normalized to have unit variance before averaging. The average temporal responses obtained with each stimulus class are similar to those from the single neurons shown in Figures 3 and 4. Natural vision movies reliably evoke a biphasic temporal response: an initial excitatory component followed by a late negative component approximately one-third as large (Fig. 6*A*, solid line). In contrast, grating sequences evoke a monophasic positive

response with no substantial late negative component (Fig. 6*B*, solid line).

The running integral of the temporal response function is the step response. This gives the response we would expect from stimuli with the temporal structure of natural vision movies, when stimuli enter the receptive field abruptly and remain fixed there for several hundred milliseconds (Fig. 6*A*,*B*, dashed lines). The step response evoked by natural vision movies shows the transient behavior characteristic of responses to natural vision movies (Fig. 1*B*). In contrast, the step response evoked by grating sequences is sustained and shows little evidence of a transient response.

To quantify the difference in temporal response between stimulus classes, we computed a temporal inhibition index for each temporal response function. Index values were computed from the ratio of negative to total power in the temporal response function (Eq. 14). STRFs with a biphasic temporal response will have temporal inhibition indices near 0.5, whereas those that are monophasic will have indices near 0. Figure 6*C* compares indices obtained using the two stimulus classes. Values for natural vision temporal response functions are distributed broadly around a mean of 0.37. Indices for grating temporal response functions have a mean of 0.18 and are almost always lower than the corresponding index for natural vision movies. This difference is significant (randomized paired *t* test, $p < 0.001$), confirming that temporal inhibition is stronger during stimulation by natural vision movies than by grating sequences. The magnitude of the change in inhibition, however, varies substantially across neurons.

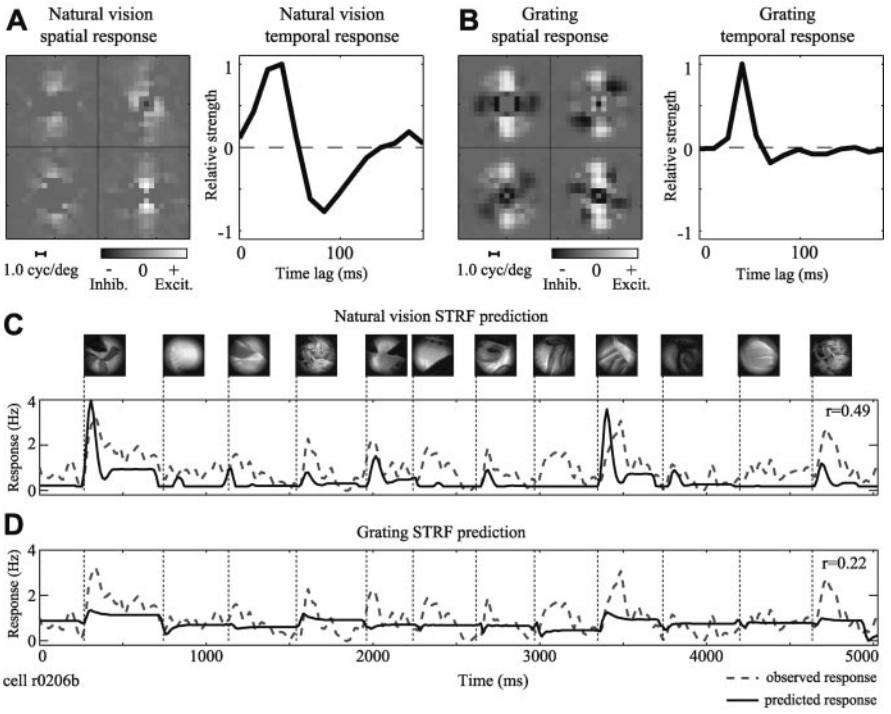

Figure 4. Natural stimulus statistics influence spatial tuning in some neurons. *A*, Spatial and temporal response functions estimated using natural vision movie data indicate a preference for horizontal stimuli at 1.0 cycle/°. Peak latency is 49 msec, followed by an inhibitory response (77–91 msec). *B*, The grating spatial response function also shows excitatory tuning to horizontal stimuli, but spatial frequency tuning is higher (2.4 cycles/°). Inhibitory tuning is also clearly present. The time course is quite brief (peak latency, 49 msec) and lacks a strong negative component. *C*, The natural vision STRF predicts a substantial portion of the observed PSTH ($r = 0.49$). *D*, The grating STRF predicts with significantly lower accuracy ($r = 0.22$; $p < 0.05$, randomized paired *t* test). In addition to missing transient responses, it also fails to predict modulation between fixations as well as the natural vision STRF. Inhib., Inhibition; Excit., excitation.

Contribution of nonlinear temporal responses to predictions

The preceding analyses demonstrate that natural visual stimuli evoke a strong, late inhibitory response that is absent during stimulation with grating sequences. Does this change in temporal response explain the fact that natural vision STRFs predict natural visual responses better than grating STRFs? To answer this question, we constructed a third, hybrid STRF for each neuron. Each hybrid STRF was created by combining the grating sequence spatial response function with the natural vision temporal response (Eq. 4). For example, to construct a hybrid STRF for the neuron shown in Figure 3, we combined the grating spatial response function from Figure 3*B* with the natural vision temporal response function from Figure 3*A*. Because the hybrid STRF followed the temporal profile of the natural vision STRF, any difference in its predictive power relative to the natural vision STRF must reflect differences in spatial response properties evoked by the two stimulus classes.

Figure 3 illustrates a case for which the hybrid STRF has substantially better predictive power than the grating STRF. For this neuron, saccadic transitions in the natural vision movie elicit large transient responses. The biphasic structure of the natural vision temporal response enables it to predict the transients, whereas the monophasic grating temporal response fails to do so (Fig. 3, compare *C*, *D*). The hybrid STRF, which incorporates the biphasic temporal response, predicts responses to natural vision movies with a correlation of $r = 0.66$, significantly better than the grating STRF ($r = 0.36$; $p < 0.05$, randomized paired *t* test). This correlation is greater, although not significantly different from the prediction correlation obtained using the natural vision STRF ($r = 0.55$). Thus, for this neuron, the difference in the ability of natural vision and grating STRFs to predict natural visual responses stems primarily from a temporal nonlinearity evoked differently by the two stimulus classes.

For the neuron in Figure 4, a nonlinear temporal response does not account entirely for the discrepancy between natural vision and grating STRFs. The poor performance of the grating STRF can be attributed only partially to its inability to predict transient responses. The hybrid STRF prediction ($r = 0.33$) is better than that of the grating STRF ($r = 0.22$; $p < 0.05$) but still significantly lower than that of the natural vision STRF ($r = 0.49$; $p < 0.05$). Instead, the difference in the ability of natural vision and grating STRFs to predict natural visual responses reflects both spatial and temporal nonlinearities evoked differently by the two stimulus classes.

In general, hybrid STRFs predict natural vision movie responses better than grating STRFs. The mean prediction correlation for hybrid STRFs is 0.30, whereas the mean prediction correlation for grating STRFs is 0.19, a significant difference ($p < 0.001$, randomized paired *t* test). Despite their improvement, however, hybrid STRFs do not predict natural vision movie responses as accurately as natural vision STRFs on average (Fig. 5*B*). The mean prediction correlation for hybrid STRFs is signif-

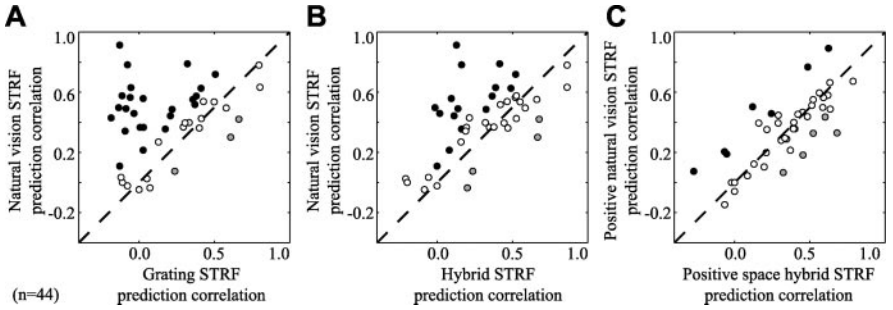

Figure 5. Predictive power depends on estimation stimulus class. *A*, Comparison of natural vision STRF and grating STRF predictions of natural vision movie validation responses. The position on the *x*-axis shows the correlation (Pearson's *r*) between the grating STRF prediction and the observed response. The *y*-axis shows the prediction correlation for natural vision STRFs. Filled points above the dashed line correspond to neurons with significantly more accurate natural vision STRF predictions, whereas shaded points below the line indicate more accurate grating STRF predictions ($p < 0.05$, randomized paired *t* test). Natural vision STRFs predict responses (mean $r = 0.42$) consistently better than grating STRFs (mean $r = 0.19$; $p < 0.001$, randomized paired *t* test; $n = 44$). *B*, Comparison of natural vision movie predictions by hybrid STRFs and natural vision STRFs. Hybrid STRFs are composed of grating spatial response functions and natural vision temporal response functions. Natural vision STRFs predict responses (mean $r = 0.42$) significantly better than hybrid STRFs (mean $r = 0.30$; $p < 0.001$, randomized paired *t* test). Thus, incorporating natural vision temporal responses into both STRF classes decreases but does not account entirely for the gap in predictions. *C*, Comparison of natural vision movie predictions by positive space hybrid STRFs and positive space natural vision STRFs. These STRFs have excitatory spatial tuning estimated using grating sequence and natural vision movie data, respectively. However, both incorporate natural vision temporal responses and have negative coefficients removed from their spatial response functions. Positive space natural vision movie STRFs (mean $r = 0.34$) predict responses no better than positive space hybrid STRFs (mean $r = 0.35$; not significant, randomized *t* test). Thus, changes in temporal response and spatially tuned inhibition account for the differences in predictive power of STRFs estimated using the two stimulus classes.

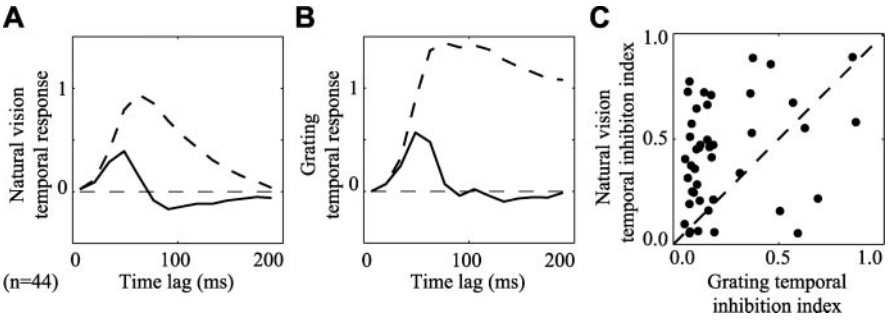

Figure 6. Statistics of natural stimuli affect temporal response properties. *A*, The mean natural vision temporal response function (solid line) shows a strong biphasic pattern: excitation at early latencies, peaking at 49 msec, followed by a negative component. The step response (dashed line), found by integrating the temporal response in time, predicts transient responses to fixation epochs. *B*, The mean grating temporal response function is primarily monophasic. The time course of the excitatory component is similar to that for natural vision movies, but it lacks the late inhibitory component. The step response predicts a sustained response without a transient. *C*, Scatterplot compares temporal inhibition indices (Eq. 14) measured using grating sequences (*x*-axis) with those measured using natural vision movies (*y*-axis). Larger index values correspond to a stronger negative component in a temporal response function. Most points lie above the line of unity slope, indicating that temporal inhibition index values were larger for natural vision movie responses. The mean inhibition index for natural vision movies (0.37) is significantly higher than for grating sequences (0.18; $p < 0.001$, randomized paired *t* test; $n = 44$).

icantly lower than for natural vision STRFs (mean, 0.30 vs 0.42, respectively; $p < 0.001$, randomized paired *t* test).

Spatial response properties during natural visual stimulation

The results in the previous section demonstrate that differences in temporal response properties account for some of the deficit in predictive power observed in grating STRFs, compared with natural vision STRFs. However, hybrid STRFs, which incorporate natural vision temporal response functions but preserve grating spatial response functions, still predict less accurately than STRFs estimated entirely using natural vision movies in 36% (16 of 44) of the neurons in our study (Fig. 5*B*, filled points). This suggests that the statistical properties of natural vision movies also drive changes in spatial response functions.

Spatial response functions estimated using the two stimulus classes for the same neuron are compared in Figure 7. The grating spatial response functions have been normalized to match the residual correlation bias in the natural vision spatial response functions according to Equation 15. In the first example (Fig. 7*A*), both excitatory and inhibitory tuning are similar in the two spatial response functions. We compared spatial response functions according to the spatial similarity index defined in Equation 16. Index values near 1 indicate a high degree of similarity between the functions, whereas values near 0 indicate little similarity. The similarity index of these spatial response functions is 0.88.

Not all neurons show such a high degree of similarity. The pattern in Figure 7*B* is typical of neurons showing differences in spatial tuning. Excitatory tuning is similar in both spatial response functions. However, inhibition is more tightly tuned in the natural vision spatial response function (left) than the grating spatial response function (right). This pair of spatial response functions has a similarity index of 0.68, lower than the previous example. Because the majority of difference is in inhibitory tuning, we also measured the similarity index separately for the positive and negative components of the functions. For this neuron, the similarity index of the positive spatial response functions is 0.85, whereas the similarity of the negative spatial response functions is just 0.38.

The histogram in Figure 8*A* plots the distribution of similarity indices between spatial response functions. Index values are distributed broadly around a mean of 0.37. Figure 8*B* shows a histogram of similarity indices for the positive spatial response functions. Relative to the full spatial response comparison, the distribution is shifted significantly toward greater similarity, with a mean of 0.49 ($p < 0.001$, randomized paired *t* test). In contrast, the distribution of index values for the negative components of spatial response functions (Fig. 8*C*) is biased toward lower values, with a mean of 0.30 ($p < 0.05$, randomized paired *t* test). Thus, similarity is greater for excitatory tuning rather than inhibitory tuning across the entire set of neurons.

Contribution of nonlinear spatial inhibition to predictions

The comparison of spatial tuning properties for different stimulus classes (Fig. 8) suggests that differences in inhibitory tuning may explain the superior predictive power of natural vision STRFs over hybrid STRFs. If this hypothesis is true, then removing the negative component of the spatial response function before constructing space–time separable STRFs should cancel the difference in predictive power. To test this, we compared the predictive power of two final STRF classes that controlled for the differences in spatially

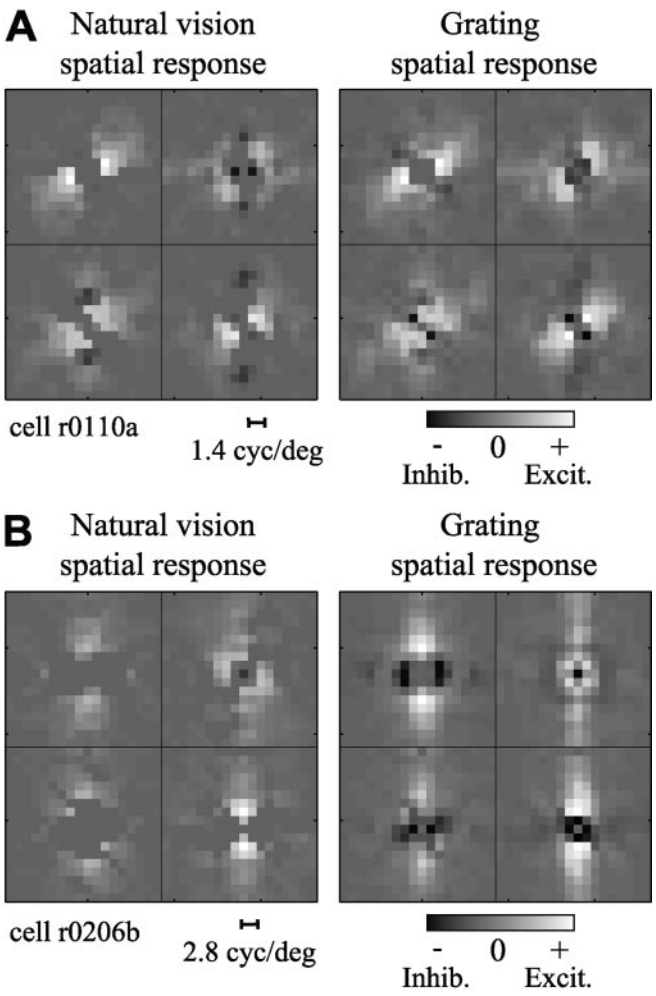

Figure 7. Spatial response functions compared between stimulus classes. *A*, Natural vision (left) and grating (right) spatial response functions estimated for a single neuron. Both have excitatory tuning to diagonal orientations at 3.0 cycles/°. Inhibitory tuning is localized to a small range of spatial channels. Residual correlation bias has been normalized for this spatial response function pair. *B*, Other neurons in V1 reveal stimulus-dependent patterns in their spatial response functions. The natural vision and grating spatial response functions for this neuron have similar excitatory tuning, but inhibition is much broader in the grating spatial response. The effect of residual bias normalization can be seen by comparing this grating spatial response function with Figure 4*B*. Inhib., Inhibition; Excit., excitation.

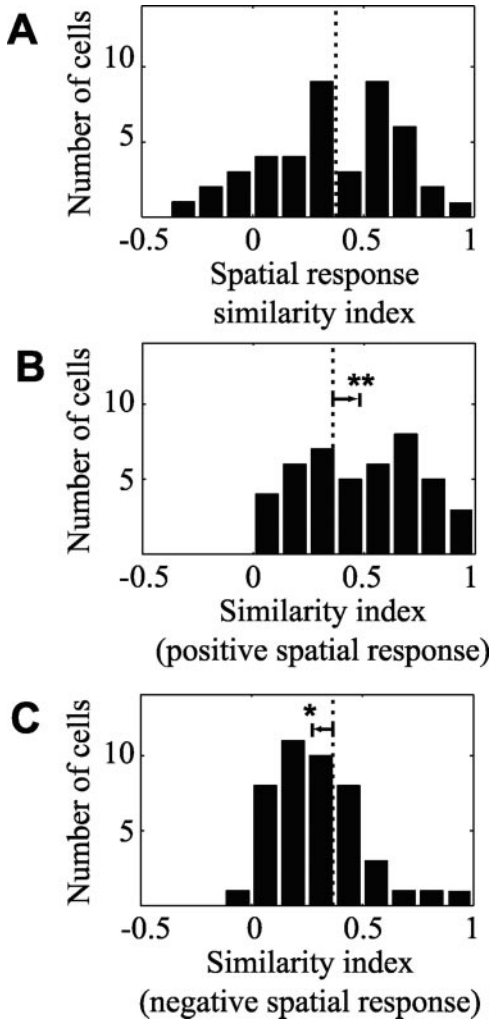

Figure 8. Natural stimuli affect spatially tuned inhibition. *A*, Histogram of similarity index values (Eq. 16) between spatial response functions estimated using natural vision movies and grating sequences. The distribution of index values has a mean of 0.37 (dotted line; $n = 44$). *B*, Histogram of similarity index values between only the positive coefficients of the same spatial response functions. The distribution is shifted significantly toward higher values relative to *A* (arrow), with a mean of 0.49 ($p < 0.001$, randomized paired *t* test). *C*, In contrast, the histogram of spatial similarity between the negative coefficients reveals a significantly lower mean of 0.30 ($p < 0.05$, randomized paired *t* test), suggesting that the greater difference between spatial response functions lies in inhibitory tuning.

tuned inhibition. First, the positive space natural vision STRF was estimated by extracting only the positive component of a natural vision spatial response function (Eq. 5) before estimating the temporal response function using natural vision movie data. Second, the positive space hybrid STRF was estimated using only the positive component of a grating spatial response function and the natural vision temporal response function.

The effect of removing inhibitory spatial tuning from natural vision and hybrid STRFs is summarized in the scatterplot in Figure 5*C*. Positive space hybrid STRFs predict with a mean correlation of 0.35, which is not significantly different from the mean correlation of 0.34 for positive space natural vision STRFs (randomized paired *t* test; $n = 44$). This result reflects primarily a decrease in predictive power for natural vision STRFs, whereas STRFs with spatial tuning estimated using grating sequences show a small but insignificant improvement. The distribution of points in Figure 5*C* is clustered along the line of unity slope, indicating that both models perform similarly for most neurons. Thus, increased late temporal inhibition and changes in spatially tuned inhibition account for the differences in response properties we observed under the two stimulus conditions.

Effects of natural spatial versus temporal stimulus statistics

The preceding results demonstrate that synthetic gratings and natural visual stimuli elicit different spatial and temporal response properties in V1 neurons. Ultimately, understanding the nonlinear mechanisms that underlie these differences in response properties will require identifying the specific features of natural stimuli that cause them. This is a complicated issue because natural vision movies and grating sequences differ in many ways and the statistical structure of natural images is only partially understood. It may be possible to discover general principles, however, such as how basic differences in spatial and temporal stimulus statistics influence spatial and temporal response properties.

In theory, differences in either spatial or temporal stimulus statistics could cause changes in the spatial response properties of

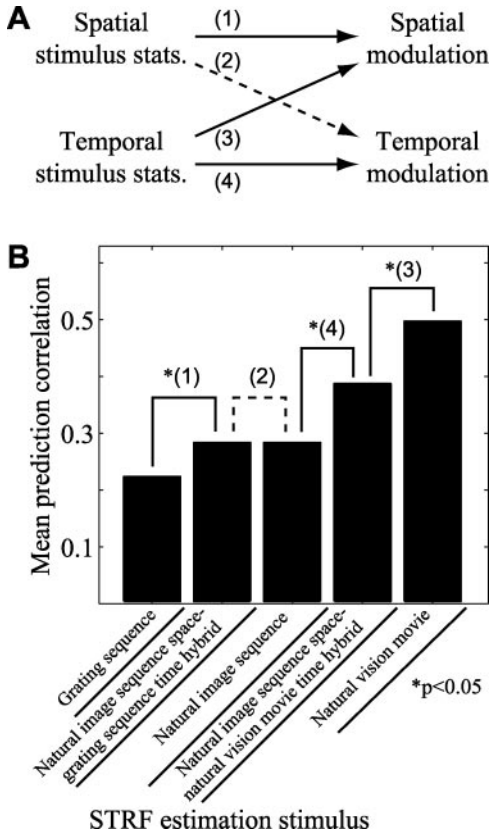

Figure 9. Sources of stimulus-dependent modulation. *A*, Differences in either spatial or temporal stimulus statistics (stats.) could potentially modulate spatial and temporal response properties in V1. The four possible relationships between stimulus statistics and response modulation are indicated by arrows 1–4. Solid arrows indicate observed relationships. *B*, We assessed these potential relationships by comparing predictions of hybrid STRFs estimated using natural vision movies, grating sequences, and natural image sequences ($n = 21$ neurons). Labels marking each comparison indicate the relevant relationship in *A*. Solid lines indicate significant improvements in prediction accuracy ($p < 0.05$). These comparisons indicate that natural temporal stimulus statistics modulate both spatial and temporal response properties. Natural spatial statistics modulate spatial response properties, but they do not influence temporal response properties (dashed line). For details, see Results.

V1 neurons, changes in their temporal response properties, or changes in both temporal and spatial properties (Fig. 9*A*). To explore these possibilities, we collected data using a third stimulus class, natural image sequences. Natural image sequences consist of a stream of natural image patches chosen randomly for each 14 msec frame. Thus, they share spatial statistics with natural vision movies but share temporal statistics with grating sequences.

Effects of spatial stimulus statistics on response properties

To determine whether the spatial statistics of natural images influence responses in V1 in any way (Fig. 9*A*, arrows 1, 2), we compared the predictive power of STRFs estimated using natural image sequences with those obtained with grating sequences. Predictions were evaluated against responses obtained with natural vision movies. We reasoned that if STRFs estimated using stimuli with natural spatial statistics are different from those estimated from gratings, then natural image sequence STRFs should predict natural visual responses more accurately.

For the 21 neurons with appropriate data, natural image sequence STRFs predict with a mean correlation of 0.28, whereas grating STRFs predict with a mean of just 0.22, a significant difference ($p < 0.05$, randomized paired t test) (Fig. 9*B*). Thus, differences in the spatial statistics of gratings and natural stimuli elicit different response properties in area V1.

Relationship between spatial stimulus statistics and temporal response properties

We next determined whether the spatial statistics of natural images affect the temporal response properties of V1 neurons (Fig. 9*A*, arrow 2). If this is true, then a hybrid STRF that combines the spatial response function obtained with natural image sequences and the temporal response function obtained with grating sequences should predict natural visual responses less accurately than the original natural image sequence STRF.

However, STRFs estimated using natural image sequences predict responses to natural vision movies no better than these hybrid STRFs (mean correlation, 0.28 vs 0.27, respectively; not significant, randomized paired t test) (Fig. 9*B*). The previous section demonstrated that natural image sequence STRFs predict natural visual responses better than grating STRFs. The difference in predictive power is eliminated when the spatial response function evoked by grating sequences is replaced with the spatial response evoked by natural image sequences. Therefore, we conclude that differences between the spatial statistics of natural stimuli and gratings have a significant influence on the spatial response properties of V1 neurons (Fig. 9*A*, arrow 1) but do not influence their temporal response properties (Fig. 9*A*, arrow 2).

Relationship between temporal stimulus statistics and spatial response properties

We performed a similar analysis to determine whether natural temporal statistics affect the spatial response properties of V1 neurons (Fig. 9*A*, arrow 3). If this is true, then a hybrid STRF that combines the spatial response function obtained with natural image sequences and the temporal response function obtained with natural vision movies should not predict natural visual responses as well as the original natural vision movie STRF.

Across the sample of 21 neurons, these hybrid STRFs predict natural visual responses with a mean correlation of 0.38, whereas the original natural vision STRFs show a mean correlation of 0.49. This difference is significant ($p < 0.05$, randomized paired t test) (Fig. 9*B*). These STRFs were estimated using stimuli having the same spatial statistics but different temporal statistics, and they use a common temporal response function. Therefore, this difference in predictive power must reflect nonlinear modulation of the spatial response properties of V1 neurons by temporal stimulus statistics.

Relationship between temporal stimulus statistics and temporal response properties

Results presented above (see Fig. 5) have already demonstrated that temporal response properties of V1 neurons depend on temporal stimulus statistics (Fig. 9*A*, arrow 4). For completeness, we also addressed this issue using data acquired with natural image sequences. If temporal stimulus statistics influence temporal responses, then a hybrid STRF that combines the spatial response profile obtained with natural image sequences and the temporal response profile obtained using natural vision movies should predict natural visual responses better than the original natural image sequence STRF.

As noted earlier, STRFs estimated using natural image sequences predict responses to natural vision movies with a mean correlation of 0.28. The hybrid STRFs predict with a mean of 0.38, a significant difference ($p < 0.05$, randomized paired t test) (Fig. 9*B*). This confirms that natural temporal statistics evoke

temporal response properties that are not found with stimuli that have different temporal statistics.

In summary, the analyses presented here demonstrate that spatial and temporal stimulus statistics have specific effects on both the spatial and temporal response properties of V1 neurons. Spatial statistics influence spatial response properties but do not affect temporal response properties. In contrast, temporal stimulus statistics influence both spatial and temporal response properties of cells in area V1. Some caution should be exercised when comparing the effects of estimation stimulus properties in terms of absolute difference in prediction correlation. For example, when STRFs are estimated using stimuli without natural temporal statistics, the improvement in prediction correlation from natural spatial statistics may be relatively small. For many neurons, STRFs estimated without natural temporal statistics predict with no better than random accuracy (see Fig. 5*A*). For these neurons, using stimuli with natural spatial statistics to estimate STRFs may provide little or no improvement to predictions. Introducing natural spatial statistics to STRFs that have been estimated using stimuli with natural temporal statistics, in which a greater number of STRFs have better than random predictions (Fig. 5*B*), might lead to a larger increase in prediction correlation. Thus, the absolute difference in prediction correlation reported in each comparison in Figure 9 should not be interpreted strictly as the size of the effect but instead as a measure of whether the influence of a particular stimulus property on STRF structure is significant.

Discussion

Natural vision and neural response properties

This study demonstrates that both the spatial and temporal tuning properties of V1 neurons differ under natural and synthetic stimulus conditions. The dependence of tuning on stimulus statistics is functionally important: a model that incorporates spatial and temporal tuning observed during natural visual stimulation is able to predict novel natural visual responses significantly better than a model incorporating tuning observed during stimulation by synthetic gratings. We identified two major components of the tuning changes. During natural vision, neurons in V1 show increased inhibition at late time lags and complex shifts in the spatial tuning of inhibition. These tuning changes reflect differential activation of nonlinear response properties by the two stimulus classes. Therefore, natural tuning properties cannot be predicted from responses to grating sequences and vice versa.

Temporal inhibition

V1 STRFs estimated using natural visual stimuli show substantial late inhibition that gives rise to transient responses. This inhibition is much weaker or even absent in STRFs estimated using grating sequences. The changes in temporal inhibition depend on differences in temporal stimulus statistics and are consistent with previous observations of nonlinear temporal summation (Tolhurst et al., 1980; Mancini et al., 1990; Reid et al., 1992). Grating sequences are temporally white (up to 72 Hz), whereas natural vision movies are biased toward saccade frequencies (3–4 Hz). This difference in temporal stimulus statistics has a substantial enough effect on tuning properties to dramatically influence the predictive power of STRFs.

This temporal nonlinearity is found in most neurons in our sample, although some neurons have a linear temporal response that does not vary appreciably with stimulus class. Transient responses could reflect depression of subcortical input to V1 (Chance et al., 1998) or local inhibition by intracortical circuitry (Troyer et al., 1998; Müller et al., 1999). In either case, our observations are consistent with the idea that V1 neurons adapt their temporal filtering properties to accommodate changing stimulus statistics and thereby increase information transmission and efficiency (Dan et al., 1996; Richmond et al., 1999; Fairhall et al., 2001; Lesica et al., 2003).

Spatially tuned inhibition

Our STRF estimation procedure reveals both spatially tuned excitation and inhibition in V1 neurons. In the phase-separated Fourier model, excitatory and inhibitory channels each represent patterns at a single orientation, spatial frequency, and phase that increase and decrease spiking responses, respectively; they do not correspond to "on" and "off" subfields described in image domain models (Jones et al., 1987; DeAngelis et al., 1993). Our study demonstrates that the excitatory spatial channels are consistent, regardless of whether natural stimuli or gratings are used. However, inhibitory channels vary substantially with stimulus statistics.

Inhibitory channels likely reflect suppressive influences such as cross-orientation inhibition (Bonds, 1989; Carandini et al., 1997), off-peak suppression (Bauman and Bonds, 1991; Shapley et al., 2003), contrast gain control (Wilson and Humanski 1993; Carandini et al., 1997), and short-term adaptation (Müller et al., 1999). Nonlinear modulation has been shown to vary across the classical and nonclassical receptive field (Gilbert and Wiesel, 1990; Walker et al., 1999; Vinje and Gallant, 2000) and may depend on the relative phases of spatial frequencies composing the stimulus (Mechler et al., 1998). The variability we observed in inhibitory tuning may result from the fact that patterns of suppression are stimulus-dependent. For example, cross-orientation inhibition is observed when two gratings are presented simultaneously (Carandini et al., 1997), but this effect may be absent or attenuated for a rapidly changing sequence of individual gratings. The pattern of inhibition found using natural vision movies reflects the specific influence of the statistical properties of natural visual stimuli.

The differences in temporal and spatial inhibition observed with natural vision movies and grating sequences are attributable to differences in both spatial and temporal stimulus statistics. A more detailed understanding of activity in V1 during natural vision will require identifying the specific differences that give rise to these changes (e.g., stimulus power spectrum, phase spectrum, or bandwidth in space or time). Furthermore, there is a possibility that the statistical properties of the stimulus might interact to produce unanticipated changes in spatial and temporal tuning. For example, nonlinear interactions in the local circuitry of V1 could depend on both the time course and the spatial composition of the stimulus. In that case, natural spatial and temporal statistics could combine synergistically to produce a pattern of tuning not predicted by stimuli with only natural spatial or temporal statistics.

Phase-separated Fourier model

The linearized STRF framework used here is closely related to classical white noise analysis (Marmarelis and Marmarelis, 1978). In white noise analysis, each STRF is composed of a series of kernel functions of increasing polynomial order. The first-order kernel provides the best linear model of response properties in the domain of stimulation, and nonlinear response properties are captured by the second- and higher-order kernels. In principal, such models can predict responses to arbitrary stimuli. Linear white noise analysis has proved valuable at early stages of sensory

processing, in which a first-order kernel explains a large portion of response variance (Jones et al., 1987; DeAngelis et al., 1993). However, this approach has proved less successful when applied to nonlinear neurons [e.g., complex cells (DeAngelis et al., 1995)] or at higher stages of visual processing (Mazer et al., 2000). Specialized algorithms have been developed to estimate second-order kernels (Mancini et al., 1990; Touryan et al., 2002), but these methods require spatially and temporally restricted stimuli to achieve adequate signal-to-noise levels.

Our linearized STRF approach directly addresses the limitations of classical linear kernel estimation in two ways. First, by placing a nonlinear transformation at the input stage of the model, we estimate nonlinear STRFs without a drastic increase in the number of model parameters. This avoids the exponential increase in parameters required to estimate higher-order kernels in the image domain using other methods. Second, our approach permits the use of natural stimuli that drive neurons at all stages of visual processing. Natural stimuli contain high-order statistical properties that the visual system has evolved to exploit (Barlow, 1961; Field, 1987; Simoncelli and Olshausen, 2001). The space of natural stimuli is far from small, but it is much more compact than the space of white noise typically used for STRF estimation. Linearized STRFs estimated in this natural stimulus domain provide a more accurate description of tuning properties during natural vision than STRFs estimated using another stimulus.

By appropriate choice of a nonlinear transformation at the input stage, the linearized STRF framework can account for a wide variety of nonlinear mechanisms. For example, the classical energy model could be incorporated more fully by including a temporal Fourier transform at the input stage. This would provide an explicit model of direction selectivity and might increase predictive power. However, incorporating additional nonlinear parameters also introduces greater susceptibility to estimation noise. Without careful attention to data limitations, a model with more parameters suffers greater risk of overfitting and even reduced prediction accuracy.

Overview of the prediction framework

This study is the first attempt to evaluate a model of V1 in terms of how well it predicts time-varying activity during natural vision. Our approach includes two innovations that impose stricter criteria on measurements of prediction accuracy than many previous studies. These innovations represent critical steps toward developing a concrete understanding of how well existing models describe the actual function of V1.

First, we maintained a complete segregation between estimation and validation data. This approach avoids entirely the possibility of overfitting to the data used for evaluating predictions. Thus, the correlations reported here give an unbiased measure of how the model generalizes to an arbitrary novel stimulus.

Second, predictions from different models were evaluated against a common time-varying response to a natural stimulus. This provides an absolute metric of how well models describe both spatial and temporal response properties to natural stimuli (Gallant, 2003; Olshausen and Field, 2004). To the extent that the visual system is nonlinear, neurons may respond differently to synthetic stimuli than during natural vision. Therefore, any model, whether produced with a synthetic or natural stimulus, must ultimately be tested under natural viewing conditions. This provides a powerful tool for comparing models. In this study, we estimated STRFs with arbitrary data (e.g., natural vision movies or grating sequences) and with different model constraints (e.g., positive spatial response functions) and compared them in an unbiased manner.

Because of the introduction of these criteria into our methods, a direct comparison of correlation values in these results with the typically higher values reported in other studies would not be meaningful. The definitive model of V1 response properties would predict with perfect correlation values of 1 and, at the same time, take natural vision properties into account. Our absolute prediction scores are well below unity. Although phase-separated Fourier STRFs represent a significant advance in the ongoing effort to produce the definitive model of V1, they cannot be interpreted as a realization of that ideal.

The natural vision estimation–prediction approach used here will be useful for developing models of higher visual processing in extrastriate cortical areas. Neurons in these areas respond poorly to synthetic stimuli such as bars or sinusoidal gratings but often give substantial responses to more complex stimuli (Gallant, 2003). The prediction framework also offers a powerful tool for estimating and comparing models of arbitrary complexity.

References

Adelson EH, Bergen JR (1985) Spatiotemporal energy models for the perception of motion. J Opt Soc Am 2:284–299.

Aertsen AMHJ, Johannesma PIM (1981) A comparison of the spectro-temporal sensitivity of auditory neurons to tonal and natural stimuli. Biol Cybern 42:145–156.

Albrecht DG, Geisler WS (1991) Motion selectivity and the contrast-response function of simple cells in the visual cortex. Vis Neurosci 7:531–546.

Barlow HB (1961) Possible principles underlying the transformation of sensory messages. In: Sensory communication (Rosenbluth WA, ed), pp 217–234. Cambridge, MA: MIT.

Bauman LA, Bonds AB (1991) Inhibitory refinement of spatial frequency selectivity in single cells of the cat striate cortex. Vision Res 31:933–944.

Bonds AB (1989) Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex. Vis Neurosci 2:41–55.

Brillinger DJ (1996) Some uses of cumulants in wavelet analysis. J Nonparametric Stat 6:93–114.

Carandini M, Heeger DJ, Movshon JA (1997) Linearity and normalization in simple cells of the macaque primary visual cortex. J Neurosci 17:8621–8644.

Chance FS, Nelson SB, Abbott LF (1998) Synaptic depression and the temporal response characteristics of V1 cells. J Neurosci 18:4785–4799.

Dan Y, Atick JJ, Reid RC (1996) Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. J Neurosci 16:3351–3362.

Daugman JG (1980) Two-dimensional spectral analysis of cortical receptive field profiles. Vision Res 20:847–856.

David SV, Vinje WE, Gallant JL (1999) Natural image reverse correlation in awake behaving primates. Soc Neurosci Abstr 25:767.22.

DeAngelis GC, Ohzawa I, Freeman RD (1993) 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.

DeAngelis GC, Ohzawa I, Freeman RD (1995) Receptive field dynamics in the central visual pathways. Trends Neurosci 18:451–458.

DeBoer E, Kuyper P (1968) Triggered correlation. IEEE Trans Biomed Eng 15:159–179.

DeValois RL, Albrecht DG, Thorell LG (1982) Spatial frequency selectivity of cells in macaque visual cortex. Vision Res 22:545–559.

Efron B, Tibshirani R (1986) Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat Sci 1:54–77.

Fairhall AL, Lewen GD, Bialek W, de Ruyter Van Steveninck RR (2001) Efficiency and ambiguity in an adaptive neural code. Nature 412:787–792.

Field DJ (1987) Relations between the statistics of natural images and the response properties of cortical cells. J Opt Soc Am [A] 4:2379–2394.

Field DJ (1993) Scale-invariance and self-similar "wavelet" transforms: an analysis of natural scenes and mammalian visual systems. In: Wavelets, fractals, and Fourier transforms (Farge M, Hunt JCR, Vascillicos JC, eds), pp 151–193. New York: Clarendon.

Gallant JL (2003) Neural mechanisms of natural scene perception. In: The visual neurosciences (Chalupa LM, Werner JS, eds), pp 1590–1602. Boston: MIT.

Gilbert CD, Wiesel TN (1990) The influence of contextual stimuli on the orientation selectivity of cells in primary visual cortex of the cat. Vision Res 30:1689–1701.

Hubel DH, Wiesel TN (1959) Receptive fields of single neurones in the cat's striate cortex. J Physiol (Lond) 148:574–591.

Jones JP, Stepnoski A, Palmer LA (1987) The two-dimensional spectral structure of simple receptive fields in cat striate cortex. J Neurophysiol 58:1212–1232.

Lesica NA, Boloori AS, Stanley GB (2003) Adaptive encoding in the visual pathway. Network 14:119–135.

Machens CK, Wehr MS, Zador AM (2004) Linearity of cortical receptive fields measured with natural sounds. J Neurosci 24:1089–1100.

Mancini M, Madden BC, Emerson RC (1990) White noise analysis of temporal properties in simple receptive fields of cat cortex. Biol Cybern 63:209–219.

Marmarelis PZ, Marmarelis VZ (1978) Analysis of physiological systems: the white noise approach. New York: Plenum.

Mazer JA, David SV, Gallant JL (2000) Spatiotemporal receptive field estimation during free viewing visual search in macaque striate and extrastriate cortex. Soc Neurosci Abstr 26:53.15.

Mazer JA, Vinje WE, McDermott J, Schiller PH, Gallant JL (2002) Spatial frequency and orientation tuning dynamics in area V1. Proc Natl Acad Sci USA 99:1645–1650.

Mechler F, Victor JD, Purpura KP, Shapley R (1998) Robust temporal coding of contrast by V1 neurons for transient but not steady-state stimuli. J Neurosci 18:6583–6598.

Movshon JA, Thompson ID, Tolhurst DJ (1978) Spatial summation in the receptive fields of simple cells in the cat's striate cortex. J Physiol (Lond) 283:53–77.

Müller JR, Metha AB, Krauskopf J, Lennie P (1999) Rapid adaptation in visual cortex to the structure of images. Science 285:1405–1408.

Olshausen BA, Field DJ (1997) Sparse coding with an overcomplete basis set: a strategy employed by V1? Vision Res 23:3311–3325.

Olshausen BA, Field DJ (2004) What is the other 85% of V1 doing? In: Problems in systems neuroscience (Sejnowski TJ, van Hemmen L, eds). Oxford: Oxford UP, in press.

Pollen DA, Ronner SF (1983) Visual cortical neurons as localized spatial frequency filters. IEEE Trans System Man Cybern 13:907–916.

Reid RC, Victor JD, Shapley RM (1992) Broadband temporal stimuli decrease the integration time of neurons in cat striate cortex. Vis Neurosci 9:39–45.

Richmond BJ, Hertz JA, Gawne TJ (1999) The relation between human V1 neuronal responses and eye movements-like stimulus presentations. Neurocomputing 26:247–254.

Ringach DL, Sapiro G, Shapley R (1997) A subspace reverse-correlation technique for the study of visual neurons. Vision Res 17:2455–2464.

Ringach DL, Hawken MJ, Shapley R (2002) Receptive field structure of neurons in monkey primary visual cortex revealed by stimulation with natural image sequences. J Vision 2:12–24.

Schwartz O, Simoncelli EP (2001) Natural signal statistics and sensory gain control. Nat Neurosci 4:819–825.

Shapley R, Hawken M, Ringach DL (2003) Dynamics of orientation selectivity in the primary visual cortex and the importance of cortical inhibition. Neuron 38:689–699.

Simoncelli EP, Olshausen BA (2001) Statistical properties of natural images. Annu Rev Neurosci 24:1193–1216.

Smyth D, Willmore B, Baker GE, Thompson ID, Tolhurst DJ (2003) The receptive-field organization of simple cells in primary visual cortex of ferrets under natural scene stimulation. J Neurosci 23:4746–4759.

Theunissen FE, Sen K, Doupe AJ (2000) Spectral-temporal receptive fields of non-linear auditory neurons obtained using natural sounds. J Neurosci 20:2315–2331.

Theunissen FE, David SV, Singh NC, Hsu A, Vinje WE, Gallant JL (2001) Estimating spatial temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. Network 12:289–316.

Tolhurst DJ, Walker NS, Thompson ID, S. DA (1980) Non-linearities of temporal summation in neurones in area 17 of the cat. Exp Brain Res 38:431–435.

Tolhurst DJ, Movshon JA, Dean AF (1983) The statistical reliability of signals in single neurons in cat and monkey visual cortex. Vision Res 23:775–785.

Touryan J, Lau B, Dan Y (2002) Isolation of relevant visual features from random stimuli for cortical complex cells. J Neurosci 22:10811–10818.

Troyer TW, Krukowski AE, Priebe NJ, Miller KD (1998) Contrast-invariant orientation tuning in cat visual cortex: thalamocortical input tuning and correlation-based intracortical connectivity. J Neurosci 18:5908–5927.

Vinje WE, Gallant JL (1998) Modeling complex cells in an awake macaque during natural image viewing. In: Advances in neural information processing systems, Vol 10 (Jordan MI, Kearns MJ, Solla SA, eds), pp 236–242. Cambridge, MA: MIT.

Vinje WE, Gallant JL (2000) Sparse coding and decorrelation in primary visual cortex during natural vision. Science 287:1273–1276.

Vinje WE, Gallant JL (2002) Natural stimulation of the non-classical receptive field increases information transmission efficiency in V1. J Neurosci 22:2904–2915.

Walker GA, Ohzawa I, Freeman RD (1999) Asymmetric suppression outside the classical receptive field of the visual cortex. J Neurosci 19:10536–10553.

Weliky M, Fiser J, Hunt RH, Wagner DN (2003) Coding of natural scenes in primary visual cortex. Neuron 37:703–718.

Willmore B, Smyth D (2003) Methods for first-order kernel estimation: simple-cell receptive fields from responses to natural scenes. Network 14:533–577.

Wilson HR, Humanski R (1993) Spatial frequency adaptation and gain control. Vision Res 33:1133–1149.

Woods M, Stringer KM, Dong DW (2001) Visual input statistics during free-viewing of natural time-varying images: a comparison across viewers and scenes. Soc Neurosci Abstr 27:821.28.

Zetzsche C, Barth E, Wegmann B (1993) The importance of intrinsically two-dimensional image features in biological vision and picture coding. In: Digital images and human vision (Watson A, ed), pp 109–138. Boston: MIT.