Network: Computation in Neural Systems
June/September 2005; 16(2/3): 239–260

Taylor & Francis
Taylor & Francis Group

Predicting neuronal responses during natural vision

STEPHEN V. DAVID[1,3] & JACK L. GALLANT[2,3]

[1]*Group in Bioengineering, University of California, Berkeley, USA,* [2]*Department of Psychology, University of California, Berkeley, USA, and* [3]*Helen Wills Neuroscience Institute, University of California, Berkeley, USA*

(Received 17 December 2004; accepted 28 September 2005)

Abstract
A model that fully describes the response properties of visual neurons must be able to predict their activity during natural vision. While many models have been proposed for the visual system, few have ever been tested against this criterion. To address this issue, we have developed a general framework for fitting and validating nonlinear models of visual neurons using natural visual stimuli. Our approach derives from linear spatiotemporal receptive field (STRF) analysis, which has frequently been used to study the visual system. However, prior to the linear filtering stage typical of STRFs, a linearizing transformation is applied to the stimulus to account for nonlinear response properties. We used this approach to compare two models for neurons in primary visual cortex: a nonlinear Fourier power model, which accounts for spatial phase invariant tuning, and a traditional linear model. We characterized prediction accuracy in terms of the total explainable variance, given intrinsic experimental noise. On average, Fourier power STRFs predicted 40% of explainable variance while linear STRFs were able to predict only 21% of explainable variance. The performance of the Fourier power model provides a benchmark for evaluating more sophisticated models in the future.

Keywords: *Visual cortex, receptive field, nonlinear model, prediction*

Introduction

Natural visual stimuli are complex and highly variable, but they do contain many systematic properties (Field 1987; Dong & Atick 1995; Olshausen & Field 1997; Simoncelli & Olshausen 2001). It has been suggested that the visual system has evolved specifically to take advantage of these properties (Barlow 2001). At the same time, functional models of visual neurons have been developed largely using simple synthetic stimuli, such as bars, sine wave gratings and white noise (Hubel & Wiesel 1959; DeValois et al. 1982; Jones et al. 1987). The question of how well models developed using synthetic stimuli generalize to natural vision has received little attention (Vinje & Gallant 1998; David et al. 2004). This is an issue of fundamental importance for understanding visual processing. A model that describes important neural response properties must be able to predict responses to an arbitrary natural stimulus, not just the synthetic stimuli chosen for experimental convenience.

Studies using synthetic stimuli have shown that neurons in visual cortex display a constellation of nonlinear response properties, including nonlinear temporal summation (Tolhurst

Correspondence: Jack L. Gallant, 3210 Tolman Hall #1650, University of California, Berkeley, CA 94720-1650, USA. Tel: 1 510 643 3573. Fax: 1 510 5425293. E-mail: gallant@socrates.berkeley.edu

ISSN: 0954-898X print / ISSN 1361-6536 online
DOI: 10.1080/09548980500464030

et al. 1980), contrast gain control (Carandini et al. 1997) and non-classical receptive field modulation (Gilbert & Wiesel 1990). Any response evoked during natural vision is likely to reflect the combined influence of these nonlinear mechanisms. Moreover, some nonlinearities are likely to have more influence than others during natural vision, and their relative influence may differ from synthetic stimulus conditions (David et al. 2004). For this reason, the influence of nonlinearities and interactions between them must be validated under natural stimulus conditions.

In this study, we develop a general method for validating and comparing nonlinear models in terms of their ability to predict neural responses to natural stimuli. This approach characterizes neural responses in terms of a spatiotemporal receptive field (STRF), a function that maps time-varying visual inputs to neural responses (Marmarelis & Marmarelis 1978; Theunissen et al. 2001; David et al. 2004). Classically, STRFs have been used to implement linear and sometimes second-order models of neural responses (Jones et al. 1987; Rust et al. 2005; Touryan et al. 2005). The framework employed in this study uses *linearization* to implement a much wider range of nonlinear models in the STRF framework. A linearized STRF applies a nonlinear transformation to the stimulus and then applies a linear filter to the transformed stimulus.

We used linearized STRFs to compare one nonlinear model, the Fourier power model, to the classical linear model for neurons in primary visual cortex (V1) (Jones et al. 1987). The Fourier power model can account for spatial phase-invariant tuning, a nonlinear response property observed in V1 complex cells. Because of phase-invariance, linear STRFs are unable to characterize the tuning of complex cells (DeAngelis et al. 1995). Thus, we predicted that Fourier power STRFs would be able to characterize the tuning properties of complex cells, for which linear STRFs would fail.

Performance of the linear and linearized STRFs was measured in terms of their ability to predict neural responses to natural vision movies, stimuli that simulate natural visual stimulation. We took into account the effects of experimental noise in the prediction measurements and report predictions in terms of the fraction of explainable response variance predicted by each model. It was crucial to measure the effect of noise on predictions because it could vary between models, even for the same estimation data. By accounting for noise, we established an absolute metric for evaluating how well a model generalizes to natural visual activity. Thus, our prediction measurements provide a benchmark for comparing predictions by other models in the future.

Methods

Data acquisition

Data were acquired from 72 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). Stimuli were presented on a CRT display using custom software. 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 & Gallant 2002).

During recording, 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 more

than 0.35 degrees from fixation. The location and size of the classical receptive field for each neuron was measured using a procedure described in detail elsewhere (David et al. 2004).

Natural vision movies

In order to simulate natural visual stimulation, *natural vision movies* were presented in the receptive field of the neuron while the animal fixated. These stimuli mimicked the stimulation occurring in an area three to four times the CRF diameter during inspection of a natural scene with voluntary eye movements (Vinje & Gallant 2000; David et al. 2004). A brief segment taken from a natural vision movie appears at the top of Figure 1A. The sequence of

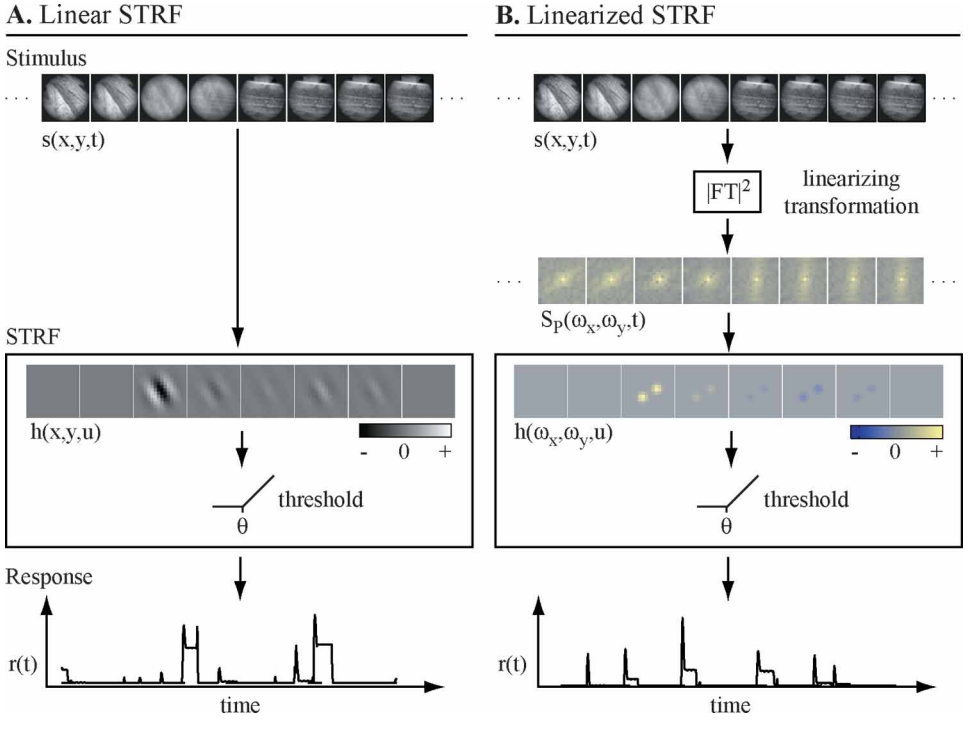

Figure 1. Linear vs. linearized spatiotemporal receptive fields (STRFs) for V1 neurons. (A) Linear, image domain STRF. Visual input is represented as a time-varying sequence of grayscale images, $s(x, y, t)$ (top row). The stimulus is convolved with a linear spatiotemporal filter according to Equation 1 (box). The filter is shown as a series of subpanels corresponding to different time lags, where the gray level indicates the magnitude of relative excitation (white) or inhibition (black) at a given position in the receptive field. The output of the filter is thresholded to produce the instantaneous firing rate, $r(t)$ (bottom row). This model is spatially linear and can account for the responses of simple cells in V1. (B) Linearized, Fourier power STRF. At the input stage, the stimulus is linearized by transforming into the Fourier power domain (second row) according to Equation 2. The transformed stimulus is then convolved with a linear spatiotemporal filter (box). Each filter time slice shows spatial frequency tuning in polar coordinates, where distance from the center corresponds to increasing spatial frequency and angle about the center corresponds to orientation. Yellow indicates relative excitation; blue indicates relative inhibition. The spatial frequency and orientation tuning of this STRF are the same as the linear image domain STRF shown in panel A. The output of the filter is thresholded to produce the instantaneous firing rate, $r(t)$ (bottom row). The Fourier power transformation models the nonlinear phase invariant responses of V1 complex cells.

panels shows each 14 ms frame during the transition between two simulated fixations. Each stimulus sequence was divided into several five-second segments. Different segments were presented on successive fixation trials in random order. To avoid transient trial onset effects, the first 196 ms of data acquired on each trial were discarded before analysis.

STRF models for V1 neurons

The spatiotemporal receptive field (STRF) is a function that maps a visual stimulus to a neural response. A stimulus and response can be described by two time-varying signals: the spatiotemporal stimulus, $s(x_i, t)$, and instantaneous firing rate of a neuron, $r(t)$, sampled at discrete points in space, $x_i \in \{x_1, x_2, \ldots, x_N\}$, and time, $t = 1 \ldots T$.

Sensory neurons employing a rate code have often been described by a rectified linear filter (De Boer 1968; Marmarelis & Marmarelis 1978; Theunissen et al. 2001),

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

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 the response for larger values of s, while negative values indicate inhibitory channels that decrease the response. We modeled the spiking threshold observed in cortical neurons by half-wave rectification, $|x|^{+} = \max(O, x), \ldots,$ with threshold specified by the scalar θ (Albrecht & Geisler 1991). The residual, $\varepsilon(t)$, represents deviations from linear behavior, due either to noise or unmodeled nonlinear response properties.

Image domain model. Area V1 simple cells respond to stimuli having appropriate orientation, spatial frequency and spatial phase (Hubel & Wiesel 1959; DeValois et al. 1982). These cells can be modeled as linear spatiotemporal filters applied to the stimulus (Jones et al. 1987; DeAngelis et al. 1995). According to this *image domain model*, the input channels to the STRF are the luminance at each retinotopic position, $x_i = (x, y)$. A schematic of the image domain model appears in Figure 1A with a simulated STRF in the central box. Each panel of the STRF indicates spatial tuning at successive time lags. Within each panel, the brightness map indicates the relative gain of the receptive field at each point in space. White indicates positive coefficients (excitation), and black indicates negative coefficients (inhibition).

Fourier power model. V1 complex cells have tuning properties similar to those of simple cells, except that they are insensitive to spatial phase (Hubel & Wiesel 1959; DeValois et al. 1982). Luminance at a point in space may be either excitatory or inhibitory to a phase invariant neuron, depending on the luminance at nearby locations. The image domain model requires consistent excitation or inhibition at each retinotopic position and cannot be used to estimate the STRFs of phase invariant complex cells (DeAngelis et al. 1995; Touryan et al. 2005).

Several models have been proposed to account for complex cell responses. The most common of these is the energy model, which assumes that the response of a complex cell is proportional to the spatial Fourier energy falling within its orientation and spatial frequency passband (Pollen & Ronner 1983; Adelson & Bergen 1985). We developed a nonlinear transformation consistent with the energy model that permitted efficient estimation of neuronal

STRFs. In our scheme, a spatial Fourier power transformation is introduced at the input stage (Theunissen et al. 2001). The input to the STRF is defined as the time-varying Fourier power of the stimulus, S_P (ω_x, ω_y, t),

$$S_P(\omega_x, \omega_y, t) = |S(\omega_x, \omega_y, t)|^2. \tag{2}$$

Here, $S(\omega_x, \omega_y, t)$ is the spatial Fourier transform of the stimulus. In this *Fourier power model*, each input channel to the STRF corresponds to a two dimensional spatial frequency, $x_i = (\omega_x, \omega_y)$. The Fourier transformation removes spatial phase but preserves information about stimulus orientation and spatial frequency. Thus, for a neuron that obeys the energy model, the Fourier power model will indicate excitatory tuning within a small range of input channels (Figure 1B). However, the Fourier power model is actually more general than the energy model because it allows multiple spatial frequency channels to excite or inhibit the neuron and it does not constrain spatial tuning to be the same at every time lag.

After the initial linearizing transformation, the remaining stages of the Fourier power model are identical to the image domain model. A schematic of the Fourier power model appears in Figure 1B with a simulated STRF in the central box. Each panel shows a polar plot of spatial frequency (radial component) and orientation (angular component) tuning. In this and other Fourier power STRFs, yellow indicates positive, excitatory coefficients and blue indicates negative, inhibitory coefficients. This STRF has the same orientation and spatial frequency tuning as the image domain STRF in Figure 1A. However, the Fourier power STRF does not preserve the phase tuning that is visible in the image domain STRF.

Procedure for STRF estimation

The STRF estimation procedure used here requires fitting many model parameters. In such cases, accurate predictions can only be obtained if care is taken to avoid bias from overfitting. 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 approximately 90% of the available data (repeated or single trial). The validation set contained the remaining 10% of the data (5–10 seconds per neuron, repeated trials) and was reserved exclusively for evaluating predictions. This strict segregation avoided any possibility of artificially inflating the accuracy with which STRFs could predict natural visual responses in the final validation procedure.

Normalized reverse correlation. We used normalized reverse correlation to estimate linear and linearized STRFs from the estimation data (Theunissen et al. 2001; David et al. 2004). Here, we present the minimum mean-squared error solution for the STRF from Theunissen et al. (2001), modified slightly to facilitate the discussion of prediction noise ceilings.

The STRF specified in Equation 1 can be re-written in linear algebraic form. Suppose we have T samples of stimulus and response, i.e., $t = 1 \ldots T$. To generate the stimulus matrix, S, we define S_u to be the $N \times T$ matrix,

$$S_u = \begin{bmatrix} s(1, 1-u) & s(2, 1-u) & \cdots & s(N, 1-u) \\ s(1, 2-u) & s(2, 2-u) & & s(N, 2-u) \\ \vdots & & \ddots & \vdots \\ s(1, T-u) & s(2, T-u) & \cdots & s(N, T-u) \end{bmatrix} \tag{3}$$

Each row of S_u contains N coefficients describing the stimulus at a lag u before each point in time. The full stimulus matrix is the concatenation of S_u over all relevant time lags,

$$S = [S_0 \quad S_1 \quad \cdots \quad S_U] \tag{4}$$

Thus, each row of S contains all the stimulus information that contributes to the response at a single point in time. S is an $N \times Y$ matrix, where $Y = XU$. We define the STRF, h, to be the $Y \times 1$ vector, where each value indicates the STRF gain for the corresponding spatial channel and time lag. The response and residual are described by $T \times 1$ vectors, r and ε, respectively, which indicate their values at each moment in time. Using these variables, Equation 1 can be re-written,

$$r = |Sh - \theta + \varepsilon|^{+}. \tag{5}$$

Given the stimulus matrix, S, and response vector, r, the minimum mean-squared error estimate of the STRF, h, is the cross correlation between stimulus and response, normalized by the inverse of the stimulus autocorrelation (Marmarelis & Marmarelis 1978; Theunissen et al. 2001; Smyth et al. 2003),

$$h = \frac{1}{T} C_{ss}^{-1} S^T r. \tag{6}$$

Here, C_{ss}^{-1} is the inverse of the stimulus autocorrelation matrix, $C_{ss} = S^T S / T$, and the superscript ^{T} indicates the transpose operation.

This solution ignores the threshold in the model. When stimuli have Gaussian statistics, a static output nonlinearity, such as a threshold, only affects the overall gain in reverse correlation estimates of h (De Boer 1968). Recent work has suggested that, for non-Gaussian stimuli such as natural scenes, an output nonlinearity can bias the shape of h (Sharpee et al. 2004). However, it has not been determined how large that bias is in practice.

Approximate stimulus autocorrelation inverse to reduce estimation error. The autocorrelation matrix, C_{ss}, for natural vision movies is nearly singular. If the STRF is computed simply by applying the inverse of C_{ss} to the stimulus-response cross correlation (Equation 6), noise is amplified excessively for stimulus channels with low power. More accurate STRF estimates can be obtained by an approximation of this inverse that does not amplify noise as severely. We used singular value decomposition (SVD) to construct a pseudo-inverse of the stimulus autocorrelation matrix, C_{approx}^{-1} (Theunissen et al. 2001; David et al. 2004). This procedure ranks stimulus principle components according to their power and sets STRF parameters to zero for stimulus channels where noise is too strong to allow for their accurate estimation.

Pseudo-inverse construction requires selecting a tolerance value, β, to determine the total fraction of stimulus variance to remove from the inverse. We determined the optimal value of β by a jackknife procedure. Twenty jackknife data sets were generated by excluding distinct 5% segments from the estimation data set. For each jackknife set, STRFs were estimated over a range of β (30 values, from 0.1 to 0.00001). For each β and each jackknife, the STRFs were used to predict the responses in the 5% of excluded data. Predictions were concatenated into a prediction of the entire response for each value of β. The value of β that produced the minimum mean-squared error prediction was chosen as the best STRF estimate.

Shrinkage filter to reduce estimation error. After selecting the optimal pseudo-inverse threshold, we applied a shrinkage filter to the STRF estimates to reduce noise further (Brillinger 1996). A shrinkage filter scales down STRF coefficients according to their uncertainty. This process is qualitatively related to Bayesian methods for regularization, such as automatic relevancy determination, that remove model parameters with low probability of being non-zero (Sahani & Linden 2003). The mean STRF, $\bar{h}$, and standard error, $\hat{\sigma}$, were computed across the jackknife estimates (Efron & Tibshirani 1986). The ratio of mean to standard error was taken as the signal to noise ratio for each channel. Each STRF coefficient was scaled according to a shrinkage filter to produce a final STRF estimate (Brillinger 1996):

$$h = \bar{h}\sqrt{|1 - \gamma\hat{\sigma}^2/\bar{h}^2|^+}. \tag{7}$$

The brackets, $|\ldots|^+$, indicate half-wave rectification, and γ is a minimum noise threshold. The optimal value of γ was selected by sampling over a range of 7 values, from 0.8 to 2.0, and finding the value that produced the minimum mean-squared error prediction of responses in the estimation data set.

Nonlinear threshold estimation. The STRF model also contains a nonlinearity that represents response threshold, θ in Equation 5. The threshold for each STRF was selected after estimating h. The output of the linear component of the STRF, Sh, indicated the range of possible values of θ. The threshold was chosen from a dense sampling between the smallest and largest of these values to minimize the mean squared error prediction of responses in the estimation data set.

Data pre-processing. To estimate STRFs the stimulus was first cropped with a square window circumscribing twice the CRF diameter. The window was a fixed multiple of the CRF diameter regardless of the original stimulus size, which ranged from three 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 a spatial tuning profile from responses to sine wave gratings (David et al. 2004). In theory, an analysis using higher spatial resolution could 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. For the Fourier power model, edge artifacts were minimized by applying a Hanning window to each stimulus frame before applying the Fourier transform. Because the window tapered toward zero at the edge of the stimulus, features in the central CRF diameter of the stimulus were emphasized in the Fourier power representation. Fourier power STRFs estimated with the Hanning window were substantially less noisy and had significantly greater predictive power than estimates without a window.

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 ms resolution) were binned at 14 ms, synchronized with the 72 Hz refresh cycle of the CRT. STRFs were calculated across time lags ranging from 0 to 196 ms ($U = 14$ time bins).

Generating and evaluating predictions

Predicted responses were generated according to the procedure outlined in Figure 1. The validation stimuli were cropped and downsampled to match the size used for estimation. Responses were binned at 14 ms in the same manner as the estimation data. For the Fourier power model, stimuli were transformed according to the nonlinearity in Equation 3. Then the estimated STRF was convolved with the stimulus in time, summed over space and thresholded according to Equation 1 to produce the predicted response, r_{pred}. Prediction accuracy was quantified by subtracting the mean from r_{pred} and the observed response, r_{obs}, and computing the correlation coefficient, ρ, between them.

In theory, squared correlation coefficient, ρ^2, indicates the percent variance of the observed response explained by the predicted response. However, experimental noise introduces error in the STRF estimate and in the validation data, both of which bias measurements of ρ^2 to lower values (see Appendix). We measured the effects of both sources of noise on predictions so that we could adjust measurements of ρ^2 to reflect the amount of variance that would be explained in the absence of noise.

Measuring validation noise bounds on predictions. Validation data were acquired by recording responses to repeated presentations of a natural vision movie and averaging them to obtain a PSTH. Because the PSTH was averaged over a finite number of repetitions, it contained noise that could not be predicted by the STRF (see Appendix). We used the model in Equation 15 to quantify the effect of validation noise on predictions,

$$\frac{1}{\rho^2} = \frac{1}{\rho^2_{\text{valmax}}} + \frac{A}{M}. \tag{8}$$

Here, M is the number of repetitions, A is a constant reflecting trial-to-trial variability, and ρ^2_{valmax} is the fraction of variance that would be explained if there was no noise in the validation PSTH.

To measure ρ^2_{valmax}, we divided the validation data into independent subsets (5%, 10%, and 85% of M available trials) and measured $\rho^2$for each. We then found the least-squares fit for Equation 8 to the values of ρ^2 obtained for each subset. By using independent subsets of the validation data to measure ρ^2, we avoided biasing the estimate of ρ^2_{valmax} toward low values. This procedure was repeated for 20 resamplings of the validatation data in order to achieve a reliable estimate of ρ^2_{valmax} (Efron & Tibshirani 1986).

Measuring estimation noise bounds on predictions. Error in the STRF estimate is caused by the finite sampling of stimulus-response samples available for estimation (see Appendix). We used the model in Equation 20 to measure the effects of estimation noise on predictions. For a large number of samples, T, the fraction of response variance explained, ρ^2, is proportional to $1/T$,

$$\frac{1}{\rho^2_{\text{valmax}}} = \frac{1}{\rho^2_{\text{ideal}}} + \frac{B}{T}. \tag{9}$$

Here, B is a constant reflecting noise and the nonlinear residual that must be measured empirically. To determine ρ^2_{ideal}, we estimated separate STRFs using independent subsets of the T available samples (5%, 10%, 25% and 60% of the estimation data), and measured ρ^2 for each. We corrected each measurement of ρ^2 for validation noise and determined

corresponding ρ^2_{valmax} (see above). We then fit Equation 9 to find B and ρ^2_{ideal}. As in the case of validation noise, it was critical that values of ρ^2 used to fit the noise model were measured using independent subsets of the estimation data. This procedure was repeated for 20 resamplings of the estimation data in order to achieve a reliable estimate of ρ^2_{ideal} (Efron & Tibshirani 1986).

Results

Image domain versus Fourier power models

We compared two functional models of primary visual cortex (V1) by evaluating how well each could predict neuronal responses during natural vision. To measure natural visual activity, we recorded the responses of V1 neurons to natural vision movies. These stimuli contain the spatial and temporal pattern of stimulation that occurs in a receptive field during free inspection of a natural scene (see Figure 1 and Vinje & Gallant 2000; David et al. 2004).

We characterized the functional properties of each neuron by estimating the spatiotemporal receptive field (STRF) from the natural vision movie data. Intrinsic to any STRF is a functional model that constrains the response properties that can be captured by the STRF. Previous studies have used the linear, image domain model to estimate STRFs of simple cells in V1 of anesthetized animals (Jones et al. 1987; DeAngelis et al. 1993; Smyth et al. 2003). The image domain model may not appropriate for describing V1 neurons in the awake macaque, where a large fraction of neurons are phase invariant complex cells (Snodderly et al. 2001). Therefore, we developed an alternative nonlinear model, the Fourier power model, which can account for phase invariant tuning (see Figure 1 and Methods).

We compared the image domain model directly the Fourier power model by estimating STRFs using each model from the same natural vision movie data. Figure 2A shows the image domain STRF estimated for one neuron. The STRF reveals the spatial structure of the receptive field, including both inhibitory (white) and excitatory (black) regions. The cell is tuned to horizontal orientations and low spatial frequencies, and has a latency of about 35 ms (time lags indicate the center of each 14 ms time bin). At about 75 ms, spatial tuning changes polarity, indicating that the cell has a biphasic temporal response. Figure 2B shows the Fourier power STRF estimated for the same cell, using the same data as used for the image domain STRF. The Fourier power STRF reveals spatial tuning properties similar to the image domain STRF: a preference for horizontal orientations and low spatial frequencies at the same latency, followed by strong, broadly tuned inhibition.

STRFs estimated using the image domain and Fourier power models provide two alternative descriptions of the same neuron. Because the two models differ primarily in their treatment of spatial phase, any difference in their predictive power must reflect the influence of the spatial phase nonlinearity. Predictions by the STRFs in Figure 2 appear in Figure 3. Each panel shows the predicted response to the validation stimulus (solid line) overlaid on the observed validation response (dashed line). We assessed predictive power by measuring the correlation coefficient between observed and predicted responses to validation data. The image domain STRF predicts with greater accuracy ($\rho = 0.54$, Figure 3A) than the Fourier power STRF ($\rho = 0.38$, Figure 3B, $p < 0.05$, randomized paired t-test).

It is also useful to examine the square of the correlation coefficient, ρ^2, which gives the proportion of variance in the observed response that is explained by the STRF. The percent response variance explained ($100 \times \rho^2$) provides a more natural measure for assessing noise bounds on model performance (see below). For this neuron, the image domain STRF explains 29% of the variance in the validation response, and the Fourier power STRF explains

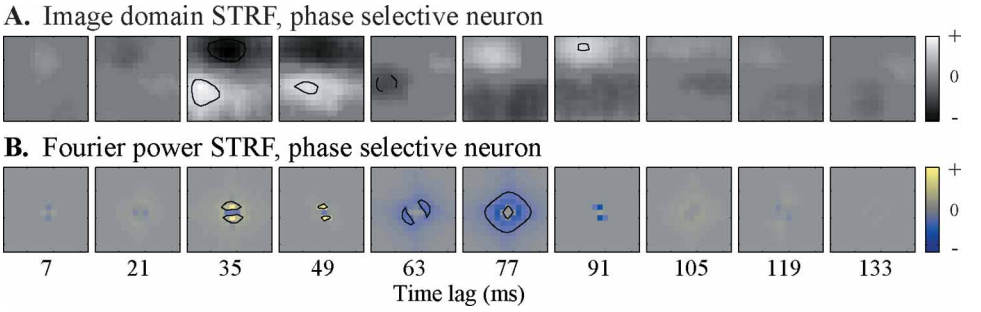

Figure 2. Comparison of image domain and Fourier power STRFs estimated for one V1 neuron stimulated with natural vision movies. (A) Linear image domain STRF. Each subpanel shows spatial tuning at progressively later time lags. Gray level corresponds to the magnitude of relative excitation (white) or inhibition (black) at a given position in the receptive field. Contours indicate points in the STRF that are two standard deviations above or below zero. The STRF shows tuning for horizontal low frequency stimuli (~0.7 cyc/deg) with a latency of 35 ms. At later time lags (63–91 ms) the STRF undergoes a 180-degree spatial phase inversion, suggesting the presence of a late inhibitory component. (B) Linearized Fourier power STRF estimated for the same neuron as in panel A. Each subpanel shows a polar plot of spatial frequency tuning at progressively greater time lags. Distance from the center gives spatial frequency and angle gives orientation. Yellow indicates relative excitation; blue indicates relative inhibition. This STRF shows a preference for horizontal low frequency stimuli (0.7 cycles/deg) at a latency of 35 ms, and strong relative inhibition appears at greater time lags (63–91 ms). The similar tuning for both image domain and Fourier power STRFs suggests that this is a simple cell.

14% of response variance. The superior predictive power of the image domain STRF suggests that this neuron is sensitive to spatial phase and thus that it is a simple cell (Hubel & Wiesel 1959; DeValois et al. 1982).

It may seem surprising that, for the simple cell in Figure 2, the Fourier power STRF is able to predict responses at all. The Fourier power model does not account for spatial phase tuning, a basic property of simple cells. In fact, after the Fourier power transformation, stimuli with different phases are indistinguishable. Thus, one might expect the Fourier power model to confound excitatory and inhibitory (180 degree phase-shifted) stimuli. Instead, the Fourier power STRF has similar orientation and spatial frequency tuning to the image domain STRF, which does account for phase tuning. The similar tuning of both STRFs might reflect the influence of the spiking threshold present in cortical neurons. The threshold is effectively an expansive nonlinearity, which increases the gain for excitatory stimuli, relative to inhibitory stimuli (Geisler & Albrecht 1992). The amplification of excitatory spatial phases would prevent complete phase cancellation in the Fourier power STRF and permit it to account for tuning properties of both simple and complex cells.

Figure 4 shows image domain and Fourier power STRFs for a second V1 neuron. In this case, the STRFs estimated using the two models are quite different. The image domain STRF (Figure 4A) lacks any clear orientation or spatial frequency tuning, though activity increases weakly at time lags of about 50–100 ms. In contrast, the Fourier power STRF (Figure 4B) shows clear tuning for vertical orientations and medium spatial frequencies, with a peak response latency of about 65 ms. Thus, the Fourier power model reveals spatiotemporal tuning properties that the image domain model fails to recover, suggesting that this is a complex cell insensitive to spatial phase.

These two STRFs also have substantially different predictive power (Figure 5). Predictions of the image domain STRF are low ($\rho = 0.18$, 3% of response variance explained, Figure 5A), while the Fourier power STRF predictions are quite accurate ($\rho = 0.61$, 37%

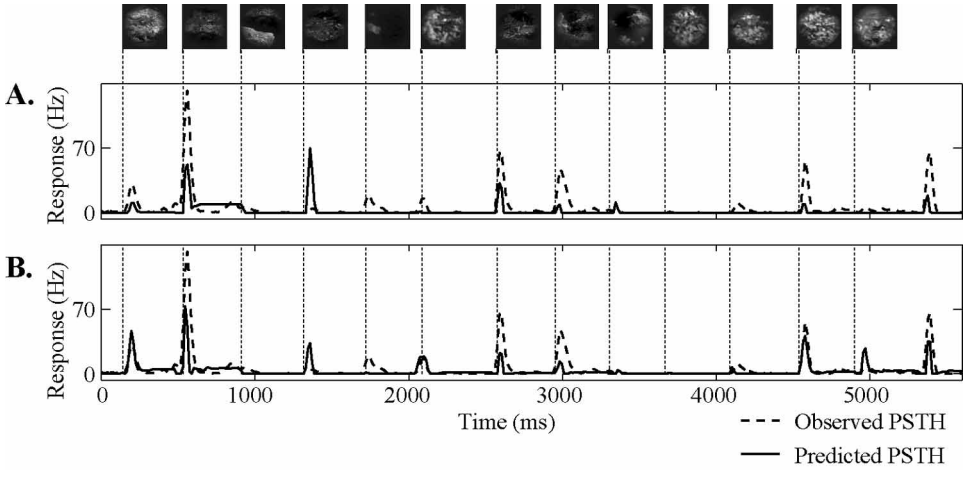

Figure 3. Predictions of image domain and Fourier power STRFs. Neuronal data are the same as in Figure 2. (A) The average response of the neuron to 10 repeated presentations of the validation stimulus (dotted line) is overlaid with the prediction of the image domain STRF (solid line). Each simulated saccade (vertical lines) evokes a brief transient response from this neuron, and the predicted PSTH matches these transients well ($\rho = 0.54$, 29% of response variance explained). (B) Same neuronal data as shown in panel A (dotted line), along with the prediction of the Fourier power STRF (solid line). The Fourier power STRF does not predict as well as does the image domain STRF ($\rho = 0.38$, 14% of response variance explained), confirming that this is a simple cell.

of response variance explained, Figure 5B). Thus, the Fourier power model, which accounts for phase invariance, provides a better prediction of this neuron's responses to natural vision movies.

The examples presented above suggest that the image domain model provides a less general description of V1 neurons than the Fourier power model: the image domain model is only applicable to V1 simple cells, while the Fourier power model can be applied to both simple and complex cells. To test this hypothesis, we compared the predictive power of the image domain and Fourier power STRFs estimated for all 72 V1 neurons in our sample (Figure 6). In 31 of these cells (43%, indicated by the black filled circles) the Fourier power STRF predicts

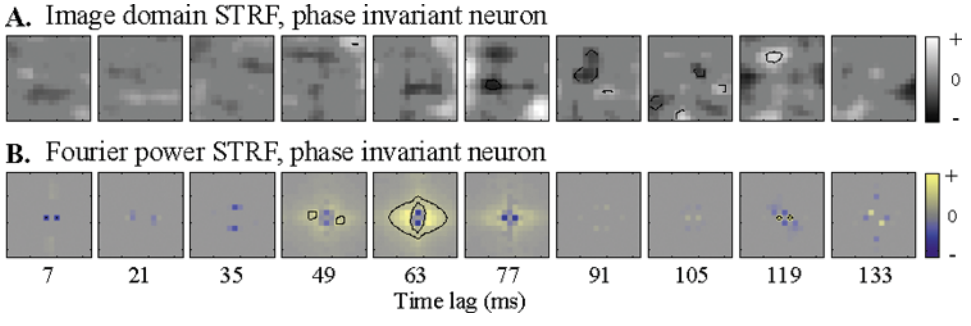

Figure 4. Comparison of image domain and Fourier power STRFs estimated for a second V1 neuron stimulated with natural vision movies. (A) Image domain STRF, format same as in Figure 2A. No spatial tuning is evident. (B) Fourier power STRF estimated for the same neuron as in panel A. This STRF shows clear tuning for vertical stimuli (1.8 cycles/deg) with a latency of 49 ms. The absence of spatial tuning in the image domain STRF and the presence of tuning in the Fourier power STRF suggests that this is a phase-invariant complex cell.

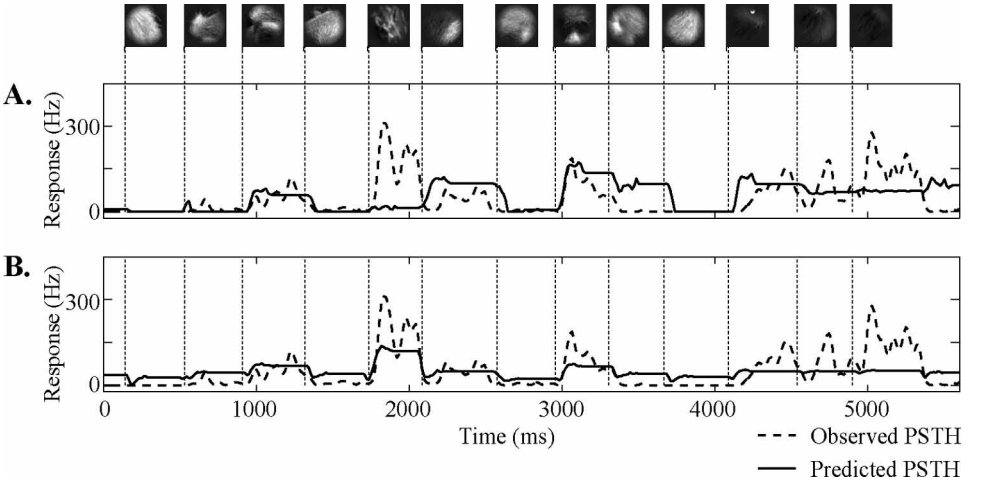

Figure 5. Predictions of image domain and Fourier power STRFs. Neuronal data and STRFs are the same as those shown in Figure 4. (A) Response observed in the validation data set (dotted line) and prediction of the image domain STRF (solid line), format same as Figure 3. The image domain STRF provides poor a prediction of the observed response ($\rho = 0.18$, 3% of response variance explained). (B) Same neuronal data as shown in panel A (dotted line) and prediction of the Fourier power STRF (solid line). The Fourier power STRF predicts the validation response much better than the image domain STRF ($\rho = 0.61$, 37% of response variance explained). The failure of the image domain STRF to predict validation responses and success of the Fourier power model confirms that this is a phase invariant complex cell.

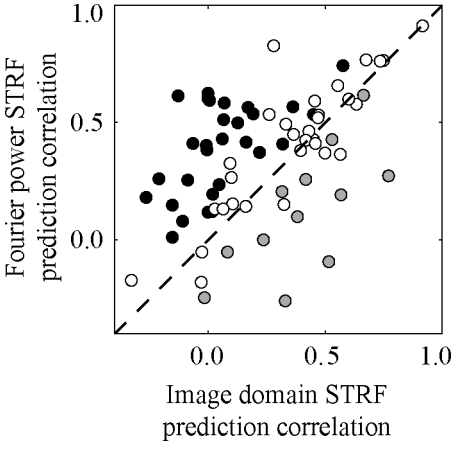

Figure 6. Comparison of image domain and Fourier power STRF predictions. The horizontal axis gives the correlation, ρ, between the response observed in the natural vision movie validation data and the prediction by the image domain STRF. The vertical axis gives the correlation between the observed validation response and the Fourier power STRF prediction. Gray points falling below the line denote cells for which the image domain STRF predicts responses significantly better than the Fourier power STRF; black points above the line denote cells for which the image domain STRF is significantly worse than the Fourier power STRF ($p < 0.05$ in both cases, randomized paired t-test). Across the sample of 72 neurons, Fourier power STRFs predict responses better than image domain STRFs (mean Fourier power prediction: $\rho = 0.37$, 19% of response variance explained; mean image domain prediction: $\rho = 0.25$, 13% of response variance explained; $p < 0.001$, randomized t-test).

responses significantly better than the image domain STRF ($p < 0.05$, randomized paired t-test). Only in 11 cases (15%, gray circles) does the image domain STRF predict significantly better. Across the sample, the mean correlation between the observed responses to natural vision movies and the image domain STRF prediction is only 0.25 (13% of response variance explained), while the mean correlation between observed responses and the Fourier power STRF prediction is 0.37 (19% of response variance explained). This difference is significant ($p < 0.001$, randomized paired t-test). (Note that for population data mean percent variance explained is not equal to the square of mean correlation, $\langle \rho^2 \rangle \neq \langle \rho \rangle^2$.)

Effects of finite sampling on predictions

The accuracy of STRF predictions is limited by two factors: failure to model nonlinear response properties of the neuron and noise due to finite sampling of stimulus-response data. The former reflects fundamental limitations of the model but the latter can be ameliorated with a larger data set. We measured the effect of finite sampling on predictions in order to accurately determine ρ^2_{ideal}, the fraction of V1 response variance that would be predicted in the absence of noise.

Validation noise. Finite sampling of validation data introduces noise into the observed responses used to evaluate predictions. In theory, the PSTH constructed from validation data reflects the underlying response modulation function of each neuron. However, this function cannot be measured directly from extracellular recordings. Instead it must be estimated by averaging noisy spike trains over repeated trials. Because estimated STRFs cannot predict the noise component of validation data, noise creates a downward bias in predictions. However, this bias is reduced if more repeated trials are used to construct the validation data set and the noise is averaged out.

Figure 7A shows the effect of varying the number of validation data trials on predictions by the Fourier power STRF shown in Figure 2B. When predictions are evaluated with only a single trial of validation data, the STRF explains just 10% of response variance. As more trials are included in the validation set, prediction accuracy increases. After correcting for the effects sampling limitations with the validation noise model in Equation 8, the STRF can explain 18% of response variance (circle at right).

Estimation noise. Prediction accuracy was also limited by noise from finite sampling of estimation data. Natural visual stimuli are complex and highly variable, and our neurophysiological experiments include at most a few thousand stimuli. Moreover, neurons respond unreliably, adding noise to the data that is available. With only a relatively small number of stimulus-response samples, estimated STRFs will necessarily contain some error.

Figure 7B shows the fraction of response variance explained by one STRF, estimated from variable fractions of the available data (squared correlation for each prediction, ρ^2_{valmax}, has been adjusted to correct for validation noise). Predictions improve as more data are included; when all 11,000 available samples are used, 18% of the response variance is explained. The solid curve shows the fit by the estimation noise model in Equation 9 (gray band in Figure 7B indicates 95% confidence intervals). Because predictions are still improving when all available data are used, there is uncertainty in the fit for large numbers of samples. Our best estimate is that this STRF would explain 25% of the variance in the validation set, were

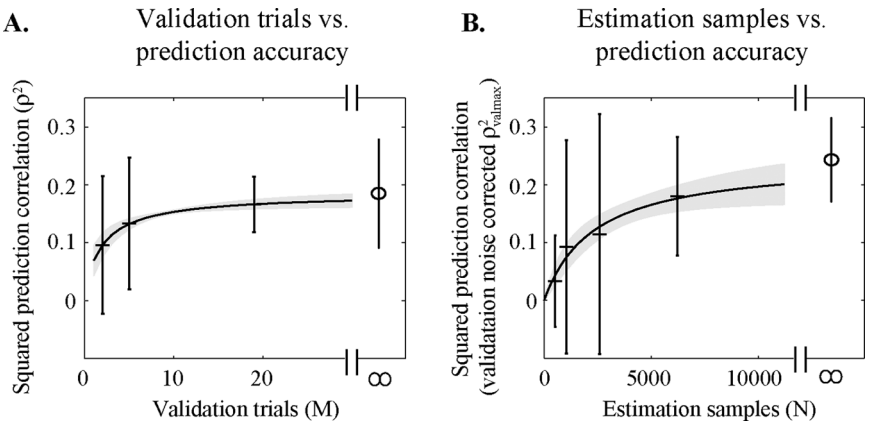

Figure 7. Effect of finite sampling on prediction accuracy for a single neuron. (A) Effects of finite validation data sampling. Squared prediction correlation, ρ^2, for a single Fourier power STRF (from Figure 2B) is plotted as a function of the number of repeated trials included in the validation data. Error bars show 95% confidence intervals on predictions. The validation noise model was fit to these data (Equation 8, curve, 95% confidence intervals indicated by shading) to estimate the value of ρ^2 that would be achieved with infinite validation trials (circle at right). For this neuron, the STRF would account for 18% of response variance if the validation data contained no noise. (B) Effects of finite estimation data sampling on predictions by the same STRF. ρ^2_{valmax} is plotted as a function of the number of data samples used to estimate the STRF. (Measured values of ρ^2 have been adjusted to ρ^2_{valmax} according to the validation noise model to correct for the effects of finite validation data sampling.) These data were fit by the estimation noise model in Equation 9 (solid curve, 95% confidence intervals indicated by shading). After correcting for finite sampling of both estimation and validation data, STRF predictions would account for 25% of response variance (circle at right).

infinite estimation data and validation data available (i.e., $\rho^2_{\text{ideal}} = 0.25$, circle at far right of Figure 7B).

Figure 8 summarizes the effect of correcting for sampling limitations on predictions by image domain and Fourier power STRFs. The bars labeled ρ^2_{ideal} indicate the average expected fraction of response variance predicted after applying the validation and estimation noise models. This analysis includes only the 49 neurons with more than 2000 samples in their estimation data sets. Smaller data sets were not reliably fit by the estimation noise model.

After fully correcting for sampling limitations, the average image domain STRF can explain 21% of variance in responses to natural vision movies. Fourier power STRFs perform substantially better, explaining an average of 40% of response variance. However, even with infinite data, the Fourier power STRF leaves an average of 60% of response variance unexplained. This remaining component of the response reflects the contribution of unmodeled nonlinear response properties that are functionally important during natural vision.

Discussion

This study compared two functional models of primary visual cortex by their ability to predict responses to natural visual stimuli. The linear image domain model predicts natural vision movie responses relatively poorly; the mean correlation between observed and predicted responses is 0.25, explaining 13% of the total response variance (21% of explainable variance). The linearized Fourier power model predicts responses significantly better; the mean correlation between observed and predicted responses is 0.37, accounting for 19% of response

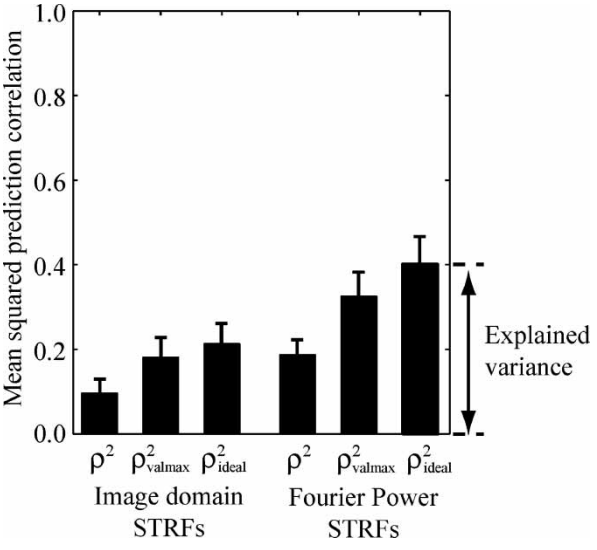

Figure 8. Summary of the effect of finite sampling on predictions. Each bar indicates mean squared prediction correlation for the 49 neurons with greater than 2000 stimulus-response samples (error bars indicate standard error). Data for image domain STRFs are shown at left, and for Fourier power STRFs at right. ρ^2 indicates the mean squared prediction correlation actually measured for the STRFs. ρ^2_{valmax} indicates mean prediction after correcting for finite sampling of validation data. ρ^2_{ideal} indicates the mean prediction after correcting for finite sampling of both estimation and validation data. Fourier power STRFs perform consistently better than image domain STRFs. After correcting for sampling limitations, Fourier power STRFs can account for an average of 40% of the response variance in V1. The remaining portion of the response results from nonlinear response properties ('unexplained variance') not included in the Fourier power model.

variance (40% of explainable variance). The remaining, unexplained variance must reflect nonlinear response properties. The performance of the Fourier power model presents a bar to be surpassed by more sophisticated models that account for other important nonlinearities.

Nonlinear neurons and natural vision

Even after correcting for sampling limitations, the prediction accuracy of the Fourier power model might appear surprisingly low to readers familiar with other modeling studies. This is because we have adopted strict, principled criteria for evaluating prediction accuracy. First, we used stimuli that simulate natural vision to evaluate predictions. Most previous studies have used simple, synthetic stimuli to evaluate model predictions. Second, we used a strict validation procedure, splitting the data into two independent subsets: one used only to fit model parameters and the other used only to evaluate predictions. The use of separate estimation and validation data sets avoids bias in measurements of prediction accuracy that results from overfitting to noise in the estimation data. This validation procedure is still rarely used to fit models to neurophysiological data. As a consequence, many published models actually perform more poorly than is claimed.

The fact that the Fourier power model can only predict 40% of response variance suggests that natural visual processing is substantially influenced by nonlinear mechanisms not included in the Fourier power model. There are many known nonlinear response properties that could affect responses. The phase-separated Fourier model, which is presented elsewhere (David et al. 2004), can account for different orientation and spatial frequency tuning at each spatial phase. Thus, it can describe both phase-tuned and phase-invariant neurons,

as well as neurons with intermediate phase tuning. This model should predict as well or better than the Fourier power model. Other nonlinear mechanisms that would likely improve predictions include nonlinear temporal summation (Tolhurst et al. 1980; Reid et al. 1992), contrast gain control (Wilson & Humanski 1993; Carandini et al. 1997) and spatially localized modulation by the non-classical receptive field (Walker et al. 1999; Vinje & Gallant 2000). Some nonlinear response properties, such as contrast gain control, are approximated by negative coefficients in the Fourier power STRF, but models that capture divisive and other nonlinear modulation explicitly may prove more accurate.

Exactly which unmodeled nonlinear mechanisms would most improve model performance remains an open question, but one that can be addressed using the framework developed here by comparing alternative models in terms of their ability to predict natural visual responses. At this time the Fourier power model provides the best predictions of natural visual responses.

Functional properties of V1 neurons

Comparison of predictive power between models can be viewed in terms of classical hypothesis testing. According to this view the image domain and Fourier power models instantiate explicit hypotheses about neuronal processing during natural vision. These alternative hypotheses are tested by evaluating each model's ability to predict natural visual responses. The prediction framework is actually more powerful than the classical method of assessing statistical significance because it provides simultaneous measures of significance and importance.

The Fourier power model provides an effective description of phase invariant complex cells during natural vision. Its predictive power is equivalent to or better than that of the image domain model for 85% of the cells in our sample. Given that the image domain model is designed to characterize simple cells and the Fourier power model is designed to characterize complex cells, one could interpret this result to mean that 85% of neurons in V1 of awake macaques are complex cells. This estimate is compatible with other findings (Snodderly et al. 2001). However, we did not observe a clear bimodal distribution of model predictions, which has been observed with other metrics, such as the F1/F0 ratio (DeValois et al. 1982). The distribution of model predictions suggests that phase tuning in V1 can vary continuously between what would be expected from classical simple and complex cells (Mechler & Ringach 2002). However, predictions for some neurons could be affected by uncontrolled microsaccades within the 0.35 degree fixation window, which could introduce apparent phase invariance when none or little is present (Snodderly et al. 2001).

Linearized STRFs and other functional models

The strategy of linearization addresses a critical limitation of classical linear STRFs. A linearizing transformation provides a means of incorporating a wide range of nonlinear response properties into STRFs. The superior predictions by the Fourier power STRFs validate linearization as a tool for describing nonlinear response properties that builds on the classical STRF framework.

Several other nonlinear models have been developed for studying the response properties of visual neurons. These include higher order Volterra series (Mancini et al. 1990; Rust et al. 2005; Touryan et al. 2005) and non-parametric methods, such as artificial neural networks (Lehky et al. 1992; Lau et al. 2002; Prenger et al. 2004). Many of these models can be formulated in terms of linearization. For example, Volterra series (including spike-triggered covariance and other second-order models) can be viewed as linearized STRFs, where the linearizing transformation is the polynomial expansion of the stimulus. In principle, any

nonlinear transformation can be used to map the stimulus into a space where the relationship between stimulus and response is more linear. Stimulus information can be discarded (smoothing), made sparse, made redundant, or transformed in any other way suggested by a model of visual processing.

An advantage of the Fourier power model is that it can be specified by fewer parameters than the linear model. The effectiveness of any modeling framework—linear or nonlinear—is limited by the amount of available data, and models with more parameters require more data to be fit accurately. An effective linearizing transformation should account for nonlinear response properties without a substantial increase in parameters. Studies in the auditory system have employed low-dimensional linearizing transformations successfully (Aertsen & Johannesma 1981; Theunissen et al. 2001). However, the potential of linearized STRF estimation remains largely unexplored as a means of comparing how well different functional models describe activity during natural vision.

Selection of an appropriate linearizing transformation requires prior knowledge or a hypothesis about the system under study. Nonparametric methods, such as neural networks, offer advantages over linearized STRFs in this regard (Lehky et al. 1992; Lau et al. 2002; Prenger et al. 2004). Nonparametric methods do not require explicit models of the nonlinear stimulus-response relationship. Instead they can recover any arbitrary nonlinear function employed by the system (given enough data). Such methods may provide a complementary tool for discovering novel nonlinear response properties that can then be validated by means of the linearized STRF framework (Prenger et al. 2004).

Sampling limitations and prediction accuracy

The absolute explanatory power of a model can be assessed only after accounting for sampling limitations and experimental noise. Any remaining unexplained variance must then be due to unmodeled response properties. The best model is simply the one that explains the largest portion of remaining response variance. Although accurate measurement of experimental noise is crucial for assessing and comparing alternative models objectively, this issue has received little attention thus far (but see Sahani & Linden 2003; Hsu et al. 2004).

Other metrics than the correlation coefficient have been proposed for measuring prediction accuracy. These may be more appropriate for neural response data, which do not have Gaussian statistics (Hsu et al. 2004). In practice, these metrics tend to be strongly correlated with the correlation coefficient, but they may provide additional information about model performance. The effects of finite sampling discussed here are also relevant to these other metrics.

Acknowledgments

This work was supported by grants to JLG from the NEI and NIMH. SVD was partially supported by an NSF fellowship. We thank William Vinje for substantial contributions to data acquisition and Benjamin Hayden and Mauro Merolle for helpful comments on the manuscript.

References

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

Aertsen AM, Johannesma PI. 1981. The spectro-temporal receptive field. A functional characteristic of auditory neurons. Biol Cybern 42:133–143.

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. 2001. Redundancy reduction revisited. Network: Comp Neural Systems 12:241–253.

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.

David SV, Vinje WE, Gallant JL. 2004. Natural stimulus statistics alter the receptive field structure of v1 neurons. J Neurosci 24:6991–7006.

De Boer E. 1968. Reverse correlation. I. A heuristic introduction to the technique of triggered correlation with the application to the analysis of compound systems. Proc K Ned Akad Wet C 71:472–486.

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 Neurophys 69:1118–1135.

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

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

Dong DW, Atick JJ. 1995. Statistics of natural time-varying images. Network: Comp Neural Systems 6:345–358.

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

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.

Geisler WS, Albrecht DG. 1992. Cortical neurons: Isolation of contrast gain control. Vis Res 8:1409–1410.

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

Hsu A, Borst A, Theunissen FE. 2004. Quantifying variability in neural responses and its application for the validation of model predictions. Network: Comp Neural Systems 15:91–109.

Hubel DH, Wiesel TN. 1959. Receptive fields of single neurones in the cat's striate cortex. J Physiol (London) 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(4):1212–32.

Lau B, Stanley GB, Dan Y. 2002. Computational subunits of visual cortical neurons revealed by artificial neural networks. Proc Natl Acad Sci USA 99:8974–8979.

Lehky SR, Sejnowski TJ, Desimone R. 1992. Predicting responses of nonlinear neurons in monkey striate cortex to complex patterns. J Neurosci 9:3566–3581.

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

Marmarelis PZ, Marmarelis VZ. 1978. Analysis of physiological systems: The white noise approach. New York, NY: Plenum.

Mechler F, Ringach DL. 2002. On the classification of simple and complex cells. Vis Res 42:1017–1033.

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

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

Prenger RJ, Wu MC-K, David SV, Gallant JL. 2004. Nonlinear V1 responses to natural scenes revealed by neural network analysis. Neural Networks: 663–679.

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

Rust NC, Schwartz O, Movshon JA, Simoncelli EP. 2005. Spatiotemporal elements of macaque v1 receptive fields. Neuron 46:945–956.

Sahani M, Linden JF. 2003. How linear are auditory cortical responses? In: Becker S, Thrun S, Obermayer K (editors). Advances in neural information processing systems 15. Cambridge, MA: MIT Press. pp 301–308.

Sharpee T, Rust NC, Bialek W. 2004. Analyzing neural responses to natural signals: maximally informative dimensions. Neural Comput 16:223–250.

Simoncelli EP, Olshausen BA. 2001. Statistical properties of natural images. Ann Rev Neurosci 25:1193–1216.

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.

Snodderly DM, Kagan I, Gur M. 2001. Selective activation of visual cortex neurons by fixational eye movements: Implications for neural coding. Vis Neurosci 18:259–277.

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: Comp Neural Systems 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.

Touryan J, Felsen G, Dan Y. 2005. Spatial structure of complex cell receptive fields measured with natural images. Neuron 45: 781–91.

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

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.

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

Appendix: Noise bounds on predictions

Interpreting squared correlation

To measure the accuracy of STRF predictions, we computed the correlation coefficient (Pearson's r) between the predicted response, r_{pred}, and observed response, r_{obs}, from the validation data set,

$$\rho = \sqrt{\frac{\langle r_{\text{pred}} r_{\text{obs}} \rangle}{\langle r^2_{\text{pred}} \rangle \langle r^2_{\text{obs}} \rangle}}. \tag{10}$$

(Here we refer to the correlation coefficient as ρ to avoid confusion with response vectors.) One convenient feature of the correlation coefficient is that it provides an intuitive assessment of the portion of the observed response predicted by the STRF. According to the model in Equation 5, a neural response can be decomposed into its linear and residual components. In ideal conditions without any noise, the observed response to a validation stimulus, S_{val}, can be written,

$$\begin{aligned} r_{\text{obs}} &= S_{\text{val}} h + \varepsilon_{\text{nl}} \\ &= r_{\text{lin}} + \varepsilon_{\text{nl}} \end{aligned} \tag{11}$$

The residual, ε_{nl}, results exclusively from nonlinear response properties. In these noise-free conditions, the STRF can be estimated perfectly, and the predicted response, r_{pred}, is identical to the linear component of the observed response. Equation 10 then simplifies to,

$$\rho^2_{\text{ideal}} = \frac{\langle r^2_{\text{lin}} \rangle}{\langle r^2_{\text{lin}} \rangle + \langle \varepsilon^2_{\text{nl}} \rangle}. \tag{12}$$

The value of ρ^2_{ideal} is the ratio of variance in the linear response to the total variance in the observed response, or the fraction of response variance explained by the STRF. The subscript 'ideal' reflects that, in this case, there is no noise in the STRF estimate or in the observed neural response.

In practice, however, two sources of noise can adversely affect measurements of ρ^2. The observed response contains noise because of finite sampling of validation data, and the STRF itself contains noise because of finite sampling of estimation data. In order to obtain accurate

estimates of ρ^2_{ideal} it is necessary to account for both of these effects. These two sources of noise originate from independent data sets. Thus their effects can be addressed separately.

Error from finite validation data

In experimental data, the residual response, ε from Equation 5, is the sum of a nonlinear response, ε_{nl}, and noise, $\varepsilon_{\text{noise}}$, $\varepsilon = \varepsilon_{\text{nl}} + \varepsilon_{\text{noise}}$. These two components represent distinct portions of the validation data. Improving the accuracy of the model (by appropriate linearization) can reduce ε_{nl} to 0, but $\varepsilon_{\text{noise}}$ is independent of the stimulus and cannot be predicted. Thus, non-zero values of $\varepsilon_{\text{noise}}$ necessarily bias measurements of ρ^2 toward values lower than ρ^2_{ideal},

$$\rho^2 = \frac{\langle r^2_{\text{lin}} \rangle}{\langle r^2_{\text{lin}} \rangle + \langle \varepsilon^2_{\text{nl}} \rangle + \langle \varepsilon^2_{\text{noise}} \rangle}. \tag{13}$$

(Here we present the theoretical situation with a noise-free STRF estimate that predicts the linear component of the response accurately. Estimation noise is discussed in the following section.)

The effect of noise on measurements of ρ^2 can be reduced by averaging validation responses over repeated presentations of the same stimulus. The linear and nonlinear components of the response are the same on each repetition and are not affected by averaging. Noise, on the other hand, is independent of the stimulus. If responses are averaged over M trials, then noise variance is reduced proportional to $1/M$,

$$\rho^2 = \frac{\langle r^2_{\text{lin}} \rangle}{\langle r^2_{\text{lin}} \rangle + \langle \varepsilon^2_{\text{nl}} \rangle + \frac{1}{M} \langle \varepsilon^2_{\text{noise}} \rangle}. \tag{14}$$

We can rearrange Equation 14 and substitute in Equation 12 to obtain ρ^2_{ideal} in terms of the measured correlation,

$$\frac{1}{\rho^2} = \frac{1}{\rho^2_{\text{ideal}}} + \frac{1}{M} \left(\frac{\langle \varepsilon^2_{\text{noise}} \rangle}{\langle r^2_{\text{lin}} \rangle} \right). \tag{15}$$

Thus, as M increases, the measured ρ^2 will asymptote to the ideal value.

Figure 9A illustrates the relationship between ρ^2 measured with finite validation data and ρ^2_{ideal} for simulated neural data. The estimation and validation stimuli consisted of 10 channels of Gaussian white noise with mean 0 and variance 1. The linear component of the neural response depended only on the first channel: $h(1) = 0.5$, and $h(2) \ldots h(8) = 0$. This specified a linear response variance of $\langle r^2_{\text{lin}} \rangle = 0.25$. Observed responses also contained additive Gaussian white noise with variance, $\langle \varepsilon^2_{\text{noise}} \rangle = 0.7$, and a nonlinear component with variance, $\langle \varepsilon^2_{\text{nl}} \rangle = 0.04$. These parameters specify the prediction correlation in the absence of noise to be $\rho^2_{\text{ideal}} = 0.86$. In Figure 9A, values of ρ^2 are measured for predictions against validation data averaged over a variable number of independent trials, M. Error bars indicate one standard error around the mean for predictions by STRFs estimated with independent sets of $T = 12$, 24, or 180 samples (bottom to top). As M increases, validation noise is reduced, and predictions asymptote toward a maximum. The maximum is specified by the unmodeled nonlinear component and by error resulting from finite estimation data (see below). When estimation noise is small (e.g., $T = 180$), the measured value of ρ^2 approaches the theoretical maximum for the linear model, ρ^2_{ideal} (dashed line). Measurements of ρ^2 for variable numbers of validation trials are well matched to values predicted by Equation 15 (solid lines).

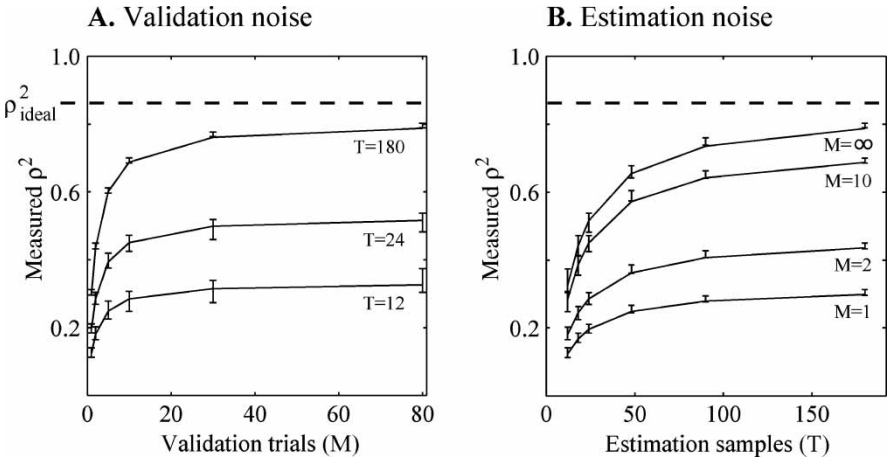

Figure 9. Effects of noise on measurements of prediction accuracy. Data are shown for a simulated system with additive linear, nonlinear and noise components. Linear STRFs were estimated using a variable number of samples, T, and prediction accuracy was evaluated for each STRF using validation sets averaged over variable numbers of repeated trials, M. (A) Effects of finite validation data sampling on prediction accuracy. Each point shows ρ^2 as a function of the number of repeated trials, M, included in the validation data. Curves indicate prediction by the noise model in Equation 15 for (bottom to top) $T = 12$, 24 and 120 estimation samples. For $T = 120$ and large M, ρ^2 nearly reaches the theoretical maximum, ρ^2_{ideal} (dashed line). (B) Effects of finite estimation data sampling on prediction accuracy. Each point shows squared prediction correlation, ρ^2, as a function of T. Solid curves show predictions by the noise model in Equation 20. Each curve corresponds to a different number of repeated validation trials (from bottom to top, $M = 1$, 2, 10 and infinity). As more data is included in the STRF estimate, prediction accuracy increases asymptotically toward the maximum achievable value. For infinite validation trials (top curve), ρ^2 asymptotes to the theoretical maximum (dashed line).

Error from finite estimation data

Finite sampling of estimation data introduces noise into STRF estimates. Because the residual component of the response, ε, is not correlated with the stimulus, it should not, in theory, affect STRF estimates. With finite sampling, however, the residual response does not cancel out exactly in the cross correlation between stimulus and response (Equation 6). Thus the STRF that best describes the stimulus-response relationship can only be approximated.

The estimated STRF, h_{est}, is a sum of the true STRF, h, and the estimation error, h_{err}: $h_{\text{est}} = h + h_{\text{err}}$. The magnitude of h_{err} depends on the residual response, ε, and inversely on T, the number of stimulus-response samples available for estimation,

$$h_{\text{err}} = \frac{1}{T} C_{ss}^{-1} S^T \varepsilon. \tag{16}$$

Because of its dependence on T, the approximation becomes asymptotically more accurate as the size of the estimation data set increases. Estimation error propagates directly to predicted response,

$$\begin{aligned} r_{\text{pred}} &= S_{\text{val}}(h + h_{\text{err}}) \\ &= r_{\text{lin}} + \varepsilon_{\text{pred}} \end{aligned} \tag{17}$$

and introduces a prediction error, $\varepsilon_{\text{pred}}$. We can substitute the STRF error into the prediction and find the variance of the prediction error,

$$\left\langle \varepsilon^2_{\text{pred}} \right\rangle = \frac{1}{T} \langle \varepsilon^2 \rangle. \tag{18}$$

(This assumes that the estimation and validation stimuli have the same autocorrelation, which is the case when natural vision movies are used both for estimation and validation.)

The prediction error biases measurements of correlation to lower values. If we substitute the noisy prediction into the definition of ρ^2, we find,

$$\rho^2 = \left(\frac{\langle r_{\text{lin}}^2 \rangle}{\langle r_{\text{lin}}^2 \rangle + \langle \varepsilon_{\text{pred}}^2 \rangle} \right) \rho_{\text{ideal}}^2. \tag{19}$$

(Here we assume that there is no validation noise. In practice, the corrections for noise can be applied sequentially. Values of ρ^2 corrected for validation noise by the model in Equation 15 can be substituted into Equation 19.)

Thus, for large h_{err}, the variance of $\varepsilon_{\text{pred}}$ is large relative to r_{lin} and biases measurements of ρ^2 to be lower than ρ_{ideal}^2. To determine ρ_{ideal}^2, we must find the contribution of prediction error to ρ^2. It is difficult to directly estimate the size of the residual response in the estimation data. Instead, Equation 18 can be substituted into Equation 19 and rewritten,

$$\frac{1}{\rho^2} = \frac{1}{\rho_{\text{ideal}}^2} + \frac{1}{T} \left(\frac{\langle \varepsilon^2 \rangle}{\langle r_{\text{lin}}^2 \rangle} \frac{1}{\rho_{\text{ideal}}^2} \right). \tag{20}$$

For large T, the second term decreases toward zero, and the measurement of ρ^2 asymptotes to ρ_{ideal}^2. The relative variances of ε and r_{lin} depend on the functional form of the model being validated. Thus, the behavior of this function depends on the particular model represented by the STRF in addition to the size of the estimation data set.

Figure 9B shows the effect of finite estimation sampling on predictions for the same simulated system shown in Figure 9A. Measured values of ρ^2 are shown for STRFs estimated with increasing numbers of samples, T. Each STRF was estimated with independently sampled data in order to avoid correlation artifacts in the fit of the noise model. The solid curves show predictions by Equation 20, given the parameters of the simulated system. As T grows larger, estimation noise is reduced, and predictions improve. The top curve shows prediction accuracy when there is no noise in the validation data. In this case, measurements of ρ^2 asymptote to the theoretical maximum of $\rho_{\text{ideal}}^2 = 0.86$ (dashed line at top).