Spike-triggered neural characterization

Odelia Schwartz The Salk Institute, La Jolla, CA, USA, & Howard Hughes Medical Institute, USA AQ1

Jonathan W. Pillow Gatsby Computational Neuroscience Unit, University College London, London, UK AQ2

Nicole C. Rust Department of Brain and Cognitive Sciences and McGovern Institute for Brain Research, MIT, Cambridge, MA, USA AQ2

Eero P. Simoncelli Howard Hughes Medical Institute, USA, & Center for Neural Science, New York University, New York, NY, USA AQ3

Neural responses are commonly described using receptive fields. This description may be formalized in a model that operates with a small set of linear filters whose outputs are nonlinearly combined to determine the instantaneous firing rate. Spike-triggered average and covariance analyses can be used to estimate the filters and nonlinear combination rule from extracellular experimental data. We describe this methodology, demonstrating it with simulated model neuron examples that emphasize practical issues that arise in experimental situations.

AQ4 Keywords: spike-triggered analysis, characterization, neural response

Introduction

A fundamental goal of sensory systems neuroscience is the characterization of the functional relationship between stimuli and neural responses. The purpose of such a characterization is to elucidate the computation being performed by the system. Many electrophysiological studies in sensory areas describe neural firing rates in response to highly restricted sets of stimuli that are parameterized by one or perhaps two stimulus parameters. Although such "tuning curve" measurements have led to considerable understanding of neural coding, they provide only a partial glimpse of the full neural response function. On the other hand, it is not feasible to measure neural responses to *all* stimuli. One way to make progress is to restrict the response function to a particular model (or class of models). In this modeling approach, the problem is reduced to developing a set of stimuli along with a methodology for fitting the model to measurements of neural responses to those stimuli. One wants a model that is flexible enough to provide a good description of neural responses but simple enough that the fitting is both tractable and well constrained under realistic experimental data conditions.

One class of solutions, which we refer to as "spike-triggered analysis," has received considerable attention in recent years due to a variety of new methodologies, improvements in stimulus generation technology, and demonstration of physiological results. In these methods, one generally assumes that the probability of a neuron eliciting a spike (i.e., the instantaneous firing rate) is governed only by recent sensory stimuli. More specifically, the response model is assumed to be an inhomogeneous Poisson process whose rate is a function of the stimuli presented during a recent temporal window of fixed duration. In the forward neural response model, the stimuli are mapped to a scalar value that determines the instantaneous firing rate of a Poisson spike generator. Our job in the analysis is to work backward: From the stimuli that elicited spikes, we aim to estimate this firing rate function. The analysis of experimental data is thus reduced to examining the properties of the stimuli within temporal windows preceding each recorded spike, known as the *spike-triggered stimulus ensemble* (Figure 1A).

Understanding how the spike-triggered distribution differs from the raw stimuli is key to determining the firing rate function. It is often useful to visualize the analysis problem geometrically (Figure 1B). Consider input stimuli, which at each time step consist of an array of randomly chosen pixel values (8 pixels in this example). The neural response at any particular moment in time is assumed to be completely determined by the stimulus segment that occurred during a prespecified interval in the past (6 time steps in this example). The overall stimulus dimensionality is high (48 dimensions here), but we can depict a projection of the stimuli onto two space–time vectors. The raw stimulus ensemble and the spike-

doi: 10.1167/0.0.1 Received September 4, 2005; published 0 0, 2006 ISSN 1534-7362 © ARVO

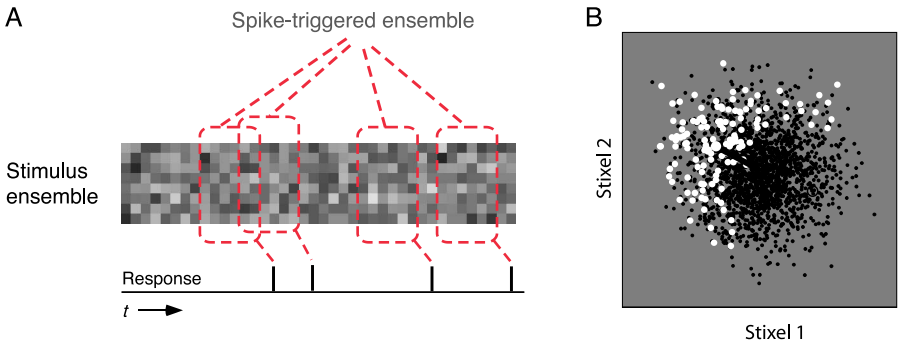

Figure 1. The spike-triggered stimulus ensemble. (A) Discretized stimulus sequence and observed neural response (spike train). On each time step, the stimulus consists of an array of randomly chosen values (eight, for this example). These could represent, for example, the intensities of a fixed set of individual pixels on the screen or the contrast of each of a set of fixed sinusoidal gratings that are additively superimposed. The neural response at any particular moment in time is assumed to be completely determined by the stimulus segment that occurred during a prespecified interval in the past. In this figure, the segment covers six time steps and lags three time steps behind the current time (to account for response latency). The spike-triggered ensemble consists of the set of segments associated with spikes. (B) Geometric (vector space) view of the spike-triggered ensemble. Stimuli (here, 48-dimensional) are projected onto two space–time vectors. In this example, each of the two vectors contained 1 stixel (space–time pixel) set to a value of 1, and the other 47 stixels were set to 0. For these given vectors, the projection is equivalent to the intensity of the corresponding stixel in the stimulus. More generally, one can project the stimuli onto any two 48-dimensional vectors. The spike-triggered stimulus segments (white points) constitute a subset of all stimulus segments presented (black points).

triggered ensemble are then two clouds of points in this space. Intuitively, the task of estimating the neural response function corresponds to describing the ways in which these two clouds differ. In practice, when the input stimulus space is of high dimensionality, one cannot estimate the neural response function without further assumptions.

Spike-triggered analysis has been employed to estimate the terms of a Wiener/Volterra expansion (Korenberg, Sakai, & Naka, 1989; Marmarelis & Marmarelis, 1978; Volterra, 1959; Wiener, 1958), in which the mapping from stimuli to firing rate is described using a low-order polynomial (see Dayan & Abbott, 2001; Rieke, Warland, de Ruyter van Steveninck, & Bialek, 1997 for a review). Although any reasonable function can be approximated as a polynomial, the firing rate nonlinearities found in the responses of sensory neurons (e.g., half-wave rectified, rapidly accelerating and saturating) tend to require a polynomial with many terms (see, e.g., Rieke et al., 1997). However, the amount of data needed for accurate estimation grows rapidly with the number of terms. Therefore, in an experimental setting where one can estimate only the first few terms of the expansion, the polynomial places a strong restriction on the nonlinearity.

As an alternative to the polynomial approximation, one can assume that the response function operates on a low-dimensional *linear subspace* of the full stimulus space (Bialek & de Ruyter van Steveninck, 2005; de Ruyter van Steveninck & Bialek, 1988). That is, the response of a neuron is modeled with a small set of linear filters whose outputs are combined nonlinearly to generate the instantaneous firing rate. Stated differently, although the stimulus space is high dimensional, it is assumed that the neuron only cares about a small set of dimensions. This is in contrast to the Wiener/Volterra approach, which in general does not restrict the subspace but places a restriction on the nonlinearity.[1] By concentrating the data into a space of reduced dimensionality, the neural response can be fit with less restriction on the form of the nonlinearity.

A number of techniques have been developed to estimate the linear subspace and, subsequently, the nonlinearity. In the most widely used form of this analysis, the linear front end is limited to a single filter that serves as an explicit representation of the "receptive field" of the neuron, but the nonlinearity is essentially unrestricted. With the right choice of stimuli, this linear filter may be estimated by computing the spike-triggered average (STA) stimulus (i.e., the mean stimulus that elicited a spike). The STA has been widely used in studying auditory neurons (e.g., Eggermont, Johannesma, & Aertsen, 1983). In the visual system, STA has been used to characterize retinal ganglion cells (e.g., Meister, Pine, & Baylor, 1994; Sakai & Naka, 1987), lateral geniculate neurons (e.g., Reid & Alonzo, 1995), and simple cells in primary visual cortex (V1; e.g., DeAngelis, Ohzawa, & Freeman, 1993; Jones & Palmer, 1987; McLean & Palmer, 1989). Given the STA filter, one typically has enough experimental data to construct a nonparametric estimate of the nonlinearity (i.e., a lookup table; Anzai, Ohzawa, & Freeman, 1999; Chichilnisky, 2001; deBoer & Kuyper, 1968; Eggermont et al., 1983). For some classes of nonlinearity, it has also been shown that one can write down a closed-form solution for the estimates of the linear filter and nonlinearity in a single step (Nykamp & Ringach, 2002).

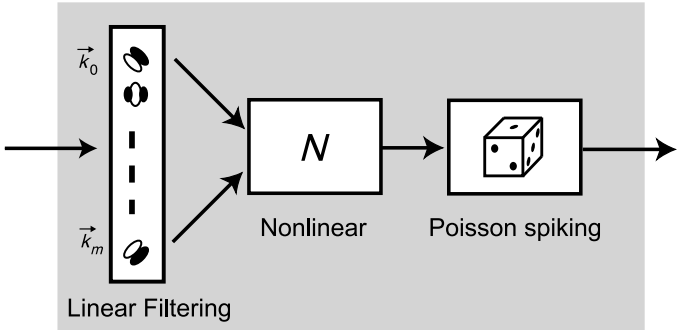

Figure 2. Block diagram of the LNP model. On each time step, the components of the stimulus vector are linearly filtered by $\vec{k}_0 \ldots \vec{k}_m$. The responses of the linear filters are then passed through a nonlinear function $N(\cdot)$, whose output determines the instantaneous firing rate of a Poisson spike generator.

This methodology may be extended to the recovery of multiple filters (i.e., a low-dimensional subspace) and the nonlinear combination rule. One approach to finding a low-dimensional subspace is the spike-triggered covariance (STC; Bialek & de Ruyter van Steveninck, 2005; de Ruyter van Steveninck & Bialek, 1988). STC has been used to characterize multidimensional models and a nonlinear combination rule in systems ranging from the invertebrate motion system (Bialek & de Ruyter van Steveninck, 2005; Brenner, Bialek & de Ruyter van Steveninck, 2000; de Ruyter van Steveninck & Bialek, 1988) to songbird forebrain auditory neurons (Sen, Wright, Doupe, & Bialek, 2000) to the vertebrate retina cells (Pillow, Simoncelli, & Chichilnisky, 2003; Schwartz, Chichilnisky, & Simoncelli, 2002) and mammalian cortex (Horwitz, Chichilnisky, & Albright, 2005; Rust, Schwartz, Movshon, & Simoncelli, 2004, 2005; Touryan, Lau, & Dan, 2002). In addition, several authors have recently developed subspace estimation methods that use higher order statistical measures (Paninski, 2003; Sharpee, Rust, & Bialek, 2003, 2004). A review of spike-triggered subspace approaches may also be found in Ringach (2004) and Simoncelli, Pillow, Paninski, & Schwartz (2004).

Despite the theoretical elegance and experimental applicability of the subspace methods, there are a host of issues that an experimentalist is likely to confront when attempting to use them: How should one choose the stimulus space? How many spikes does one need to collect? How does one know if the recovered filters are significant? How should one interpret the filters? How do the filter responses relate to the nonlinear firing rate function? and so on. In this article, we describe the family of spike-triggered subspace methods in some detail, placing emphasis on practical experimental issues, and demonstrating these (where possible) with simulations. We focus our discussion on the STA and STC analyses, which have become quite widely used experimentally. A software implementation of the methods described is available on the Internet at http://www.cns.nyu.edu/~lcv/stc/.

The linear–nonlinear Poisson (LNP) model

Experimental approaches to characterizing neurons are generally based on an underlying response model. Here, we assume a model constructed from a cascade of three operations:

1. a set of linear filters, $\{\vec{k}_1 \ldots \vec{k}_m\}$,
2. a nonlinear transformation that maps the instantaneous responses of these filters to a scalar firing rate, and
3. a Poisson spike generation process, whose instantaneous firing rate is determined by the output of the nonlinear stage.

This LNP cascade is illustrated in Figure 2. The third stage, which essentially amounts to an assumption that the generation of spikes depends only on the recent stimulus (and not on the history of previous spike times), is often not stated explicitly but is critical to the analysis.

If we assume a discretized stimulus space, we can express the instantaneous firing rate of the model as:

$$r(t) = N(\vec{k}_1 \cdot \vec{s}(t), \vec{k}_2 \cdot \vec{s}(t), \ldots \vec{k}_m \cdot \vec{s}(t)), \qquad (1)$$

where $\vec{s}(t)$ is a vector containing the stimuli over an appropriate temporal window preceding the time t. Here, the linear response of filter i (i.e., the projection or dot product of the filter $\vec{k}_i$ with the stimuli $\vec{s}(t)$) is given by $\vec{k}_i \cdot \vec{s}(t)$. The nonlinear transformation $N(\cdot)$ operates over the linear filter responses.

Spike-triggered analysis

We aim to characterize the LNP model by analyzing the spike-triggered stimulus ensemble. The spike-triggered analysis techniques proceed as follows:

1. Estimate the low-dimensional linear subspace (set of filters). This effectively projects the high-dimension stimulus into a low-dimensional subspace that the neuron cares about.
2. Compute the filter responses for the stimulus, and estimate the nonlinear firing rate function based on these responses. As noted earlier, typical physiological data sets allow nonparametric estimates of the nonlinearity for one or two filters but require more model restrictions as the number of filters increases.

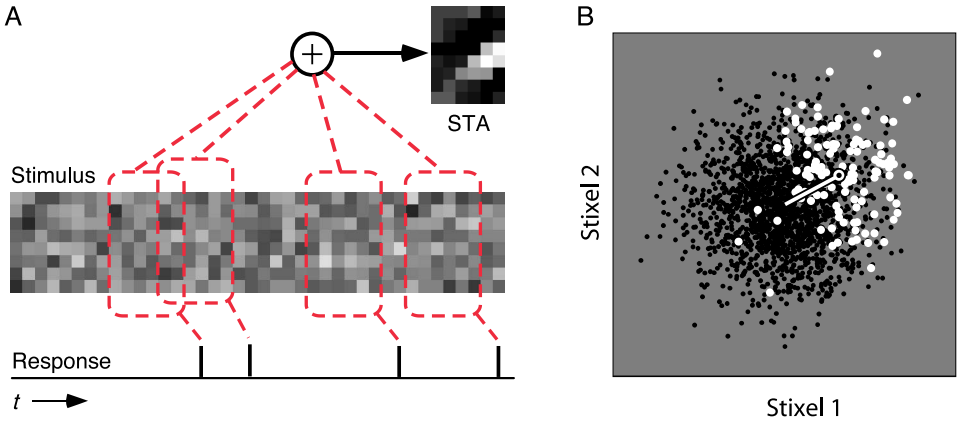

Figure 3. Two alternative illustrations of STA. (A) The STA is constructed by averaging the spike-triggered stimulus segments in red boxes (and subtracting off the average over the full set of stimulus segments). (B) Geometric (vector space) depiction of spike-triggered averaging in two dimensions. Black points indicate raw stimuli. White points indicate stimuli eliciting a spike. The STA, indicated by the line in the diagram, corresponds to the difference between the mean (center of mass) of the spike-triggered ensemble and the mean of the raw stimulus ensemble.

In the following subsections, we describe these steps in detail. In the Experimental issues section, we also stress the importance of an additional step: validating the resulting model by comparing it to neural responses from other stimuli.

Subspace (filter) estimation

In general, one can search for any deviation between the raw and spike-triggered stimulus ensembles. This can be done, for instance, using measures of information theory (Paninski, 2003; Sharpee et al., 2003, 2004). Another natural approach is to consider only changes in low-order moments between the raw and spike-triggered stimulus. Here, we focus on changes in the first and second moments, which may be computed efficiently and manipulated using a set of standard linear algebraic techniques. We also briefly discuss how the analysis relates to the Wiener/Volterra approach.

Spike-triggered average

The simplest deviation between the spike-triggered and raw stimulus distributions is a change in the mean. Assuming that the raw stimuli have zero mean, this can be estimated by computing the average of the spike-triggered ensemble (STA):

$$\hat{A} = \frac{1}{N}\sum_{n=1}^{N} \vec{s}(t_n), \tag{2}$$

where t_n is the time of the nth spike, $\vec{s}(t_n)$ is a vector representing the stimuli presented during the temporal window preceding that time, and N is the total number of spikes. In practice, the times t_n are binned. If there is more than one spike in a bin, then the stimulus vector for that time bin is multiplied by the number of spikes that occurred. The STA is illustrated in Figure 3A.

For an LNP model with a single linear filter, the STA provides an unbiased estimate of this filter,[2] provided that the input stimuli are spherically symmetric (Bussgang, 1952; Chichilnisky, 2001; Paninski, 2003), and the nonlinearity of the model is such that it leads to a shift in the mean of the spike-triggered ensemble relative to the raw ensemble (see Limitations and potential failures section and Experimental issues section). This last requirement rules out, for example, a model with a symmetric nonlinearity such as full-wave rectification or squaring.

For an LNP model with multiple filters, the STA provides an estimate of a particular *linear combination* of the model filters, subject to the same restrictions on input stimuli and the form of the nonlinearity given above (Paninski, 2003; Schwartz et al., 2002). That is, the STA lies in the subspace spanned by the filters, but one cannot assume that it will exactly represent any particular filter in the model.

Spike-triggered covariance

The STA only recovers a single filter. Additional filters may be recovered seeking directions in the stimulus space in which the *variance* of the spike-triggered ensemble differs from that of the raw ensemble. Assuming that the raw stimuli have spherical covariance, this is achieved by computing the STC matrix:

$$\hat{C} = \frac{1}{N-1}\sum_{n=1}^{N} \left[\vec{s}(t_n) - \hat{A}\right]\left[\vec{s}(t_n) - \hat{A}\right]^T, \tag{3}$$

where the $[\cdot]^T$ indicates the transpose of the vector. Again, the t_n are binned in practice, and this means that each term should be multiplied by the number of spikes occurring in the associated time bin.

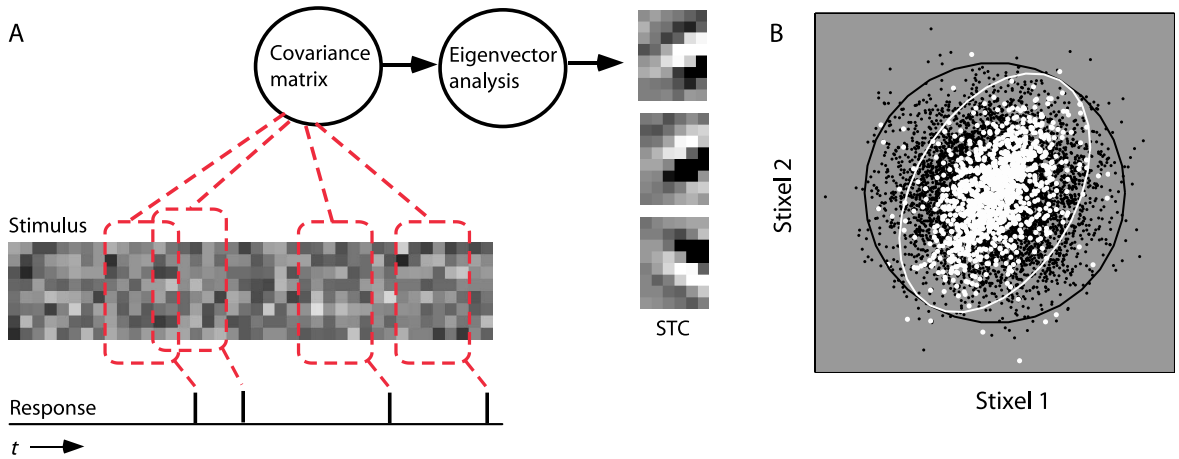

Figure 4. Two alternative illustrations of STC. (A) The STC is determined by constructing the covariance of the spike-triggered stimuli (relative to the raw stimuli), followed by an eigenvector analysis of the covariance matrix. This can result in multiple filters that represent directions in stimulus space for which the spike-triggered stimuli have lower or higher variance than the raw stimuli. (B) Geometric depiction of STC. Black points indicate raw stimuli. White points indicate stimuli eliciting a spike. Ellipses represent the covariance of each ensemble. Specifically, the distance from the origin to the ellipse along any particular direction is the standard deviation of the ensemble in that direction. Raw stimuli are distributed in a circular (Gaussian) fashion. Spike-triggered stimuli are elliptically distributed, with a reduced variance (relative to the raw stimuli) along the minor axis. The minor axis of the ellipse corresponds to a suppressive direction: Stimuli that have a large component along this direction (either positive or negative) are less likely to elicit a spike. The variance of the major axis of the ellipse matches that of the raw stimuli and, thus, corresponds to a direction in stimulus space that does not affect the neuron's firing rate.

The STC matrix embodies the multidimensional variance structure of the spike-triggered ensemble. Specifically, the variance of the ensemble in any direction specified by a unit vector, $\hat{u}$, is simply $\hat{u}^T\hat{C}\hat{u}$. The surface swept out by all such unit vectors scaled by the square root of their associated variance is a multidimensional ellipsoid. The principle axes of this ellipsoid, along with the associated variances, may be recovered as the eigenvectors and associated eigenvalues of the STC matrix. This is illustrated in Figure 4. The consistency of the STC estimate is guaranteed, provided that the input stimuli are Gaussian (Paninski, 2003) and the nonlinearity of the model is such that it leads to a change in the variance of the spike-triggered ensemble relative to the raw ensemble. Note that the Gaussianity is a more severe requirement than the spherical symmetry required for STA analysis (see Limitations and potential failures section and Experimental issues section).

The STA and STC filters together form a low-dimensional linear subspace of the neural response. A number of groups have presented different approaches for combining the STA and STC analyses; in practice, these variants all converge to the same estimated subspace.[3] Usually, the STA is subtracted prior to computing the STC filters (Brenner, Bialek & de Ruyter van Steveninck, 2000; de Ruyter van Steveninck & Bialek, 1988). It is often (but not always) the case that the STA will lie within the subspace spanned by the significant STC axes. Depending on the nonlinear properties of the response, it could coincide with either high- or low-variance STC axes. To simplify visualization and interpretation of the axes, we have chosen for all of our examples to perform the STC analysis in a subspace orthogonal to the STA. Specifically, we compute STC on a set of stimuli from which the STA has been projected:

$$\vec{s} = \vec{s} - [\vec{s}^T\hat{A}]\hat{A}/|\hat{A}|^2. \quad (4)$$

Comparison to Wiener/Volterra analysis

The STA provides an estimate of the first (linear) term in a polynomial series expansion of the system response function and, thus, is the first term of the Wiener/Volterra series. Whereas the Wiener/Volterra approach assumes that the nonlinearity is literally a polynomial, in the STA subspace approach, the nonlinearity is essentially unrestricted. For nonlinearities such as a sigmoid, the Wiener/Volterra expansion would require many terms to capture the neural response function. An example of STA analysis for characterizing a model with a single filter and sigmoidal nonlinearity is presented in the model simulations below.

The second-order term in the Wiener series expansion describes the response as a weighted sum over all pairwise products of components in the stimulus vector. The weights of this sum (the second-order Wiener kernel) may be estimated from the STC matrix. However, the STC method

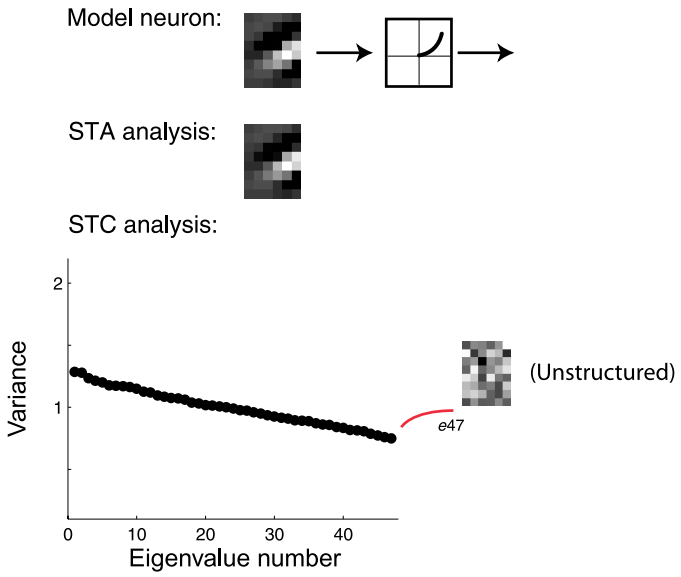

Figure 5. Eigenvalues and eigenvectors for an LNP model with a single linear filter followed by a point nonlinearity. The simulation is based on a sequence of 50,000 stimuli, with a response containing 1,891 spikes. Top: Model filter and nonlinearity. As in Figure 1, filters are 6 × 8 and, thus, live in a 48-dimensional space. The nonlinearity cartoon represents half squaring: Positive filter responses are squared, and negative filter responses are set to zero. Bottom: STA filter, and sorted eigenvalues of covariance matrix of stimuli eliciting spikes (STC). We plot the first 47 eigenvalues and omit the last eigenvalue, which is zero due to projecting out the STA (see Equation 4). The eigenvalues are gradually descending, and corresponding eigenvectors appear unstructured.

is *not* just a specific implementation of a second-order Wiener/Volterra model. The STC approach uses the STC matrix as a means to obtain a linear subspace, within which the nonlinearity is much less restricted. In contrast, the second-order Wiener/Volterra approach assumes a quadratic nonlinearity: This is suitable for characterizing nonlinearities such as the "energy model" (Adelson & Bergen, 1985) of complex cells in primary visual cortex (e.g., Emerson, Bergen, & Adelson, 1992; Emerson, Citron, Vaughn, & Klein, 1987; Szulborski & Palmer, 1990); however, it cannot describe response functions with nonlinearities such as divisive gain control (Albrecht & Geisler, 1991; Heeger, 1992) because these cannot be formulated as sums (or differences) of squared terms. An STA/STC approach is more flexible in capturing such nonlinearities (Rust, Schwartz, et al., 2005; Schwartz et al., 2002), as we demonstrate in the next section.

Simulations of example model neurons

We simulate an example ideal simple cell model, for which there is only a single filter, followed by half-wave rectification and then squaring. Specifically, the instantaneous firing rate is determined by:

$$g(\vec{s}) = r\left[\lfloor \vec{k} \cdot \vec{s} \rfloor^2\right]. \tag{5}$$

The spike-triggered analysis results are shown in Figure 5. The spike-triggered ensemble exhibits a change in the mean relative to the raw stimulus ensemble due to the asymmetric nonlinearity. We recover the STA filter by computing the change in the mean (Equation 2). Next, we consider changes in the variance between the raw and spike-triggered stimulus ensemble. For this model neuron, there is no further relationship between the stimulus space and spikes. In the limit of infinite data, the spike-triggered ensemble would be a randomly selected subset of the raw stimulus ensemble, and the variance in any direction would be identical to that of the raw stimulus set. In an experimental setting, the finiteness of the spike-triggered ensemble produces random fluctuation of the variance in different directions. As a result, there are small random increases or decreases in variance of the spike-triggered ensemble relative to the raw stimulus set. This is reflected in the eigenvalue analysis of Figure 5. Due to the random fluctuations, the sorted eigenvalues cover a range around a constant value of 1 (i.e., the variance of the raw stimulus ensemble) but are not exactly equal to this constant value.

Now, consider an example model neuron, for which there is more than a single filter. We simulate an ideal V1 complex cell model (see also simulations in Sakai & Tanaka, 2000). The model is constructed from two space–time-oriented linear receptive fields, one symmetric and the other antisymmetric (Adelson & Bergen, 1985). The linear responses of these two filters are squared and summed, and the resulting signal then determines the instantaneous firing rate:

$$g(\vec{s}) = r\left[(\vec{k}_1 \cdot \vec{s})^2 + (\vec{k}_2 \cdot \vec{s})^2\right]. \tag{6}$$

Spike-triggered analysis on the model neuron is shown in Figure 6. The STA is close to zero. This occurs because for every stimulus, there is a stimulus of opposite polarity (corresponding to a vector on opposite sides of the origin) that is equally likely to elicit a spike, and thus, the average stimulus eliciting a spike will be zero. The recovered eigenvalues indicate that two directions within this space have substantially higher variance than the others. The eigenvectors associated with these two eigenvalues correspond to the two filters in the model (formally, they span the same subspace). In contrast, eigenvectors corresponding to eigenvalues in the gradually descending region appear arbitrary in their structure.

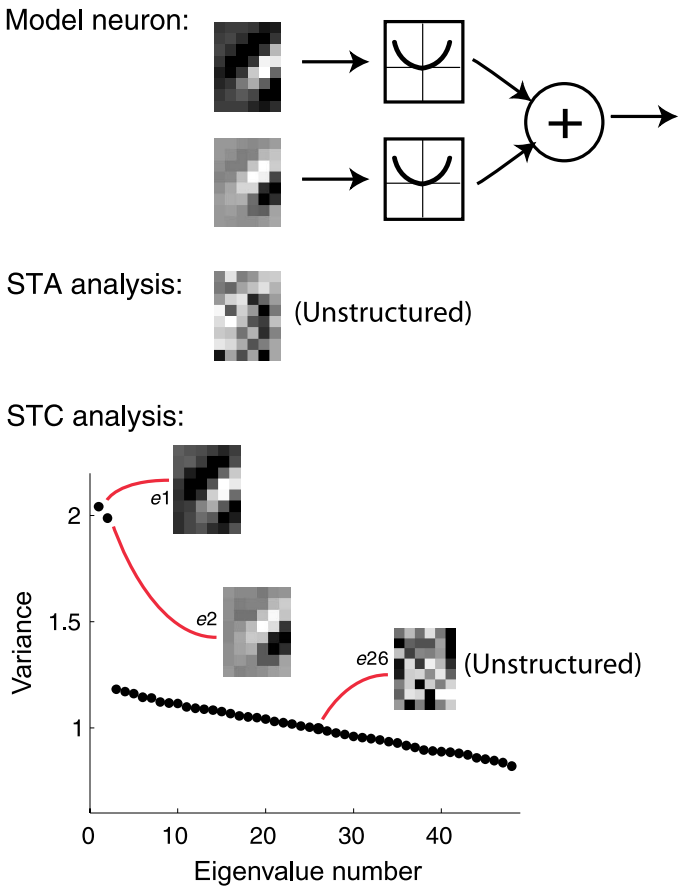

Figure 6. Eigenvalues and eigenvectors for an LNP ideal complex cell model. In this model, the Poisson spike generator is driven by the sum of squares of two oriented linear filter responses. As in Figure 1, filters are 6 × 8 and, thus, live in a 48-dimensional space. The simulation is based on a sequence of 50,000 raw stimuli, with a response containing 4,298 spikes. Top: Model, including two input filters, nonlinearities, and Poisson spiking. Bottom: STA filter is unstructured for the ideal complex cell. The plot also shows the eigenvalues, sorted in descending order. We plot the first 47 eigenvalues and omit the last eigenvalue which is zero due to projecting out the STA (see Equation 4). Two of the eigenvalues are substantially larger than the others and indicate the presence of two directions in the stimulus space along which the model responds. The others correspond to stimulus directions that the model ignores. Also shown are three example eigenvectors (6 × 8 linear filters), two of which are structured while one is unstructured.

Finally, we consider a version of a divisive gain control model (e.g., Geisler, 1992; Heeger, 1992):

$$g(\vec{s}) = r\frac{1 + \lfloor \vec{k}_1 \cdot \vec{s} \rfloor^2}{1 + (\vec{k}_2 \cdot \vec{s})^2 + .4(\vec{k}_3 \cdot \vec{s})^2}. \quad (7)$$

The analysis results are shown in Figure 7. First, we recover the STA filter, which is nonzero due to the half squaring in the numerator. A nonsymmetrical nonlinearity of this sort is captured by changes in the mean. Next, we examine the sorted eigenvalues obtained from the STC analysis. Most of the eigenvalues descend gradually, but the last two eigenvalues lie significantly below the rest, and their associated eigenvalues span approximately the same subspace as the actual simulation filters.

Significance testing

How do we know if the recovered STA and STC filters are significant? In some cases, such as a prototypical complex cell in primary visual cortex, there is essentially no difference between the mean of the raw and spike-triggered stimuli (Rust, Schwartz, et al., 2005; Touryan et al., 2002), which leads to a weak STA. To quantify this, we test the hypothesis that the difference between the mean of the raw and spike-triggered stimulus is no different than what one would expect by chance. We specifically test whether the magnitude of the true spike-triggered stimulus STA is smaller or equal to what would be expected by chance. More specifically, we generate a distribution of random STA filters by bootstrapping: We randomly time-shift the spike train relative to the raw stimulus sequence, gather the resulting spike-triggered stimulus ensemble, and perform the STA analysis. The randomly time-shifted spike train retains all temporal structure that is present in the original spike train. We repeat this 1,000 times, each time computing the average of the stimulus subset. We can then set a significance criterion (e.g., the 95% confidence interval) within which we deem the magnitude of the true STA to be insignificant.

The issue of significance is also of importance for the STC filters. Although the low-variance eigenvalues are clearly below the gradually descending region in the illustrated example of Figure 7, the distinction is not so obvious in some experimental situations. An example in which the significance cutoff is not clear-cut is shown in Figure 8. A significance test should allow us to determine the number of eigenvector axes (filters) corresponding to significant increases or decreases in variance. That is, we would like to find changes in variance in the spike-triggered ensemble that are not just due to chance (because of the finiteness of the number of spikes) but that relate to actual neural response characteristics.

The significance testing must be done in a nested fashion because the distribution of the lowest and highest eigenvalues under the null hypothesis depends on the dimensionality of the space. We begin by assuming that none of the eigenvalues are significant. We compare the true eigenvalues to the eigenvalues of randomly selected stimuli with the same interspike interval. If the largest true eigenvalue lies outside the range of largest eigenvalues of the randomly shifted stimuli and if the smallest true eigenvalue lies outside the range of smallest eigenvalues of the randomly

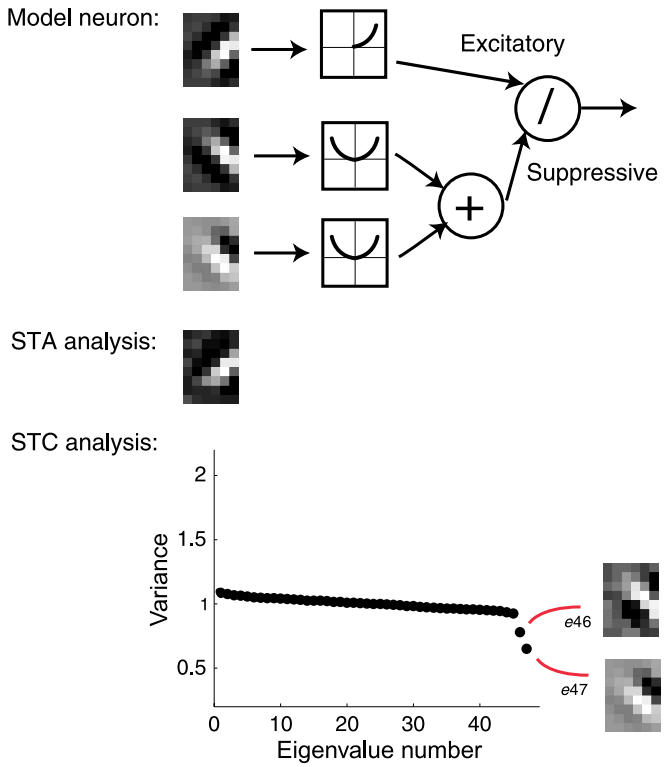

Figure 7. Eigenvalues and eigenvectors for an LNP divisive normalization model. The simulation is based on a sequence of 250,000 stimuli, with a response containing 30,444 spikes. Top: Model. Bottom: STA filter, sorted eigenvalues of covariance matrix of stimuli eliciting spikes (STC), and eigenvectors. Two of the eigenvalues are substantially lower than the others and indicate the presence of two suppressive directions in the stimulus space.

shifted stimuli, then we conclude that none of our axes are significant and accept the hypothesis. More specifically, to compute the randomly selected eigenvalues, we generate distributions of minimal/maximal eigenvalues by bootstrapping: We randomly time-shift the spike train relative to the raw stimulus sequence, gather the resulting spike-triggered stimulus ensemble, perform the STA and STC analysis on the spike-triggered ensemble, and extract the minimum and maximum eigenvalues. After repeating 1,000 times, we estimate the 95% confidence interval for both the largest and smallest eigenvalues. We then ask whether the maximal and minimal eigenvalues obtained from the true spike-triggered ensemble lie within this interval. If so, we accept the hypothesis.

Figure 8A shows that the hypothesis of no significant eigenvalues is unlikely to be correct for this example: The smallest eigenvalue lies far beyond the confidence interval. We therefore assume that the largest outlier (here, the smallest eigenvalue) has a corresponding axis that significantly affects the variance of the neural response. We thus proceed to test the hypothesis that all remaining axes are insignificant. To do so, we first project out the axis deemed significant and repeat the bootstrapping in the remaining subspace. Note that the distribution of eigenvalues (gray region in Figures 8A, 8B, and 8C) changes as the dimensionality of the remaining space decreases. We continue this process in a nested fashion, until the largest and smallest eigenvalues from the true spike-triggered ensemble lie within the estimated confidence interval. Figure 8B shows that we cannot accept the hypothesis of two significant axes. Finally, the hypothesis of four significant axes (Figure 8B) is accepted and results in eigenvalues that lie within the confidence interval.

Filter estimation accuracy

Assuming that the recovered STA and STC filters are significant, we would also like to understand how accurate they are. The accuracy of our estimated filters depends on three quantities: (1) the dimensionality of the stimulus space, d; (2) the number of spikes collected, N; and (3) the

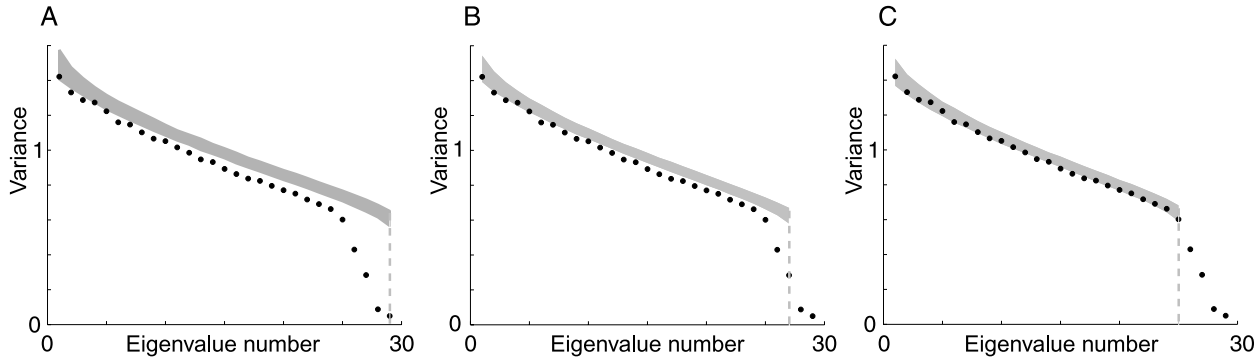

Figure 8. (A) Nested hypothesis testing. Gray solid line corresponds to 95% confidence interval, assuming no suppressive axes (B), two suppressive axes, and (C) four suppressive axes. If the hypothesis is accepted, eigenvalues should lie within the confidence interval. For the assumption of no or two suppressive axes, some eigenvalues lie below the confidence interval, indicating that the hypothesis is incorrect. In contrast, for the assumption of four suppressive axes, eigenvalues lie roughly within the confidence interval.

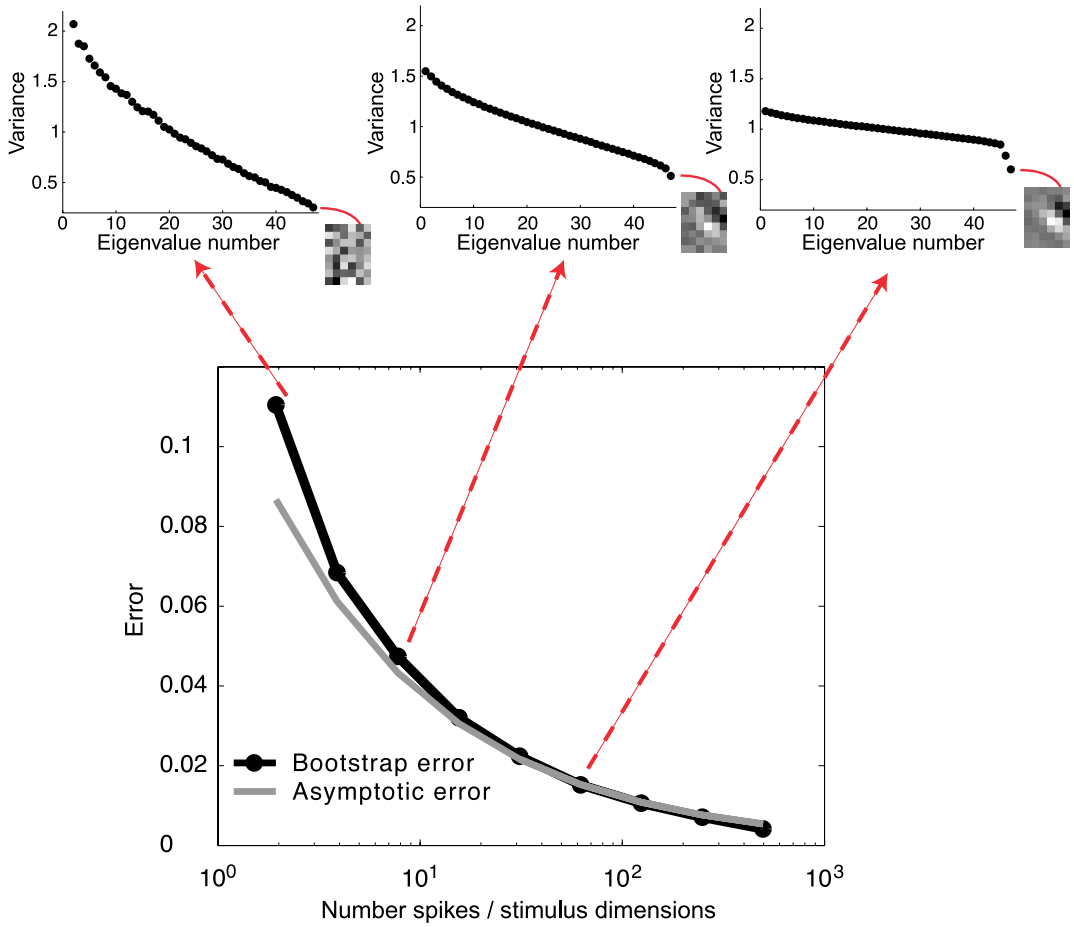

Figure 9. Accuracy in filter estimation. Simulations are shown for the divisive normalization example of Figure 7. Bottom: The error is computed as a function of the ratio of number of spikes to stimulus dimensionality. Stimulus dimensionality is held fixed for all simulations but a number of input stimuli (and thus spikes) are varied. Black line and points is the bootstrap-estimated error (mean angular error obtained from bootstrapping; see main text) of estimation of the lowest eigenvector. The gray line is the theoretical prediction of the mean angular error, computed as the square root of the stimulus dimensionality (here, 48) to number of spikes (see Equation 8 and Paninski, 2003). We multiply the theoretical prediction by a constant parameter that yields the least square error with the bootstrap-estimated error above for the last five points (because the theoretical prediction only holds for the small error regime).

strength of the response signal, relative to the standard deviation of the raw stimulus ensemble, σ.

Asymptotically, the errors decrease as (Paninski, 2003):

$$\mathrm{MAE}(\vec{k}) = \frac{\sigma}{B(\vec{k})}\sqrt{\frac{d}{N}}, \tag{8}$$

where MAE indicates the mean of the angular error (the arccosine of the normalized dot product) between the estimated filter and the true filter and $B(\vec{k})$ is a proportionality factor that depends inversely on the strength of the response signal (Paninski, 2003). For example, the strength of response signal is the length of the STA vector in the limit of infinite data. An experimentalist does not have access to the strength of response signal. However, the number of spikes and number of stimulus dimensions are known, and thus, the function of Equation 8 may be used to extrapolate the error behavior based on bootstrap estimates. To demonstrate this, we simulate an experiment on the model divisive normalization neuron.

We describe a bootstrap method to determine the error in filter estimation. We show that the bootstrap-estimated error is reasonably matched to the theoretical prediction of the error in Equation 8, when the ratio of number of spikes to number of stimulus dimensions is sufficiently high. We run a pilot experiment on the model divisive normalization neuron and collect 409,600 input samples. We consider how the ratio of stimulus dimensionality to number of spikes affects accuracy. Specifically, we hold the stimulus dimensionality fixed (which is 48 here) and vary the number of input samples (and thus spikes). For a given number of input samples, we bootstrap, drawing (with replacement) random subsets of stimuli (equal to the number of input samples). We consider the spike-triggered stimuli from this subset and compute the STA and STC.

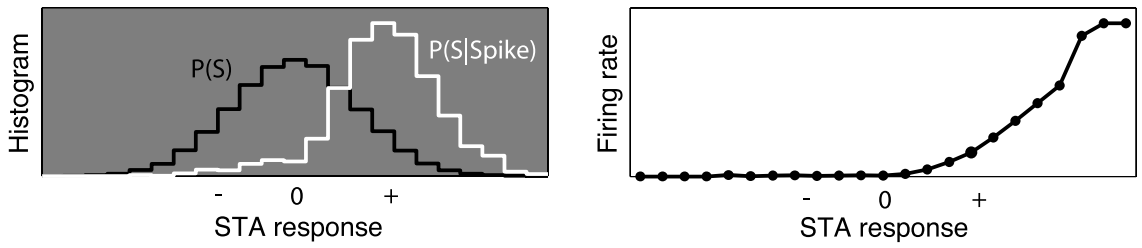

Figure 10. Nonlinearity for an LNP model with a single linear filter followed by a point nonlinearity. Left: Raw (black) and spike-triggered (white) histograms of the linear (STA) responses. Histograms have been renormalized to a maximal probability of 1. Right: The quotient of the spike-triggered and raw histograms gives an estimate of the nonlinearity that generates the firing rate.

We repeat this many times (here, 1,000) and derive an estimate of the mean angular error for a given STC filter. This is achieved by computing the mean of the 1,000 estimated filters from the bootstrapping—we will denote this the mean estimated filter, and then, for each of the 1,000 estimated filters, by computing its mean angular error with the mean estimated filter and taking an average over these computations. This analysis assumes that there are no systematic biases in the estimates (such as those shown in Figure 15).

In Figure 9, we plot the error estimates for the filter corresponding to the lowest eigenvalue. As the number of spikes to number of stimulus dimensions increases, the error is reduced. We also show, for three example ratios, the eigenvalues and the filter estimate corresponding to the lowest eigenvalue. For a low ratio of spike counts to stimulus dimensions, the eigenvalues descend gradually, and the smallest one is not separated from the rest; for a high ratio of spike counts to stimulus dimensions, the eigenvalues take on a pattern similar to Figure 7. Finally, we return to Equation 8: We fit this equation (and corresponding proportionality factor) to the errors derived from bootstrapping and obtain a rather good match for the low error regime. Such an analysis could be used in an experimental situation to determine data requirements for a given error level, by extrapolating the curve from values estimated from a pilot experiment. In the Experimental issues section, we elaborate on running a pilot experiment to choose a reasonable tradeoff between number of spikes and stimulus dimensionality.

Characterizing the nonlinearity

According to the LNP model, the firing rate of a neuron is given by a nonlinear transformation of the linear filter responses (Figure 2). Using the same set of stimuli and spike data as for the linear filter estimation, we seek to estimate the nonlinearity and, thus, characterize a neural model that specifies the full transformation from stimulus to neural firing rate. We therefore need to estimate the firing rate of the neuron as a function of the linear filter responses. To do so, it is important to recognize that the ratio of the frequency of occurrence of spike-triggered stimuli to that of raw stimuli is proportional to the instantaneous firing rate. This can be seen using Bayes rule:

$$\mathcal{P}(\text{spike}|\vec{s}) = \frac{\mathcal{P}(\text{spike})\mathcal{P}(\vec{s}|\text{spike})}{\mathcal{P}(\vec{s})}, \tag{9}$$

and therefore,

$$\mathcal{P}(\text{spike}|\vec{s}) \propto \frac{\mathcal{P}(\vec{s}|\text{spike})}{\mathcal{P}(\vec{s})}, \tag{10}$$

where $\mathcal{P}(\text{spike}|\vec{s})$ is the instantaneous firing rate, $\mathcal{P}(\vec{s}|\text{spike})$ is the frequency of occurrence of spike-triggered stimuli, and $\mathcal{P}(\vec{s})$ is the frequency of occurrence of raw stimuli.

The problem of estimating the nonlinearity can thus be described as one of estimating the ratio of two probability densities of Equation 10. The accuracy of the estimation is dependent on the dimensionality (number of filters) in the linear subspace. For one or two filters, we can use simple histograms to estimate the numerator and denominator of Equation 10. For more filters, this becomes impractical due to the so-called "curse of dimensionality": The amount of data needed to sufficiently fill the histogram bins in a d-dimensional space grows exponentially with d. In this case, we typically need to incorporate additional assumptions about the form of the nonlinearity.

Consider a model LNP neuron with only a single filter followed by a point nonlinearity. First, we estimate the linear filter by computing the STA. Then, we compute the linear filter response for each stimulus, by taking a dot product of the filter with the stimulus. We do this for all instantiations of the spike-triggered stimuli and compute a histogram estimating the numerator density $\mathcal{P}(\vec{s}|\text{spike})$; we do this for all instantiations of the raw stimuli and compute a histogram estimating the denominator density $\mathcal{P}(\vec{s})$. The nonlinearity that determines the firing rate is then the ratio of these two densities or the ratio of the histogram values in each bin. An example is shown in Figure 10 (see also Chichilnisky, 2001). We plot the histograms of the spike-triggered and raw stimuli filter

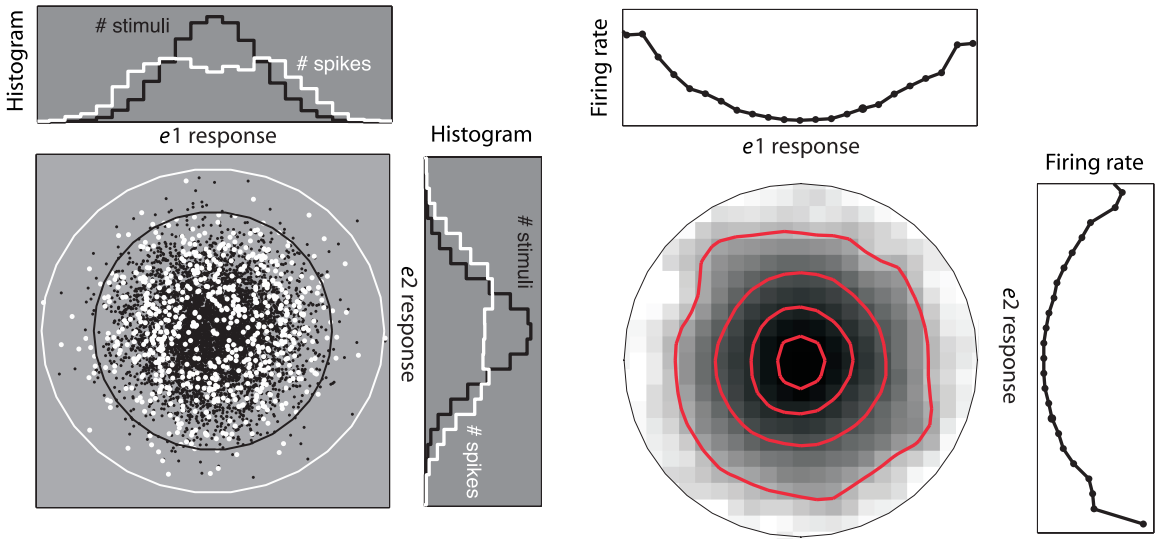

Figure 11. Nonlinearity for ideal complex cell model. This corresponds to eigenvalue and eigenvector example of Figure 6. Left: Scatter plots of stimuli projected onto estimated filters (i.e., filter responses) corresponding to first two eigenvalues (e_1 and e_2). Black points indicate the raw stimulus set. White points indicate stimuli eliciting a spike. Also shown are one-dimensional projections onto a single filter. Right: The quotient of the two-dimensional spike-triggered and raw histograms provides an estimate of the two-dimensional nonlinear firing rate function. This is shown as a circular-cropped grayscale image, where intensity is proportional to firing rate. Superimposed contours indicate four different response levels. Also shown are one-dimensional nonlinearities onto a single filter.

responses (Figure 10, left). We observe the nonlinearity by examining the ratio of these two histograms (Figure 10, right): The instantaneous firing rate grows monotonically and asymmetrically, that is, increases for stimuli to which the filter responds strongly and positively.

Note that the nonlinearity can be arbitrarily complicated (even discontinuous). The only constraint is that it must produce a change in the mean of the spike-triggered ensemble, as compared with the original stimulus ensemble. Thus, the interpretation of reverse correlation in the context of the LNP model is a significant departure from the Wiener/Volterra series expansion, in which even a simple sigmoidal nonlinearity would require the estimation of many terms for accurate characterization (Rieke et al., 1997).

Next, consider an ideal complex cell model neuron as in Equation 6. The recovered eigenvalues indicate that two directions within this space have substantially higher variance than the others (recall Figure 6). As before, we compute the raw and spike-triggered stimulus responses for each of the two filters. A two-dimensional scatter plot of these filter responses is shown in Figure 11 (left) for both the spike-triggered and raw stimuli. This is a two-dimensional depiction of samples from the numerator and denominator distributions in Equation 10. The scatter plots are similar in essence to those described in Figure 4, but the stimuli are projected onto the two filters recovered from the analysis. To estimate the two-dimensional nonlinear firing rate function (Figure 11, right), we compute the two-dimensional histogram for the spike-triggered and raw stimuli responses and calculate the ratio of the histogram values in each bin. This is analogous to the one-dimensional example shown in Figure 10. Similar pairs of excitatory axes and nonlinearities have been obtained from STC analysis of V1 cells in cat (Touryan et al., 2002) and monkey (Rust et al., 2004; Rust, Schwartz, et al., 2005).

Finally, consider the divisive normalization model in Equation 7, for which the eigenvalues and eigenvectors are shown in Figure 7. Figure 12 (left) shows a scatter plot of the STA filter response versus a suppressive filter response. The spiking stimuli lie within an ellipse, with the minor axis corresponding to the suppressive filter. This is exactly what we would expect in a suppressive system, such as that plotted in Figure 4. The two-dimensional nonlinearity is estimated by taking the quotient as before. This reveals an approximately saddle-shaped function, indicating the interaction between the excitatory and suppressive signals (Figure 12, right). Similar suppressive filters have been obtained from STC analysis of retinal ganglion cells (in both salamander and monkey; Schwartz et al., 2002) and simple and complex cells in monkey V1 (Rust et al., 2004).

For some systems, such as H1 of the blowfly (Bialek & de Ruyter van Steveninck, 2005; Brenner, Bialek & de Ruyter van Steveninck, 2000), the dimensionality of STA and STC filters is sufficiently low (and the data set sufficiently large) to calculate the quotient of Equation 10 directly (as we have shown in the simulation examples) and thus estimate the nonlinearity. But what happens when there are more than two significant filters derived from the STA and STC analyses? There is not one single recipe; rather, there are a number of ways to try and approach this

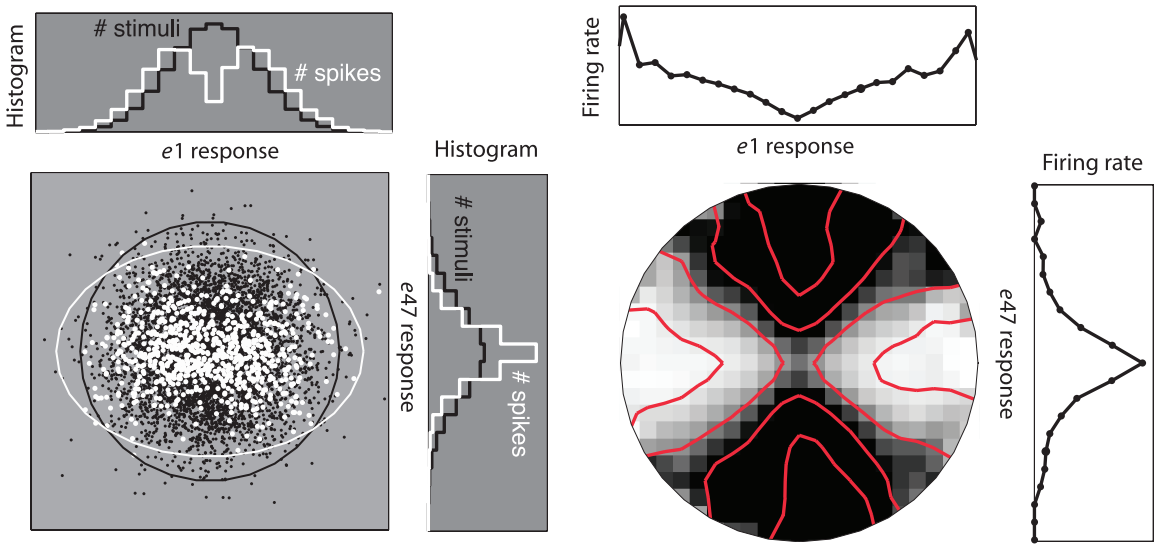

Figure 12. Nonlinearity for divisive normalization model. This corresponds to the eigenvalue and eigenvector example of Figure 7. Left: Scatter plots of stimuli projected onto estimated filters (i.e., filter responses) corresponding to STA and last suppressive eigenvector. Black points indicate the raw stimulus set. White points indicate stimuli eliciting a spike. Also shown are one-dimensional projections onto a single filter. Right: The quotient of the two-dimensional spike-triggered and raw histograms provides an estimate of the two-dimensional nonlinear firing rate function. This is shown as a circular-cropped grayscale image, where intensity is proportional to firing rate. Superimposed contours indicate four different response levels. Also shown are one-dimensional nonlinearities onto a single filter.

problem, and the answer depends on the particular system and data at hand.

One approach is to consider specific classes of LNP models that might be suitable for the particular neural area under study. For instance, in retinal ganglion cell data, it was shown that fitting a divisive normalization model to the filters recovered from STA and STC provided a reasonable characterization of the data (Schwartz et al., 2002). In another study in area V1, the dimensionality of the filters from STA and STC was too high for computing the nonlinearity within the full recovered subspace (Rust, Schwartz, et al., 2005). The form of nonlinearity was restricted by first computing squared sums of excitatory filter responses and squared sums of suppressive filter responses, and only then was the nonlinearity between these pooled excitatory and suppressive signals determined. This simplification could be made because it was observed that projections of stimuli onto the recovered filters within the excitatory or suppressive pools always resulted in elliptical contours—suggesting sum of squares operations governing the combination within each pool. An alternative approach, published in this special issue, assumes that the nonlinearity takes the form of a ratio of Gaussians (Pillow & Simoncelli, in press).

Limitations and potential failures

The STA and STC estimates depend critically on the distribution of input stimuli and on the particular nonlinearity of the neuron. For an LNP model with a single linear filter, the consistency of the STA estimator is guaranteed (e.g., irrespective of the neural nonlinearity) only if the distribution of input stimuli are spherically symmetric; that is, any two stimulus vectors with equal vector length have an equal probability of being presented (Chichilnisky, 2001). If one aims to recover a set of filters using both STA and STC, then the consistency of the estimator is guaranteed under the more stringent condition that the stimuli be Gaussian distributed (Paninski, 2003). The estimator is also guaranteed for elliptically symmetric Gaussian stimuli, in which the covariance matrix is not equal to the identity (see Appendix). For example, even if the raw stimuli are constructed as spherical Gaussian, a finite number of stimuli might, by chance, produce some axes that have (slightly) higher variance than others. There might also be interest in presenting to a neuron colored or $1/f$ noise.

Note that non-Gaussian stimulus distributions can lead to artifacts in the spike-triggered analysis, and the artifacts are dependent on how the nonlinear response properties of the neuron interact with the distribution. In the Experimental issues section, we show simulated examples with non-Gaussian stimuli, demonstrating how this could potentially impact the STA and STC in a model neuron. These examples do not indicate that experiments with non-Gaussian stimuli and STA/STC analysis will *necessarily* lead to artifacts, but because there is no general solution for eliminating artifacts that can arise from non-Gaussian stimuli, it is advisable to run experimental controls with Gaussian stimuli.

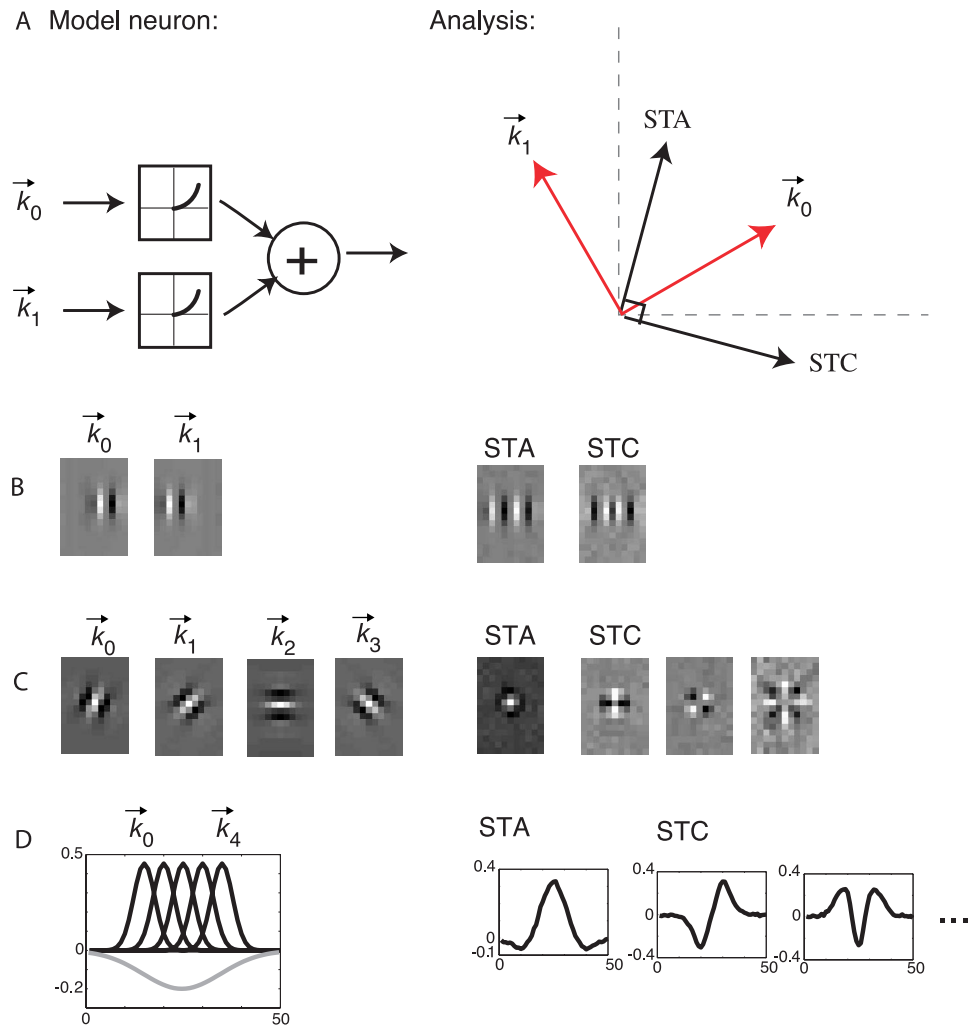

Figure 13. Interpretation issues and sum of half squares LNP model: filters. (A) Left: Model filter responses are half squared (negative values set to zero) and then added together. Note that this is different from the full squaring of the ideal complex cell. Right: Geometry of the STA and STC analysis. The STA is a vector average of the model filters. The STC is forced to be 90 deg away from the STA. Although the STA and STC filters do not equal the model filters, they do span the same subspace. (B) Example of spatially shifted model filters. Both STA and STC analysis reveal filters that are quite different from the model but span the same subspace. (C) Example of oriented filters. We extend the two-filter model to four filters that are each half squared and then added together. The STA is the average of all four filters and has a center/surround appearance rather than an oriented one. The other three STC filters are orthogonal. (D) The model neuron includes five spatially overlapping filters. The filter responses undergo a weighted sum of half squares, followed by addition of a (negative) linear surround (gray curve). The STA is a vector average of the linear filters, and the STC filters are orthogonal.

Even if one is careful to design an experiment and data analysis methodology that leads to accurate and artifact-free estimates, a spike-triggered analysis can still fail if the model assumptions are wrong. Two examples of failure of the LNP model are as follows: (1) there is no low-dimensional subspace in which the neural response may be described or (2) the neural response has a strong dependence on spike history (e.g., refractoriness, bursting, adaptation) that cannot be described by an inhomogeneous Poisson process. STA/STC analysis of data simulated using more realistic spike generation models, such as Hodgkin–Huxley (Agüera y Arcas & Fairhall, 2003; Agüera y Arcas, Fairhall, & Bialek, 2001, 2003; Pillow & Simoncelli, in press) and integrate-and-fire (Pillow & Simoncelli, 2003), produces biased estimates and artifactual filters. Although the STA/STC filters might in some cases still provide a reasonable description of a neuron's response, it is important to recognize that the LNP model provides only a crude approximation of the neural response (see Interpretation issues section).

Interpretation issues

There are a number of important issues that arise in interpreting the spike-triggered analysis. First, the number of filters recovered by STA and STC provides only a *lower*

bound on the actual number of filters. The neural response may be dependent on mechanisms not identified by the STC analysis: (1) Other filters might affect the response, but the dependence is too weak and buried in the statistical error (a possibility with any experimental method—recall Figure 9); or (2) The neural response nonlinearities may not lead to a change in the mean or variance. It should be noted that although such a nonlinearity is theoretically possible, most known physiological nonlinearities *do* affect the mean, the variance, or both.

Next, the recovered filters cannot be taken literally as physiologically instantiated mechanisms. The STC filters, together with the STA, form an *orthogonal* basis for the stimulus subspace in which the responses are generated. The analysis does not yield a unique solution: A whole family of equivalent models can be constructed (by transforming to alternative sets of filters using an invertible linear transformation), which, given the same stimulus, produce the same response. Thus, even if a neuron's response is well described by an LNP model, we cannot claim to recover the actual filters that the neuron is using to compute its response. Rather, the goal is to find a set of filters that span the proper subspace; that is, with this set of filters, one can compute the same responses as with the actual set.

Figure 13 shows a simulation for an example of model neuron in which the STA and STC do not recover the actual model filters but do span the same subspace. The model neuron responds with a rate proportional to a sum of *half squares*, as opposed to the sum of squares typical of the ideal complex cell:

$$g(\vec{s}) = r\left[[\vec{k}_1 \cdot \vec{s}]^2 + [\vec{k}_2 \cdot \vec{s}]^2\right].$$

The simulation results for different input filters are shown in Figure 13. Now, the STA does not result in a zero-weighted filter because the filter responses are not symmetric as in the ideal complex cell. Interestingly, the STA is not equal to either of the two excitatory filters of the model; rather, it is a vector average of the two filters. STC analysis on the stimuli perpendicular to the STA reveals an additional excitatory filter. Note that the two recovered filters together span the excitatory subspace of the original model filters. Figure 13C shows an example with four input filters of different orientations whose responses are half squared and summed; the STA takes on a more center–surround, unoriented appearance. Figure 13D shows an example of five overlapping *spatial* filters. These can be thought of as subunits, as has been proposed for retina (Hochstein & Shapley, 1976; see also Rust, Schwartz, et al., 2005 for cortical data). The nonlinear combination of these filters is followed by a subtraction of a linear surround. The resulting STA takes on the well-known spatial profile of retinal ganglion cells, and the STC filters are forced to be 90 deg apart and similar to what is found experimentally (Pillow, Simoncelli, & Chichilnisky, 2004). The two-dimensional depiction of the nonlinearity for the above examples is interesting: The spike-triggered stimuli form a shape that resembles a portion of an annulus (Figure 14). Neurons with nonlinearities of this flavor can be seen in area V1 of the macaque (Rust, Schwartz, et al., 2005) and in retinal ganglion cells (Schwartz & Simoncelli, 2001).

Another reason why the recovered filters should not be interpreted as a physiological mechanism is that the LNP model assumes Poisson spiking. A number of authors have demonstrated that these Poisson assumptions do not accurately capture the statistics of neural spike trains (Berry & Meister, 1998; Keat, Reinagel, Reid, & Meister, 2001; Pillow, Shlens, Paninski, Chichilnisky, & Simoncelli, 2005a; Reich, Victor, & Knight, 1998). The dependence of neural responses on spike history (e.g., refractoriness, bursting, adaptation) may be captured only indirectly in the LNP model through time-delayed suppressive STC filters (Agüera y Arcas & Fairhall, 2003; Agüera y Arcas et al., 2003; Schwartz et al., 2002). For instance, during a refractory period, a neuron will not spike, and this can be captured by an LNP model with a set of suppressive STC filters in time. The suppressive filters may still provide a reasonably accurate description of the neural response but do not reveal the mechanism of refractoriness.

Finally, the labeling of whether a filter is excitatory or suppressive is crudely based on the net change in the mean or variance and may not correspond physiologically to excitation or suppression. A given filter can indeed be both excitatory and suppressive. For example, a filter might be half square rectified, yielding a positive increase in the mean, but also include a compressive squared nonlinearity (as in divisive normalization). Because the STA and STC filters are orthogonal, the analysis will extract a single filter and label it as excitatory. As before, the analysis still finds the right subspace; one can then analyze the interaction and aim to estimate a model within the subspace.

Experimental issues

We now discuss issues that arise when designing and interpreting spike-triggered experiments.

Stimulus choice

Stimulus space

The stimuli in a spike-triggered experiment need to be restricted to lie in a finite-dimensional space, and the experimentalist must choose the fundamental components (i.e., the axes) of this space. At any moment in time, the neuron is exposed to a linear combination of this set of stimulus components. In many published examples (as well

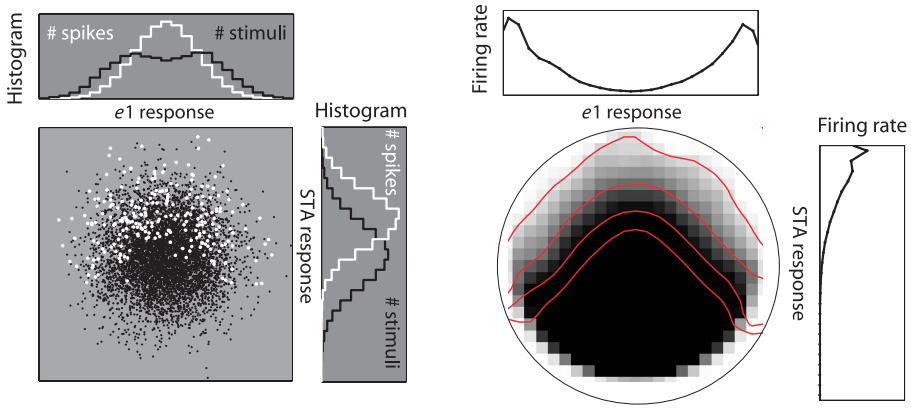

Figure 14. Interpretation issues and sum of half squares LNP model: nonlinearity. Nonlinearity is shown for model simulation of filters in Figure 13B (almost identical plots are found for Figures 13C and 13D). Left: Scatter plots of stimuli projected onto estimated filters (i.e., filter responses) corresponding to STA and first eigenvector. Black points indicate the raw stimulus set. White points indicate stimuli eliciting a spike. Also shown are one-dimensional projections onto a single filter. Right: The quotient of the two-dimensional spike-triggered and raw histograms provides an estimate of the two-dimensional nonlinear firing rate function. This is shown as a circular-cropped grayscale image, where intensity is proportional to firing rate. Superimposed contours indicate four different response levels. Also shown are one-dimensional nonlinearities onto a single filter.

as the examples shown in this article), the axes of the stimulus space corresponds to pixel (or stixel) intensities. However, the stimulus may be described in terms of other components, such as the amplitudes of a particular set of sinusoids (Ringach, Sapiro, & Shapley, 1997), the velocities of a randomly moving spatial pattern (Bair, Cavanaugh, & Movshon, 1997; Borghuis et al., 2003; Brenner, Bialek & de Ruyter van Steveninck 2000; de Ruyter van Steveninck & Bialek, 1988), or any other fixed set of functions. More generally, it is possible to do the analysis in a space that is a nonlinear function of the input stixels (David, Vinje, & Gallant, 2004; Nishimoto, Ishida, & Ohzawa, 2006; Theunissen et al., 2001). This is useful when one believes that the cells' response is LNP on these inputs (Rust, Simoncelli, & Movshon, 2005), although it may then be more difficult to interpret the results. The fundamental constraints on the choice of these components are that (1) the neuron should respond reasonably to stochastically presented combinations of these components and (2) the neuron's response should be well approximated by an LNP model operating in the space of these components.

The choice of a finite-dimensional stimulus space places a restriction on the generality of the experimental results: The response of the cell will only be characterized *within the subspace spanned by the stimulus components* (Ringach et al., 1997). Stated differently, without further assumptions, the model one constructs with STC can only predict stimuli responses that lie in the space defined by the experimental stimulus ensemble. For example, one cannot predict the responses to chromatic stimuli when using achromatic stimuli or to a full two-dimensional space when probing the neuron with only a single spatial dimension (as in the case of bars). Similarly, one cannot use the model to predict responses to stimuli that have a finer spatial or temporal resolution than that used in the characterization.

To obtain a more general characterization, one needs to increase the stimulus resolution. Unfortunately, this increases the dimensionality of the stimulus space and, thus, requires more spikes to achieve the same quality of filter estimation. At the same time, the increase in resolution typically *reduces* the responsivity of the cell (e.g., because the effective contrast is reduced), thus making it more difficult to obtain the needed spikes. Recall that the error in filter estimation is a direct consequence of the ratio of the number of spikes to stimulus dimensionality, as in the example model neuron simulation shown in Figure 9. Therefore, it is useful to run a pilot experiment to determine the proper balance between number of spikes (e.g., duration of the experiment) to stimulus dimensionality for a particular class of neurons. In practice, it useful for a physiologist to adopt a rule of thumb for the particular system at hand: In the V1 experiments, Rust, Schwartz, et al. (2005) found that at least 100 spikes per dimension were typically needed to obtain a good characterization. Other experimental methodologies or settings (e.g., recordings from an awake behaving animal) and other classes of neurons may be more limited in the number of spikes that can be collected.

Stochastic stimulus distribution

As stated earlier, the STC portion of the spike-triggered analysis is only guaranteed to work for Gaussian stimuli. The use of non-Gaussian white noise stimulus distributions (e.g., uniform, binary, sparse) is quite common experimentally, as the samples are easy to generate and the higher

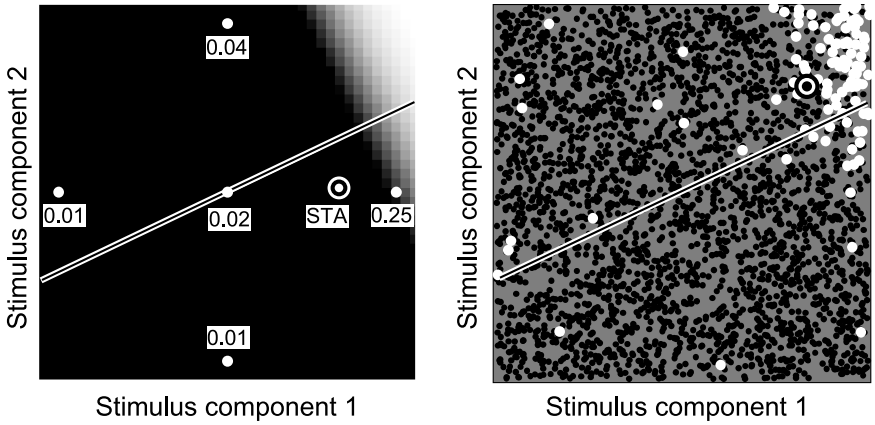

Figure 15. Simulations of an LNP model demonstrating bias in the STA for two different nonspherical stimulus distributions. The linear stage of the model neuron corresponds to an oblique axis (line in both panels), and the firing rate function is a sigmoidal nonlinearity (firing rate corresponds to intensity of the underlying grayscale image in the left panel). In both panels, the black and white "target" indicates the recovered STA. Left: Simulated response to sparse noise. The plot shows a 2-dimensional subspace of a 10-dimensional stimulus space. Each stimulus vector contains a single element with a value of ±1, whereas all other elements are zero. Numbers indicate the firing rate for each of the possible stimulus vectors. The STA is strongly biased toward the horizontal axis. Right: Simulated response of the same model to uniformly distributed noise. The STA is now biased toward the corner. Note that in both examples, the estimate will not converge to the correct answer, regardless of the amount of data collected.

contrast of the stimuli generally leads to higher average spike rates. In practice, their use is often justified by assuming that the linear filters are smooth relative to the pixel size/duration (e.g., Chichilnisky, 2001). Natural signal stimuli (such as visual scenes and auditory vocalizations) are also non-Gaussian (Daugman, 1989; Field, 1987), but their use is becoming increasingly popular (David & Gallant, 2005; David et al., 2004; Felsen & Dan, 2005; Ringach, Hawken, & Shapley, 2002; Sen et al., 2000; Smyth, Willmore, Baker, Thompson, & Tolhurst, 2003; Theunissen et al., 2001; for recent perspectives, see Felsen & Dan, 2005; Rust & Movshon, 2005). Natural signals can reveal response properties that occur less frequently under Gaussian white noise stimulation, such as bursting in the LGN (Lesica & Stanley, 2004), and they are often more effective in driving higher neural areas.

However, nonspherical stimuli can produce artifacts in the STA filters, and non-Gaussian stimuli can produce artifacts in the STC filters. Figure 15 shows two simulations of an LNP model with a single linear filter and a simple sigmoidal nonlinearity, each demonstrating that nonspherical stimulus distributions can lead to poor estimates of the linear stage. The examples are meant to emphasize the potential for bias but do not necessarily mean that an artifact will occur in experiment. Indeed, the particular nonlinear behaviors of the neural response will determine if and how much of a bias occurs. Because we do not know the nonlinearity a priori, the safest approach is to compare the experimental linear filter estimate to a control using spherically symmetric stimuli.

The first example shows a "sparse noise" experiment, in which the stimulus at each time step lies along one of the axes. As shown in the figure, the nonlinearity can result in an STA that is biased toward an axis of the space. The second example uses stimuli in which each component is drawn from a uniform distribution, which produces an estimate biased toward the "corner" of the space. Note, however, that the estimate will be unbiased in the case of a purely linear neuron or of a half-wave-rectified linear neuron (Ringach et al., 1997).

Non-Gaussian stimuli can produce strong artifacts in the STC analysis. Figure 16A (left) shows an example simulation of the divisive normalization model with binary stimuli. Note that in addition to the two "real" suppressive filters of the model, the analysis also finds two significant artifactual suppressive filters; these have a few high-intensity stixels. Similar artifacts have been found in experimental circumstances (Rust, Schwartz, et al., 2005). More intuition for the artifacts can be gained by examining two-dimensional scatter plots that include an artifactual filter response versus the STA filter response (Figure 16A, right). The raw binary stimuli are clearly not spherical in this two-dimensional view. Specifically, the set tapers as one moves in the direction of the STA. This reduction in variance of the raw stimulus happens to coincide with the stimuli that elicit spikes (i.e., those that have a large STA component). Thus, the spike-triggered analysis reveals the artifactual filter as an axis of significantly reduced variance, although it is actually not reduced relative to the raw stimuli.

There is, unfortunately, no generic recipe for reducing artifacts. From our experience with binary stimuli, we have found that the artifacts can be partially corrected by adjusting the raw stimulus such that the covariance estimated at each value of the STA is equal (conditional whitening; Rust, Schwartz, et al., 2005). Specifically, we

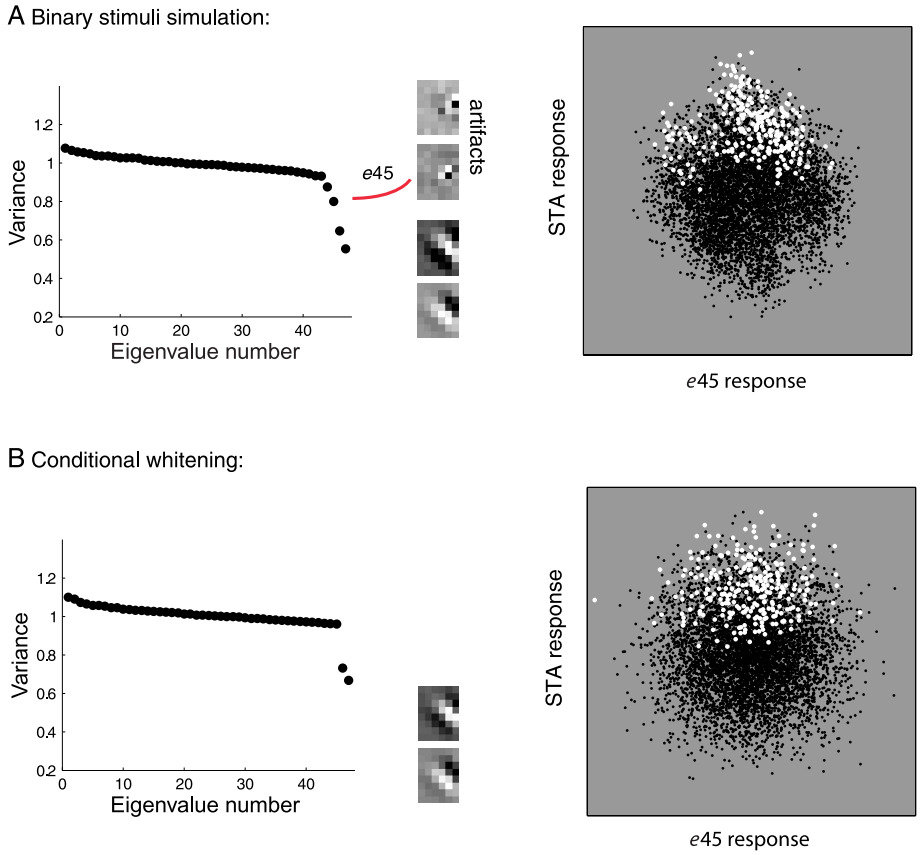

Figure 16. STC artifacts with binary stimuli. We ran the same model neuron as in Figure 7, but we replaced Gaussian stimuli with binary stimuli. (A) Left: There are four eigenvalues significantly below what one would expect by chance. Two of the corresponding eigenvectors correspond to the real model filter subspace, but two of them are artifactual. Right: Projection onto one of the artifactual filters versus the STA. The raw stimuli are nonspherical and have regions of lower variance at the top and bottom corners. The spiking stimuli appear in the upper corner because this is where the STA projection is largest. Although the variance of the raw and spike-triggered stimuli is the same when confined to this corner, the variance of the spike-triggered stimuli is significantly smaller than the variance of the entire raw ensemble, and this generates the artifactual suppressive filter (*e*45). (B) Left: After conditional whitening (see main text), there are only two significantly low eigenvalues corresponding to the model neuron subspace. Right: Projection onto the same eigenvalue as the artifactual filter above, as against the STA. The raw stimuli are now not perfectly circular but have roughly equal variance in all directions.

partition the stimuli of Figure 16A (right) into horizontal slabs according to the value of the excitatory filter response and compute the covariance matrix for each subset (C_n for the nth subset). The stimuli in each subset are whitened by multiplying them by

$$E_e E_e^T + E_0 E_n D_n^{-\frac{1}{2}} E_n^T E_0, \tag{11}$$

where E_e is a matrix containing the (orthogonal) excitatory filters (only one in this example—the STA), E_0 contains an orthogonal basis for the remainder of the stimulus space, and E_n and D_n are the eigenvectors and eigenvalues for the remainder of the conditional covariance matrix, C_n, respectively. The first term serves to preserve the component of the stimulus in the direction of the STA, while the second term depicts a whitening (by the inverse of the *raw* stimuli in that slice) in the other dimensions.

After this conditional whitening, the stimuli are recombined and STC analysis is applied on the spike-triggered set to reestimate the filters. Figure 16B shows that following conditional whitening, there are only two significant suppressive eigenvalues corresponding to the real model filter subspace.

We have described an example of binary stimulus artifacts and partially correcting for those artifacts. There is generally no known fix for artifacts, but there are several things that can be done to check for artifacts:

1. It is helpful to examine the projection of the raw and spike-triggered stimuli onto pairs of filters recovered from the analysis; if these are not spherical, then the filters can include artifacts. However, it is important to remember that the stimulus space is huge, and projection onto two dimensions might appear spherically symmetric but does not guarantee spherical symmetry in the full space.

2. It is sometimes useful to run a model neuron simulation with the given stimuli and see if artifactual filters emerge. The simplest simulation one can run is an LNP model with a single linear filter: If a significant STC filter is found, this is indicative of an artifactual axis in simulation. Here, we have demonstrated a slightly more involved example of a divisive normalization simulation. However, it is important to realize that we have control only over the stimulus ensemble; we have no control over the nonlinear behaviors of the neural response, and the artifacts depend on these nonlinearities. We can explore in simulation nonlinearities that have been attributed to neurons, and this has proved helpful in some cases.
3. It is recommended to compare experimentally the filter subspace recovered with a given stimulus ensemble with that recovered with Gaussian stimuli (recording from the same neuron); differences in the outcome between the two stimulus types could indicate estimation biases or failures of the model.

Touryan, Felsen, and Dan (2005) compared STC analysis in area V1 for white noise and natural images. To partially correct for the natural image stimuli, they first whitened the stimuli in the ensemble. Although this cannot correct for the nonspherical nature of the stimuli, they showed that the first two eigenvectors (representing complex cells in their data) were similar for white noise and natural images. The natural images required far fewer raw stimuli to achieve the same result, probably because they are more effective at eliciting spikes. They also found additional significant (and artifactual) filters that were compared with artifactual filters arising in a simulation with natural images.

Other techniques have been designed to cope directly with non-Gaussian input, such as images, and thus bypass this limitation of the STC approach. The basic idea is quite simple: Instead of relying on a particular statistical moment (e.g., mean or variance) for comparison of the spike-triggered and raw stimulus distributions, one can use a more general comparison function that can identify virtually any difference between the two distributions. A natural choice for such a function is information-theoretic: One can compare the *mutual information* between a set of filter responses and the probability of a spike occurring (Paninski, 2003; Sharpee et al., 2003, 2004). This approach is promising because it places essentially no restriction on the stimulus ensemble. A drawback is that the estimation problem is significantly more complicated; it is more expensive to compute and may get trapped in local optima. However, it has been successfully applied to estimate one- or two-dimensional subspace models in simulation and from physiological data in response to natural images (Paninski, 2003; Sharpee et al., 2003, 2004, 2006). Other techniques, based on artificial neural networks (Lau, Stanley, & Dan, 2002; Lehky, Sejnowski, & Desimone, 1992), have also been developed and applied to natural images (Prenger, Wu, David, & Gallant, 2004).

Validation

Validation is useful to evaluate the degree to which the recovered model is an accurate description of the neural response. At the very least, it is worthwhile verifying that the model, when fit to one run of white noise stimulation, can then predict responses to another run. Because the model is a rate model, this is most directly done by measuring responses to repeated stimuli and comparing their average (the PSTH) against that predicted from the model. Another possibility is to "play back" as stimuli the eigenvectors that were found in the spike-triggered analysis to verify that they affect the neuron's response as expected (Rust, Schwartz, et al., 2005; Touryan et al., 2002). This requires that one perform the analysis and stimulus generation online during the experiment. Playing back the eigenvectors is also helpful for determining the importance of the individual model components that are recovered from the analysis; for example, the weakest components might have only a minor impact on the neural response.

It is also of interest to test how well the model generalizes to other stimuli: If one characterizes the model with a set of bars, how well does the model predict the response to a single bar? If one characterizes the model with high contrast stimuli, how well does it predict the response to low contrast stimuli? Ultimately, we would like a model that predicts the response to any arbitrary stimulus. Validating the model on different stimuli can help assess the robustness of the model and when it breaks, and, in turn, can identify the need for further improving spike-triggered analysis techniques.

Discussion

We have described a set of spike-triggered techniques for characterizing the functional response properties of neurons using stochastic stimuli. In general, there is a tradeoff between restricting the subspace dimensionality (as in the STA and STC approaches) versus restricting the nonlinearity (as in the Wiener/Volterra approaches). Here, we have focused specifically on STA and STC analyses. These methods rely on an assumption that the response of the neuron is governed by an initial linear stage that serves to reduce the dimensionality of the stimulus space. The linear stage is followed by a nonlinearity upon which we place fairly minimal constraints. Having worked with these methods in both retina and V1, we have found that many experimental and analysis issues are quite tricky. We have presented examples with model neuron simulations, highlighting similarities with experiments where possible.

Estimation of the linear subspace can be corrupted by three distinct sources of error, which we have discussed in this article. First, there are errors due to the finiteness of the data. The rate at which these decrease with increasing data is given in Equation 8 and illustrated in Figure 9. Second, there are biases that can arise from the interaction of the neural nonlinearities and use of non-Gaussian stimuli. Examples are shown in Figure 15. Finally, there are errors due to model failure.

There are a number of interesting directions for future research. First, the LNP model can be extended to incorporate some spike history dependence, by recursively feeding back the spiking output into the linear input stage. This "recursive LNP" model (also referred to as a general linear model [GLM]) has appeared in recent literature (Pillow, Paninski, Uzzell, Simoncelli, & Chichilnisky, 2005; Truccolo, Eden, Fellows, Donogue, & Brown, 2005) and may allow the introduction of some adaptation effects, as well as shorter timescale effects such as refractoriness, bursting, or rapid gain adjustments. This model can no longer be directly fit to data with STA and STC and requires more complex fitting procedures. In addition, the techniques described here can be adjusted for the analysis of multineuronal interactions (e.g., Nykamp, 2003; Okatan, Wilson, & Brown, 2005; Pillow, Shlens, Paninski, Chichilnisky, & Simoncelli, 2005b). Such methods have been applied, for example, in visual cortex (Tsodyks, Kenet, Grinvald, & Arieli, 1999), motor cortex (Paninski, Fellows, Shoham, Hatsopoulos, & Donoghue, 2004), and hippocampus (Harris, Csicsvari, Hirase, Dragoi, & Buzsáki, 2003). Also, neurons adapt to stimuli over multiple timescales (Brenner, Bialek & de Ruyter van Steveninck, 2000; Fairhall, Lewen, Bialek, & de Ruyter van Steveninck, 2001), and it would be interesting to extend current approaches to incorporate adaptation. Finally, it would be desirable to develop techniques that can be applied to a cascaded series of LNP stages. This will be essential for modeling responses in higher order sensory areas, which are presumably constructed from more peripheral responses. Specifically, if the afferent responses that arrive in a particular neural area are reasonably understood, then one may be able to arrange to perform the spike-triggered analysis in the space of the afferents (Rust, Simoncelli, et al., 2005).

Appendix

We describe how to compute STA and STC for elliptically symmetric Gaussian stimuli. If the distribution of stimuli is elliptically symmetric, then a modified STA can be computed as follows (e.g., Theunissen et al., 2001):

$$\hat{A}' = C^{-1}\hat{A}, \tag{A1}$$

where

$$C = \sum_n \vec{s}(t_n)\vec{s}^T(t_n) \tag{A2}$$

is the covariance matrix of the raw stimuli such that $C = V^T D V$ (we assume that the mean stimulus is zero). Note that this solution is a regression estimate for a linear mapping from stimuli to spikes. The surprising result is that one can use linear regression on a one-dimensional LN model if the input vectors are elliptically distributed.

As in the case of STA, STC can be generalized to the case of an elliptically symmetric stimulus distribution. Here, the natural choice is to solve for stimulus dimensions in which the *ratio* of variances of the spike-triggered and raw stimulus ensembles is either large or small. Mathematically, we write this ratio in a direction specified by unit vector $\hat{u}$ as:

$$r(\hat{u}) = \frac{\hat{u}^T \hat{C} \hat{u}}{\hat{u}^T C \hat{u}}. \tag{A3}$$

The solution to this problem can be computed directly using a generalized eigenvector analysis. Specifically, we first solve for the whitening transform in the denominator, computing the eigenvalues D and eigenvectors V of the covariance matrix of the raw stimuli. We set $X = V(\sqrt{D})^{-1}$ and $\hat{u} = X\hat{v}$, obtaining:

$$r(\hat{u}) = \frac{\hat{v}^T X^T \hat{C} V \hat{v}}{\hat{v}^T \hat{v}}. \tag{A4}$$

This is now equivalent to solving a standard eigenvector problem, calculating the eigenvalues and eigenvectors of $X^T \hat{C} X$.

Acknowledgments

We are very grateful to E. J. Chichilnisky, Tony Movshon, and Liam Paninski for helpful discussions.

Commercial relationships: none.
Corresponding author: Odelia Schwartz.
Email: odelia@salk.edu.
Address: The Salk Institute, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA.

Footnotes

[1]It should be noted that a Wiener/Volterra approach has also been applied within a subspace, but under the assumption of a low—order polynomial nonlinearity (e.g., Emerson et al., 1987, 1992; Szulborski & Palmer, 1990).

[2]Note that the STA estimate is unbiased but it does *not*, in general, correspond to a maximum likelihood estimate (Dayan & Abbott, 2001).

[3]Note that recent work (Pillow & Simoncelli, 2006) suggests an information—theoretic objective that combines the STA and STC optimally.

References

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A, Optics and Image Science*, *2,* 284–299. [PubMed]

Agüera y Arcas, B., & Fairhall, A. L. (2003). What causes a neuron to spike? *Neural Computation, 15,* 1789–1807. [PubMed]

Agüera y Arcas, B., Fairhall, A. L., & Bialek, W. (2001). What can a single neuron compute? In *Adv. Neural Information Processing Systems (NIPS*00)* (Vol. 13, pp. 75–81).

Agüera y Arcas, B., Fairhall, A. L., & Bialek, W. (2003). Computation in a single neuron: Hodgkin and Huxley revisited. *Neural Computation, 15,* 1715–1749. [PubMed]

Albrecht, D. G., & Geisler, W. S. (1991). Motion sensitivity and the contrast-response function of simple cells in the visual cortex. *Visual Neuroscience, 7,* 531–546. [PubMed]

Anzai, A., Ohzawa, I., & Freeman, R. D. (1999). Neural mechanisms for processing binocular information: I. Simple cells. *Journal of Neurophysiology, 82,* 891–908. [PubMed] [Article]

Bair, W., Cavanaugh, J. R., & Movshon, J. A. (1997). Reconstructing stimulus velocity from neuronal responses in area MT. In M. C. Mozer, M. I. Jordan, & T. Petsche (Eds.), *Adv. Neural Information Processing Systems (NIPS*96)* (Vol. 9, pp. 34–40). MIT Press.

Berry, M. J. II, & Meister, M. (1998). Refractoriness and neural precision. *The Journal of Neuroscience, 18,* 2200–2211. [PubMed] [Article]

Bialek, W., & de Ruyter van Steveninck, R. (2005). Features and dimensions: Motion estimation in fly vision. http://arxiv.org/abs/q-bio/0505003

Borghuis, B. G., Perge, J. A., Vajda, I., van Wezel, R. J., van de Grind, W. A., & Lankheet, M. J. (2003). The motion reverse correlation (MRC) method: A linear systems approach in the motion domain. *Journal of Neuroscience Methods, 123,* 153–166. [PubMed]

Brenner, N., Bialek, W., & de Ruyter van Steveninck R. (2000). Adaptive rescaling maximizes information transmission. *Neuron, 26,* 695–702. [PubMed] [Article]

Bussgang, J. J. (1952). Cross-correlation functions of amplitude-distorted gaussian signals. *RLE Technical Reports, 216.*

Chichilnisky, E. J. (2001). A simple white noise analysis of neuronal light responses. *Network, 12,* 199–213. [PubMed]

Daugman, J. G. (1989). Entropy reduction and decorrelation in visual coding by oriented neural receptive fields. *IEEE Transactions on Bio-medical Engineering, 36,* 107–114. [PubMed]

David, S. V., & Gallant, J. L. (2005). Predicting neuronal responses during natural vision. *Network, 16,* 239–260. [PubMed]

David, S. V., Vinje, W. E., & Gallant, J. L. (2004). Natural stimulus statistics alter the receptive field structure of v1 neurons. *The Journal of Neuroscience, 24,* 6991–7006. [PubMed] [Article]

Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. MIT Press.

DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1993). Spatiotemporal organization of simple cell receptive fields in the cat's striate cortex: II. Linearity of temporal and spatial summation. *Journal of Neurophysiology, 69,* 1118–1135. [PubMed]

deBoer, E., & Kuyper, P. (1968). Triggered correlation. In *IEEE Transactions on Bio-medical Engineering* (Vol. 15, pp. 169–179).

de Ruyter van Steveninck, R., & Bialek, W. (1988). Coding and information transfer in short spike sequences. In *Proceedings of the Royal Society of London B, Biological Sciences* (Vol. 234, pp. 379–414).

Eggermont, J. J., Johannesma, P. M., & Aertsen, A. M. (1983). Reverse-correlation methods in auditory research. *Quarterly Reviews of Biophysics, 16,* 341–414. [PubMed]

Emerson, R. C., Bergen, J. R., & Adelson, E. H. (1992). Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vision Research, 32,* 203–218. [PubMed]

Emerson, R. C., Citron, M. C., Vaughn, W. J., & Klein, S. A. (1987). Nonlinear directionally selective subunits in complex cells of cat striate cortex. *Journal of Neurophysiology, 58,* 33–65. [PubMed]

Fairhall, A. L., Lewen, G. D., Bialek, W., & de Ruyter Van Steveninck, R. R. (2001). Efficiency and ambiguity in an adaptive neural code. *Nature, 412,* 787–792. [PubMed]

Felsen, G., & Dan, Y. (2005). A natural approach to studying vision. *Nature Neuroscience, 8,* 1643–1646. [PubMed]

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A, Optics and Image Science, 4,* 2379–2394. [PubMed]

Geisler, D. (1992). Two-tone suppression by a saturating feedback model of the cochlear partition. *Hearing Research, 63,* 203–211.

Harris, K. D., Csicsvari, J., Hirase, H., Dragoi, G., & Buzsáki, G. (2003). Organization of cell assemblies in the hippocampus. *Nature, 424,* 552–556. [PubMed]

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience, 9,* 181–197. [PubMed]

Hochstein, S., & Shapley, R. M. (1976). Linear and nonlinear spatial subunits in Y cat retinal ganglion cells. *The Journal of Physiology, 262,* 265–284. [PubMed] [Article]

Horwitz, G. D., Chichilnisky, E. J., & Albright, T. D. (2005). Blue–yellow signals are enhanced by spatiotemporal luminance contrast in macaque V1. *Journal of Neurophysiology, 93,* 2263–2278. [PubMed] [Article]

Jones, J. P., & Palmer, L. A. (1987). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology, 58,* 1187–1211. [PubMed]

Keat, J., Reinagel, P., Reid, R. C., & Meister, M. (2001). Predicting every spike: A model for the responses of visual neurons. *Neuron, 30,* 803–817. [PubMed] [Article]

Korenberg, M. J., Sakai, H. M., & Naka, K. (1989). Dissection of the neuron network in the catfish inner retina: III. Interpretation of spike kernels. *Journal of Neurophysiology, 61,* 1110–1120. [PubMed]

Lau, B., Stanley, G. B., & Dan, Y. (2002). Computational subunits of visual cortical neurons revealed by artificial neural networks. *Proceedings of the National Academy of Sciences of the United States of America, 99,* 8974–8979. [PubMed] [Article]

Lehky, S. R., Sejnowski, T. J., & Desimone, R. (1992). Predicting responses of nonlinear neurons in monkey striate cortex to complex patterns. *Journal of Neuroscience, 12,* 3568–3581. [PubMed] [Article]

Lesica, N. A., & Stanley, G. B. (2004). Encoding of natural scene movies by tonic and burst spikes in the lateral geniculate nucleus. *The Journal of Neuroscience, 24,* 10731–10740. [PubMed] [Article]

Marmarelis, P. Z., & Marmarelis, V. Z. (1978). *Analysis of physiological systems: The white noise approach.* London: Plenum Press.

McLean, J., & Palmer, L. A. (1989). Contribution of linear spatiotemporal receptive field structure to velocity selectivity of simple cells in area 17 of cat. *Vision Research, 29,* 675–679. [PubMed]

Meister, M., Pine, J., & Baylor, D. A. (1994). Multi-neuronal signals from the retina: Acquisition and analysis. *Journal of Neuroscience Methods, 51,* 95–106. [PubMed]

Nishimoto, S., Ishida, T., & Ohzawa, I. (2006). Receptive field properties of neurons in the early visual cortex revealed by local spectral reverse correlation. *The Journal of Neuroscience, 26,* 3269–3280. [PubMed]

Nykamp, D. (2003). White noise analysis of coupled linear-nonlinear systems. *SIAM Journal on Applied Mathematics, 63,* 1208–1230.

Nykamp, D. Q., & Ringach, D. L. (2002). Full identification of a linear-nonlinear system via cross-correlation analysis. *Journal of Vision, 2*(1), 1–11, http://journalofvision.org/2/1/1/, doi:10.1167/2.1.1. [PubMed] [Article]

Okatan, M., Wilson, M. A., & Brown, E. N. (2005). Analyzing functional connectivity using a network likelihood model of ensemble neural spiking activity. *Neural Computation, 17,* 1927–1961. [PubMed]

Paninski, L. (2003). Convergence properties of some spike-triggered analysis techniques. *Network, 14,* 437–464. [PubMed]

Paninski, L., Fellows, M., Shoham, S., Hatsopoulos, N., & Donoghue, J. (2004). Nonlinear population models for the encoding of dynamic hand position signals in MI. *Neurocomputing*. Paper presented at Computational Neuroscience, Alicante, Spain, July 2003.

Pillow, J. W., Paninski, L., Uzzell, V. J., Simoncelli, E. P., & Chichilnisky, E. J. (2005). Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. *The Journal of Neuroscience, 25,* 11003–11013. [PubMed] [Article]

Pillow, J. W., Shlens, J., Paninski, L., Chichilnisky, E. J., & Simoncelli, E. P. (2005a). Modeling multineuronal responses in primate retinal ganglion cells. In *Computational and Systems Neuroscience (CoSyNe)*, Salt Lake City, UT.

Pillow, J. W., Shlens, J., Paninski, L., Chichilnisky, E. J., & Simoncelli, E. P. (2005b). Modeling the correlated spike responses of a cluster of primate retinal ganglion cells. In *Annual Meeting, Neurosciences,* Washington, DC.

Pillow, J. W., & Simoncelli, E. P. (2003). Biases in white noise analysis due to non-Poisson spike generation. In *Neurocomputing* (Vol. 52–54, pp. 109–115). Elsevier. Paper presented at Computational NeuroScience (CNS*02), Chicago, 21–25 July 2002. AQ1 AQ2

Pillow, J. W., & Simoncelli, E. P. (2006). Dimensionality reduction in neural models: An information-theoretic AQ3

generalization of spike-triggered average and covariance analysis. *Journal of Vision* (in press).

Pillow, J. W., Simoncelli, E. P., & Chichilnisky, E. J. (2003). Characterization of nonlinear spatiotemporal properties of macaque retinal ganglion cells using spike-triggered covariance. In *Annual Meeting, Neurosciences,* New Orleans.

Pillow, J. W., Simoncelli, E. P., & Chichilnisky, E. J. (2004). Characterization of macaque retinal ganglion cell responses using spike-triggered covariance. In *Computational and Systems Neuroscience (CoSyNe),* Cold Spring Harbor, NY.

Prenger, R., Wu, M. C., David, S. V., & Gallant, J. L. (2004). Nonlinear V1 responses to natural scenes revealed by neural network analysis. *Neural Networks, 17,* 663–679. [PubMed]

Reich, D. S., Victor, J. D., & Knight, B. W. (1998). The power ratio and the interval map: Spiking models and extracellular recordings. *The Journal of Neuroscience, 18,* 10090–10104. [PubMed] [Article]

Reid, R. C., & Alonso, J. M. (1995). Specificity of monosynaptic connections from thalamus to visual cortex. *Nature, 378,* 281–284. [PubMed]

Rieke, F., Warland, D., de Ruyter van Steveninck, R. R., & Bialek, W. (1997). *Spikes: Exploring the neural code*. Cambridge, MA: MIT Press.

Ringach, D. L. (2004). Mapping receptive fields in primary visual cortex. *The Journal of Physiology, 558,* 717–728. [PubMed] [Article]

Ringach, D. L., Hawken, M. J., & Shapley, R. (2002). Receptive field structure of neurons in monkey primary visual cortex revealed by stimulation with natural image sequences. *Journal of Vision, 2*(1), 12–24, http://journalofvision.org/2/1/2/, doi:10.1167/2.1.2. [PubMed] [Article]

Ringach, D. L., Sapiro, G., & Shapley, R. (1997). A subspace reverse-correlation technique for the study of visual neurons. *Vision Research, 37,* 2455–2464. [PubMed]

Rust, N. C., & Movshon, J. A. (2005). In praise of artifice. *Nature Neuroscience, 8,* 1647–1650. [PubMed]

Rust, N. C., Schwartz, O., Movshon, J. A., & Simoncelli, E. P. (2004). Spike-triggered characterization of excitatory and suppressive stimulus dimensions in monkey V1. In *Neurocomputing*. Elsevier. Paper presented at Computational NeuroScience (CNS*03), Alicante Spain, July 2003. AQ12

Rust, N. C., Schwartz, O., Movshon, J. A., & Simoncelli, E. P. (2005). Spatiotemporal elements of macaque V1 receptive fields. *Neuron, 46,* 945–956. [PubMed]

Rust, N. C., Simoncelli, E. P., & Movshon, J. A. (2005). How macaque MT cells compute pattern motion. In *Annual Meeting, Neurosciences,* Washington, DC.

Sakai, H. M., & Naka, K. (1987). Signal transmission in the catfish retina: V. Sensitivity and circuit. *Journal of Neurophysiology, 58,* 1329–1350. [PubMed]

Sakai, K., & Tanaka, S. (2000). Spatial pooling in the second-order spatial structure of cortical complex cells. *Vision Research, 40,* 855–871. [PubMed]

Schwartz, O., Chichilnisky, E. J., & Simoncelli, E. P. (2002). Characterizing neural gain control using spike-triggered covariance. In T. G. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Adv. Neural Information Processing Systems (NIPS*01)* (Vol. 14, pp. 269–276). Cambridge, MA. MIT Press.

Schwartz, O., & Simoncelli, E. P. (2001). A spike-triggered covariance method for characterizing divisive normalization models. In *First Annual Meeting, Vision Sciences Society,* Sarasota.

Sen, K., Wright, B. D., Doupe, A. J., & Bialek, W. (2000). Discovering features of natural stimuli relevant to forebrain auditory neurons. In *Society for Neuroscience Abstracts*. AQ13

Sharpee, T., Rust, N. C., & Bialek, W. (2003). Maximizing informative dimensions: Analyzing neural responses to natural signals. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Adv. Neural Information Processing Systems (NIPS*02)* (Vol. 15). Cambridge, MA. MIT Press. AQ14

Sharpee, T., Rust, N. C., & Bialek, W. (2004). Analyzing neural responses to natural signals: Maximally informative dimensions. *Neural Computation, 16,* 223–250. [PubMed]

Sharpee, T. O., Sugihara, H., Kurgansky, A. V., Rebrik, S. P., Stryker, M. P., & Miller, K. D. (2006). Adaptive filtering enhances information transmission in visual cortex. *Nature, 439,* 936–942. [PubMed]

Simoncelli, E. P., Pillow, J., Paninski, L., & Schwartz, O. (2004). Characterization of neural responses with stochastic stimuli. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (3rd ed.). MIT Press. AQ15

Smyth, D., Willmore, B., Baker, G. E., Thompson, I. D., & Tolhurst, D. J. (2003). The receptive-field organization of simple cells in primary visual cortex of ferrets under natural scene stimulation. *The Journal of Neuroscience, 23,* 4746–4759. [PubMed] [Article]

Szulborski, R. G., & Palmer, L. A. (1990). The two-dimensional spatial structure of nonlinear subunits in the receptive fields of complex cells. *Vision Research, 30,* 249–254. [PubMed]

Theunissen, F. E., David, S. V., Singh, N. C., Hsu, A., Vinje, W. E., & Gallant, J. L. (2001). Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Network, 12,* 289–316. [PubMed]

Touryan, J., Felsen, G., & Dan, Y. (2005). Spatial structure of complex cell receptive fields measured with natural images. *Neuron, 45,* 781–791. [PubMed] [Article]

Touryan, J., Lau, B., & Dan, Y. (2002). Isolation of relevant visual features from random stimuli for cortical complex cells. *The Journal of Neuroscience, 22,* 10811–10818. [PubMed] [Article]

Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P., & Brown, E. N. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *Journal of Neurophysiology, 93,* 1074–1089. [PubMed] [Article]

Tsodyks, M., Kenet, T., Grinvald, A., & Arieli, A. (1999). Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science, 286,* 1943–1946. [PubMed]

Volterra, V. (1959). *Theory of functionals and of integro and integro-differential equations*. New York: Dover.

Wiener, N. (1958). *Nonlinear problems in random theory*. New York: Wiley.

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

AQ1: Please provide city and state for this affiliation, web address (URL), if available.
AQ2: Please provide web address (URL) and e-mail address, if available.
AQ3: Please provide city and state for this affiliation, web address (URL) and e-mail address if available.
AQ4: Please check if these keywords are appropriate (none were provided).
AQ5: Please check if the web address (URL) is correct.
AQ6: Please provide editor name(s) as well as publisher name and location.
AQ7: Please provide publisher location.
AQ8: Please provide publisher name and location.
AQ9: Chichilnisky (2001a) and Chichilnisky (2001b) are the same; hence, only one was captured (and was thus changed to Chichilnisky, 2001).
AQ10: Please provide editor name(s) and publisher location.
AQ11: Please provide updated publication status (e.g., volume number, page range).
AQ12: Please provide editor name(s), publisher location, and page range.
AQ13: Please provide editor name(s), page range, and publisher name and location.
AQ14: Please provide page range.
AQ15: Please provide page range and publisher location.

END OF AUTHOR QUERIES