

Presented at: The Annual Computational Neuroscience Meeting
Alicante, Spain, 5-9 July 2003

To appear in *Neurocomputing*, Elsevier, 2004.

Spike-triggered characterization of excitatory and suppressive stimulus dimensions in monkey V1

Nicole C Rust^{1*}, Odelia Schwartz², J Anthony Movshon¹ and Eero Simoncelli¹

¹Center for Neural Science, New York University, New York, NY

²Salk Institute, La Jolla, CA

Abstract

Neurons in primary visual cortex are commonly characterized using linear models, or simple extensions of linear models. Specifically, V1 simple cell responses are often characterized using a rectified linear receptive field, and complex cell responses are often described as the sum of squared responses of two linear subunits. We examined this class of model directly by applying spike-triggered covariance analysis to responses of monkey V1 neurons under binary white noise stimulation. The analysis extracts a low-dimensional subspace of the full stimulus space that is primarily responsible for generation of the neural response, including both excitatory and suppressive components. We found no fewer than two excitatory dimensions in simple cells, and as many as seven dimensions in complex cells. For all cells, we also found suppressive dimensions that were at least equal in number to the excitatory dimensions. These results suggest that extensions to standard models are required to fully describe the response properties of cells in V1.

Keywords: direction tuning; spike-triggered covariance; primary visual cortex; complex cell; simple cell

Primary visual cortex (V1) is the first visual area to contain neurons selective for the orientation and direction of a moving stimulus. It is also the first site where phase-invariant complex cells are found. Although much emphasis has been placed upon examining the details of these computations, the mechanisms underlying them are not fully characterized. Potential mechanisms can be constrained by models that accurately characterize the response properties of each cell type found in V1.

Using appropriate stochastic stimuli, one can recover a model of a neuron in terms of a spatiotemporal linear filter (the spike triggered average, STA), a nonlinear function that relates the output of this filter to firing rate, and Poisson spike generation; such a model is commonly referred to as "LNP" [1,2]. This approach is appropriate when a neuron is well described by a single linear filter and asymmetric nonlinearity (e.g. rectification) but provides an incomplete description when either condition fails. An extension to this first-order technique has been

* Corresponding author: rust@cns.nyu.edu

introduced, spike-triggered covariance (STC) [3]. Because STC considers all axes that influence the variance of firing rate statistics, this technique recovers a *set* of linear filters in the presence of both asymmetric and symmetric (e.g. squaring) nonlinearities. STC has been shown to be capable of mapping the subunits of phase-insensitive complex cells [4] as well as both suppressive and excitatory influences [2,5]. In this paper, we describe the details of the application of this technique to data obtained from V1 neurons and the LNP response models that result.

1. Recovery of the linear subspace using spike-triggered covariance (STC)

We applied a spike triggered covariance analysis (STC) to data collected from isolated directionally tuned simple cells and complex cells within primary visual cortex (V1) of anesthetized, paralyzed macaque monkeys (experimental preparation details described by [6]). Stimuli were extended temporal sequences in which each frame contained a set of parallel non-overlapping dark and bright bars with randomly assigned intensity. The orientation of the bars was aligned with the cell's preferred orientation and confined to the classical receptive field. The number of bars (8-32) was chosen to match cell's preferred spatial frequency such that 4-8 bars fell within each spatial period and the whole array spanned the classical receptive field. Each frame had a duration of 10 msec.

We denote the set of stimuli that were presented over some fixed time interval preceding the n th spike as $S_n(x,t)$. In the conventional procedure of reverse correlation, one averages these stimuli to obtain the spike-triggered average (STA):

$$STA(x,t) = \frac{1}{N} \sum_n S_n(x,t),$$

where N indicates the number of spikes. If one assumes that the neural response is generated by projection onto a single linear kernel, followed by a static nonlinearity and Poisson spike generation, the STA provides an unbiased estimate of this kernel [1,7].

But if the neural response depends on more than a single axis within the stimulus space, the STA will provide an insufficient and possibly misleading description. A number of authors have suggested the natural extension of examining higher-order statistical properties (and in particular, the covariance) of the spike-triggered ensemble of stimuli [3,4,5,8,9]. The idea is simple and intuitive: if the neural response is determined by a projection onto a low-dimensional subspace (within the space of all stimuli), an analysis of the spike-triggered covariance might allow us to recover this subspace.

In particular, consider first a stimulus direction to which the neuron is insensitive. Since the strength of the stimulus in that direction has no effect on the spiking behavior of the neuron, one expects the mean and variance (or any other statistic) of the spike-triggered stimulus ensemble along that direction to be essentially unchanged relative to that of the raw stimulus ensemble. Next consider a stimulus direction along which the cell's response grows monotonically. Here the mean of the spike-triggered stimulus ensemble will shift, relative to that of the raw ensemble. If the cell's response along this axis were symmetric (e.g., rectified, or squared), then one would expect the variance of the spike-triggered stimulus ensemble to increase. On the other hand, if the presence of a stimulus component along the direction is symmetric but suppressive, one expects the variance of the spike-triggered stimulus ensemble to decrease.

In our analysis, we first compute the STA and project it out of the stimulus ensemble. Specifically, we compute the normalized (unit vector) STA, the $nSTA$, and define:

$$S_n'(x, t) = S_n(x, t) - \left[\sum_{x, t} S_n(x, t) \cdot nSTA(x, t) \right] \cdot nSTA(x, t).$$

This ensures that the axes obtained in the STC analysis will be orthogonal to the STA, and helps to avoid unwanted interactions between the STA and STC. We then compute the spike-triggered covariance:

$$COV(x_1, x_2, t_1, t_2) = \frac{1}{N_s - 1} \sum_n S_n'(x_1, t_1) S_n'(x_2, t_2).$$

If one interprets this as a matrix, with the parameter pairs $\{x_1, t_1\}$ and $\{x_2, t_2\}$ corresponding to the two indices, then this matrix captures the variance of a collection of samples in all possible directions within the space of stimuli. The surface swept out by a vector whose length is equal to the variance in its own direction is a hyper-ellipsoid. The set of orthogonal eigenvectors of the STC matrix, obtained by a principal components analysis (PCA), correspond to the fundamental axes of this ellipsoid, and the eigenvalues give the variances along these axes.

The relevant excitatory and suppressive dimensions are identified with eigenvalues that deviate significantly from the variance expected for a randomly chosen set of stimuli blocks by nested bootstrap hypothesis testing. These high- and low-variance axes, together with the STA, constitute a coordinate system for those aspects of the stimulus that are relevant to the neuron. The range of eigenvalues one expects when randomly selecting N samples of d -dimensional noise vectors is proportional (asymptotically) to $\sqrt{d/N}$. The implication for these experiments is that mildly excitatory or suppressive axes will only be revealed when d/N is small enough. As a rough rule of thumb, we aim to gather at least $N=100*d$ spikes.

Analysis results for a representative cell are shown in figure 1A. This cell, classified as complex due to its lack of response modulation to grating stimuli, had a relatively flat STA. STC revealed 6 significant excitatory and 6 significant suppressive eigenvectors or filters (E), all with clear space-time oriented structure. The excitatory filters exhibited the same space-time orientation, indicating a similar direction preference for moving stimuli. Suppressive filters were tilted to the opposite orientation of the excitation, indicating a preference for motion in the opposite direction. Examination of the structure of the eigenvectors suggests that filters with similar structure were revealed in pairs that were phase-shifted relative to one another (E1-E2, E3-E4, E5-E6). In each case, the first pair appeared to have spatial structure centered approximately in the middle of the receptive field whereas the structure of subsequent pairs flattened in the middle and was pushed toward the edges. The recovered weights for each filter (see section 2 for details of fit) also suggested a paired arrangement.

The energy model of a complex cell proposes that the response arises from the summed squared response of the stimulus projection onto a quadrature pair of filters [10]. Due to the symmetry of the squaring operation, a zero-valued STA is predicted by this model and is realized in our results. However, STC applied to an energy model would resolve only two linear filters as relevant dimensions. As suggested by our chosen example cell, we consistently observed more than 2 excitatory filters in these cells in addition to multiple suppressive filters. Similarly, traditional

models of simple cells include a single linear filter followed by a rectifying nonlinearity, however, we consistently observed multiple excitatory and suppressive filters in these cells (data not shown). These results suggest that extensions of these traditional models are needed to fully capture firing rate statistics in these neurons.

Regarding these filters, it is important to note a few cautionary details concerning this analysis. Most importantly, the kernels (eigenvectors) should not be taken literally as physiologically instantiated mechanisms, since the analysis forces them to be orthogonal. One should think of them as a representation of a set (subspace) of stimuli that govern the response of a cell. More specifically, the labeling of any particular axis as "excitatory" or "suppressive" is based on whether the spike-triggered ensemble has a smaller or larger variance than the raw stimulus ensemble along that axis. If excitation and suppression coincide within a neuron along a common axis, the analysis might produce an axis of either label, or might even produce two orthogonal excitatory and suppressive axes. These concerns may be ameliorated by careful examination of the spike-triggered ensemble as projected on the axes obtained in the analysis. We show examples of this in the next section.

2. Recovery of the nonlinear combination function

A full model of a neuron includes not only the N filters revealed by the STC analysis but also the N -dimensional nonlinear function that describes the rule by which they are combined to generate spikes. Unfortunately, we simply cannot collect enough data to reconstruct this function, given that N is as large as 8-12 dimensions. However, the firing rate as a function of the projection onto any single dimension can be collapsed across all others and reconstructed as the quotient of the histogram of the number of spikes and the number of stimuli at each projection value. Likewise, the firing rate as a function of the joint projection onto 2 dimensions can be recovered in a analogous manner.

Upon examining these joint 2-D firing rate functions along different pairs of axes within the excitatory or suppressive pool, we observed that they took on a characteristic form: contours of constant firing rate along these pairs traced out ellipses and circles, implying that the firing rate is based on the sum of squares of the filter responses (figure 1B). As such, we assumed a model in which responses of excitatory and suppressive filters were each pooled using a weighted sum of squares (an L2 norm) where the weights determine the elongation of the axes of the hyper-ellipse. The STA was included in the excitatory pool, but was half-wave rectified before being squared and combined with other responses.

We obtained the weights for each dimension by maximizing the mutual information between the weighted sum-squared output of the joint excitatory and suppressive pools and the spikes. This approach makes no assumptions regarding the form interaction between excitation and suppression, but instead allows for the reconstruction of this function. The weights for each filter recovered from the example cell are provided in figure 1A.

The firing rate as a function of the joint output of the excitatory and suppressive pool was reconstructed as the quotient of 2-D histograms of the number of spikes and the number of total stimuli in each bin (figure 1B, right). This function reveals the nature of the interaction between excitation and suppression in V1 and thus constrains mechanistic descriptions of these neurons. If the excitatory and suppressive pools interacted in a purely subtractive manner, contours of constant firing rate would result in parallel lines. If the response could be described as the ratio of the excitatory and suppressive pools, contours of constant firing rate would emanate from the

origin. For the example cell, these contours appear to take on an intermediate behavior: the contours change in slope but fail to pass through the origin. This could be due to a combination of divisive and subtractive mechanisms or could be explained by a divisive mechanism that acts with a constant in its denominator; such equations have been used to describe contrast gain control in V1 [11].

3. Discussion

We have outlined a method for constructing a spatio-temporal model of a V1 neuron that describes the transformation of a time-varying stimulus into firing rate. In the resulting model, the predicted response to a stimulus is obtained by convolving the stimulus with each of the recovered linear filters (figure 1A), and combining the outputs of the excitatory and suppressive filters into separate pools via weighted sum of squares. The instantaneous firing rate corresponds to the appropriate bin in the reconstructed 2-D lookup table that describes the interaction between the excitatory and suppressive pools (figure 1B, right).

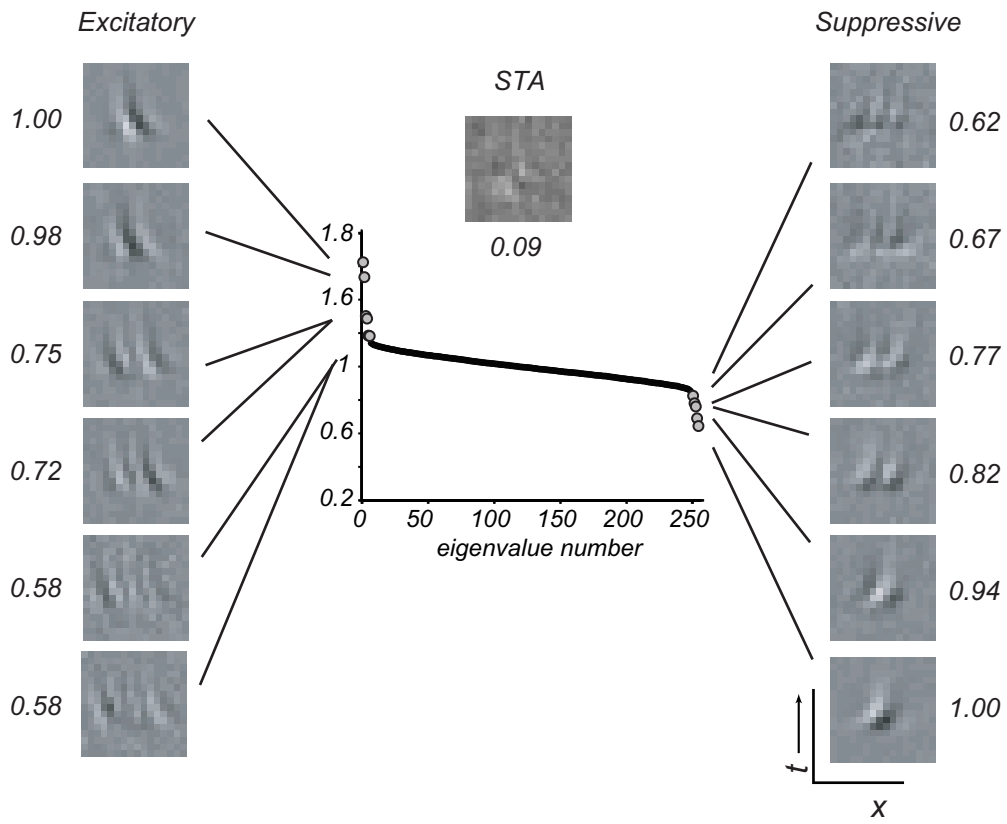
Surprisingly, application of this approach to V1 neurons revealed more linear filters than are suggested by traditional models. Although one must be cautious when relating the results of STC to mechanistic descriptions, characterization of V1 neurons in this way can potentially discriminate between models that produce similar responses to traditional stimuli. Furthermore, once the subspace that the neuron cares about is defined by STC, these results can be used to construct new stimuli optimized to discriminate between classes of models.

4. References

- [1] E.J. Chichilnisky, A simple white noise analysis of neuronal light responses, *Network* 12(2001) 199-213.
- [2] E.P. Simoncelli, J. Pillow, L. Paninski, and O. Schwartz, Characterizing receptive fields with stochastic stimuli, in: M. Gazzaniga, ed., *The newest cognitive neurosciences*, 3rd edition (MIT Press, Boston MA, 2004), to appear.
- [3] R. de Ruyter van Steveninck, W. Bialek, Real-time performance of a movement-sensitive neuron in the blowfly visual system, *Proc. R. Soc. Lond. B* 234(1988) 379-414.
- [4] J. Touryan, B. Lau, Y. Dan, Isolation of relevant features from random stimuli in cortical complex cells, *J. Neurosci.* 22(2002) 10811:10818.
- [5] O. Schwartz, E.J. Chichilnisky, E.P. Simoncelli, Characterizing gain control using spike-triggered covariance, *Advances in neural information processing systems* 14 (2002) 269-276.
- [6] J.R. Cavanaugh, W. Bair, J.A. Movshon, Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons, *J. Neurophysiol.* 88(2002) 2530-2546.
- [7] L. Paninski, Convergence properties of three spike-triggered analysis techniques, *Network* 14(2003) 437-464.
- [8] N. Brenner, W. Bialek, R. de Ruyter van Steveninck, Adaptive rescaling maximizes information transmission, *Neuron* 26(2000) 695-702.
- [9] A. Arcas, A. Fairhall, W. Bialek, Computation in a single neuron: Hodgkin and Huxley revisited, *Neural Computation* 15(2003) 1715-1749.
- [10] E.H. Adelson, J.R. Bergen, Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A* 2(1985) 284-299.
- [11] D.J. Heeger, Normalization of cell responses in cat striate cortex. *Vis. Neurosci.* 4(1992) 181-197.

Figure 1 Reconstruction of a multi-dimensional LNP response model for an example V1 complex cell. A. Spike-triggered covariance analysis. The spike-triggered average (STA) is plotted in spatial (X axis), and temporal (Y axis) coordinates where time increases relative to the events preceding a spike. The eigenvalues resulting from a PCA applied to the spike-weighted stimulus ensemble covariance matrix are plotted in rank order. Eigenvalues with gray centers correspond to axes along which the spike-triggered stimulus ensemble variance is significantly different than expected by chance as assessed by nested bootstrap hypothesis testing. The eigenvectors corresponding to significant eigenvalues are shown. Eigenvectors that have a significantly larger variance are considered excitatory; eigenvectors with a significantly smaller variance are considered suppressive. The contrast of each filter is scaled according its weight; the weight of each filter is labeled (see section 2 for details on weight fitting). Excitatory and suppressive filters are independently normalized with the strongest filter set to a value of 1. B. The left three plots illustrate firing rate as a function of the projection along different pairs of axes. Pixel intensity is set proportional to firing rate with contours of constant firing rate indicated to demonstrate their elliptical nature. The rightmost plot displays firing rate as a function of the output of the excitatory and suppressive pools where each pool is comprised of a weighted sum of squares of the filters in panel A (with the STA half-squared).

A



B

