Published in final edited form as:
Annu Rev Neurosci. 2013 July 8; 36: 103–120. doi:10.1146/annurev-neuro-062012-170253.

Computational Identification of Receptive Fields

Tatyana O. Sharpee[1,2]
Tatyana O. Sharpee: sharpee@salk.edu
[1]Computational Neurobiology Laboratories, Salk Institute for Biological Studies, La Jolla, California 92037
[2]Center for Theoretical Biological Physics, University of California at San Diego, La Jolla, California 92093

Abstract

Natural stimuli elicit robust responses of neurons throughout sensory pathways, and therefore their use provides unique opportunities for understanding sensory coding. This review describes statistical methods that can be used to characterize neural feature selectivity, focusing on the case of natural stimuli. First, we will discuss how such classic methods as reverse correlation/spike-triggered average and spike-triggered covariance can be generalized for use with natural stimuli to find the multiple "relevant" stimulus features that affect the responses of a given neuron. Second, ways to characterize neural feature selectivity while assuming that the neural responses exhibit a certain type of invariance, such as position invariance for visual neurons will be discussed. Finally, we discuss methods do not require one to make an assumption of invariance, and instead can determine the type of invariance by analyzing relationship between the multiple stimulus features that affect the neural responses.

Keywords

natural stimuli; spike-triggered average; maximum likelihood; mutual information; neural networks

INTRODUCTION

One way to understand how the brain works is to describe the function of each of its neurons. In sensory systems, to describe a neuron's function means to create (either explicitly or implicitly) a model that can predict the neural responses to novel stimuli. Ultimately, the goal is to predict a neuron's responses to "natural stimuli," i.e., stimuli that are taken from an animal's environment (or to approximate such stimuli) (Felsen & Dan 2005). However, one can gain significant understanding of the function of neural pathways by using simplified stimuli (Rust & Movshon 2005). In fact, much of our current understanding about the function of visual pathways has been obtained using reduced parametric stimuli, such as spots of light, edges and bars, curved contours, and elements of three-dimensional shapes. Studies using parametric stimuli have led to such fundamental insights as the establishment of orientation selectivity in the primary visual cortex (V1) as well as tuning for curvature and orientation in three dimensions at subsequent stages of visual processing (Anzai et al. 2007, Bakin et al. 2000, Desimone & Schein 1987, Hubel &

DISCLOSURE STATEMENT
The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

Wiesel 1968, Kuffler 1953, Kunsberg & Zucker 2012, Li & Zaidi 2004, McManus et al. 2011, Pasupathy & Connor 1999). The use of parametric stimuli has many advantages. Mainly, if the stimuli can be parameterized with a small number of parameters, then the corresponding stimulus set can be probed well experimentally, resulting in models with high predictive power, at least with the parameterized stimulus set. However, the use of parametric stimuli in some sensory areas, especially those beyond V1, presents some difficulty, namely that the relevant set of parameters is either unknown or of such high dimensionality that fully sampling it is no longer feasible. As an example, one can think of the many parameters that are needed to describe facial features and expressions. In such cases, computational approaches to characterize neural feature selectivity become indispensable.

In parallel to work in vision, research in the auditory modality since the 1950s has relied much more heavily on the use of computational approaches to describe neuronal function. The basic idea is to use stimulus sets that are defined by their statistical properties, such as the mean, variance, or correlation structure, but are otherwise unconstrained. Thus, instead of using a few parameters, such as the orientation or length of a bar, to specify each particular stimulus, investigators use a few parameters to specify the properties of an entire stimulus distribution. Examples of the resulting stimulus ensembles include “white-noise” stimuli for which the mean and variance are specified but responses at different times or across different spatial locations are uncorrelated. Adding correlations between stimulus values at different times, frequencies, or locations yields ensembles of correlated Gaussian stimuli. Again, the correlations between different stimulus values can be described by a small number of parameters, such as how fast the correlations decrease with the increasing distance between pixels.

The use of statistical stimulus sets has its own advantages. First, because the stimulus set is not optimized for a given neuron, such stimuli are ideally suited for multielectrode recordings that are becoming increasingly common. Second, the use of such stimuli enables researchers to uncover types of neural feature selectivity that were not part of the original hypothesis. The use of statistical stimulus ensembles also permits investigators to proceed without a good starting hypothesis about the types of stimulus features that are relevant in a given sensory region. As a result, compared with models obtained using parametric stimuli, models derived using statistical stimulus sets usually have better predictive power when they are applied to predict the neural responses in another stimulus context.

How do statistical approaches work in general? The basic idea common to all these methods is as follows. Prior to conducting an experiment, researchers do not know which stimulus features will modulate the responses of the neuron under consideration. The goal is to find these features, referred to as the “relevant stimulus features,” because they may either increase or decrease the neural spike probability relative to its average value. To find the relevant stimulus features, one can present a large number of stimuli (~20,000 to ~50,000 different images or sound patterns, which is roughly equivalent to a typical sensory episode of ~1 h). Although it is possible that none of these patterns exactly matches a particular neuron’s relevant feature(s), many (or at least some) of the stimuli will be sufficiently close to optimal to elicit some neural responses. Once the neuronal responses to a large number of different stimuli have been recorded, the relevant stimulus features for a given neuron, as well as its preferred optimal stimulus, can be deduced by analyzing how potentially subtle changes in the neuronal firing rate are related to changes in the corresponding stimuli. Typically, if a stimulus sequence contains ~10^4 stimuli, then ~10^3 spikes will be collected, and this will be sufficient to map out the profile of the relevant stimulus features on a grid of ~10^2 points. Notably, stimuli do not need to be repeated multiple times to enable analysis of the potentially subtle changes in the neuronal firing rate with small changes in the stimulus.

In fact, for most of the techniques described below, presenting many similar (but not identical) stimuli just once is preferable to having any chosen stimulus presented multiple times, even though multiple presentations of the same stimuli allow us to average out the neuronal noise. This is because presenting many similar stimuli allows researchers not only to average out the neuronal noise (assuming some continuity in the stimulus/response function of a given neuron), but also to better probe the stimulus/response function at intermediate points.

This article discusses the advantages and limitations of statistical techniques that are currently available for characterizing neural feature selectivity and that use noise-like and natural stimuli. Particular attention is given to the methods used to characterize neural resposes to natural stimuli, because such stimuli are often the only type that can drive robust responses in high-level sensory areas. Largely omitted from this discussion are techniques that can lead to a more effective use of experimental time. The reader is directed to a number of recent and excellent reviews (Huys & Paninski 2009, Lewi et al. 2009, Paninski et al. 2007) on how to optimize the order in which stimuli should be presented to maximize the accuracy of derived models given the limitations of the length of the recording.

RECEPTIVE FIELD

The concept of a receptive field (RF) was first introduced in somatosensation to describe a part of the body surface where the reflex can be elicited (Sherrington 1906). In sensory systems, the term became much more widely known after Hartline (1938) used it to describe the firing properties of the retinal ganglion cells (RGCs). One way to measure the RF of an RGC in its original formulation is to plot the neuronal firing rate as a function of light position. Another way is to plot the pattern of light intensities that, when presented on a screen, would elicit the maximal firing rate from this neuron. If the neuron is modeled as a linear system, then the two ways of measuring the RF are equivalent. However, because the neuronal firing rate inevitably exhibits at least some nonlinear effects, for example, because it cannot be negative, the two interpretations of the RF concept will differ. Research has shown that the second formulation, whereby the RF is interpreted as the optimal stimulus for the neuron, is much more amenable to the generalizations necessary to capture a rich variety of nonlinear and contextual effects observed for sensory neurons.

LINEAR MODEL

To predict the firing rate of a neuron to a novel stimulus using a linear model, one can compare how similar that stimulus is to the optimal pattern, i.e., the RF. Mathematically, this corresponds to multiplying stimulus values pixel by pixel by the RF values and summing across all pixels. In this interpretation, the RF becomes the weighting function according to which stimulus values are combined to obtain the firing rate (Figure 1). The linear model also includes the coefficient of proportionality between the stimulus similarity to the RF and the neural firing rate. This coefficient of proportionality, referred to as the “gain,” is the same for all stimuli.

Before discussing various ways for building nonlinear models of neural responses, it is useful to explore other ways of thinking about the RF concept. If the RF has D pixels (which could include temporal profiles), then it can also be represented as a vector in a D-dimensional space. To compare each new stimulus to the RF, stimuli should be defined on the same grid of pixel values as the RF. Then, each stimulus can be considered as a vector in the same D-dimensional space. The mathematical procedure described above of weighting each stimulus value by the RF profile corresponds to the computation of a dot product between the RF and the stimulus for which we would like to obtain the firing rate prediction. In geometrical terms, this corresponds to taking a projection of a vector that describes the

stimulus onto the vector that describes the RF (Figure 1). In other words, this procedure corresponds to either finding a stimulus component along the RF or filtering stimuli by the RF.

Spike-Triggered Average

These geometrical interpretations suggest ways to find the RF from neural recordings. When only the stimulus component along the RF affects the neural firing rate, averaging all stimuli that have elicited a spike will result in an averaging out of all the stimulus components along directions in the stimulus space other than that of the RF. This spike-triggered average (STA) should then yield a vector that is proportional to the RF (Figure 2). The STA could also contain the average stimulus that is not associated with spiking, which in many cases equals zero. If the mean of all stimuli (both those that elicited and those that did not elicit a spike) is nonzero, then this term should be subtracted from the STA to yield an estimate of the neuron's RF.

One can prove that this intuition for computing the STA is mathematically rigorous if the stimulus ensemble is "circularly symmetric," meaning that the stimulus ensemble probes the neuron's responses in all directions equally (Chichilnisky 2001). If so, the STA yields an unbiased estimate of the neuron's RF that will converge to the true RF as longer recordings are obtained. One example of a circularly symmetric stimulus distribution is the so-called white-noise stimulus ensemble, which has zero mean and independent variations along all stimulus dimensions that follow a Gaussian distribution. More generally, the probability distribution of observing a stimulus with a certain amplitude can be non-Gaussian. Such distributions are also of relevance. For example, analyses of image intensities in the natural environment may follow a Laplace distribution (Ruderman & Bialek 1994, Simoncelli & Olshausen 2001). However, as long as the distribution is circularly symmetric, the STA will correspond to the RF asymptotically.

What happens if the stimulus ensemble is not circularly symmetric? For example, in the natural environment, image intensities at nearby locations are often positively correlated with each other (Field 1987). As a consequence, the sum of intensities at the two locations will cover a much broader range of values than their difference. Stimulus ensembles with such correlations are not circularly symmetric. In other words, variance is not equal along different dimensions in the stimulus space (Figure 3*b*). How does this affect our ability to estimate the relevant stimulus features by computing the STA? Consider the following hypothetical example: The neuron's spikes are triggered when the light intensity at one location exceeds a certain threshold value. The light intensities at this location are correlated with light intensities at nearby locations. The STA will, therefore, show a peak in the light intensity at the location that is relevant for eliciting neuronal spikes. However, the nearby locations will also show significant deviations from zero (Figure 3). In technical terms, the STA provides a biased estimate of the relevant stimulus feature. Even if we collect an infinite amount of data, the STA will not provide a correct estimate of the relevant stimulus feature in cases where stimulus values are correlated across different dimensions. However, it is possible to compensate for the presence of stimulus correlations with one relatively simple step. In the linear model, the effect of stimulus correlations is limited to pairwise correlations between different stimulus values, which are specified by the stimulus covariance matrix. To obtain an unbiased estimate of the relevant stimulus dimension from the STA, one simply needs to multiply it by the inverse of the stimulus covariance matrix (Rieke et al. 1997; Theunissen et al. 2000, 2001):

$$v_i = C^{-1}_{ij}(\mathrm{STA})_j, \quad (1)$$

where v_i are components of the relevant stimulus feature, matrix C_{ij} is the stimulus covariance matrix, $(\mathrm{STA})_j$ represents components of the STA, and summation over the repeated index j is implied. The covariance matrix is obtained by averaging stimulus deviations from the mean. This procedure is known as decorrelation or deconvolution. The resultant vector is termed the decorrelated STA (Rieke et al. 1997; Theunissen et al. 2000, 2001) and is analogous to deconvolution that is sometimes done in microscopy to obtain a deblurred image by compensating for the microscope's point spread function (Press et al. 1992). The stimulus covariance matrix is the analogue of the point spread function in the context of neural coding. It captures the expected stimulus values at nearby locations given a "point source" at the chosen location.

Although correcting the STA for stimulus correlations according to Equation 1 is a simple linear operation, practical implementations can be difficult because a very strong asymmetry is often found in the range of stimulus values contained within the input ensemble. This is especially true for stimuli derived from the natural environment. Here, the intensities averaged across space (the image) or time have a much wider range of values compared with the range of values explored by their differences. More generally, components of natural stimuli at high spatial and temporal frequencies are much less well sampled than are components corresponding to low spatial and temporal frequencies (Field 1987). Because of this asymmetry, some stimulus dimensions are much less well probed than others. The process of compensating for the difference in sampling along different stimulus dimensions according to Equation 1 amplifies noise at the stimulus dimensions that were less well probed. In mathematical terms, the covariance matrix C is ill-defined, and its eigenvalues have widely different amplitudes. In the context of natural stimuli, small eigenvalues of the covariance matrix correspond to high temporal and spatial frequencies and reflect the small power found at these frequencies.

The division of the STA by the stimulus covariance matrix amplifies noise at the components that are underrepresented in the stimulus. To overcome these issues, one can perform "regularization." This process is based on the observation that at some frequencies the amount of noise will exceed the measured signal; for those frequencies, it is better to assume that the signal is equal to zero. Deciding which stimulus components are sufficiently well sampled to be included often has to be determined on a neuron by neuron basis and in some cases can introduce additional biases that are not present in the decorrelated STA computed without regularization according to Equation 1 (Sharpee et al. 2008). Several advanced statistical techniques have been developed to tackle this issue by incorporating priors such as STA smoothness and sparsity (Ahrens et al. 2008, Christianson et al. 2008, Paninski et al. 2007, Park & Pillow 2011b, Sahani & Linden 2003).

LINEAR-NONLINEAR MODEL

The linear model can often well characterize neural feature selectivity by specifying the neuron's RF (Theunissen et al. 2000, 2001). However, the linear model cannot account for many nonlinear and contextual effects in the neural response (Gilbert & Wiesel 1990, Nothdurft et al. 1999, Schwartz et al. 2007, Series et al. 2003, Sharpee & Victor 2008, Sillito & Jones 1996, Zipser et al. 1996). One way to generalize the linear model of neural feature selectivity to capture some of the nonlinear aspects of the neural computation is to assume that these nonlinear effects are in some sense weak. Then, one can write the nonlinear transformation as a power series containing linear, quadratic, and higher-order terms. This expansion corresponds to the Volterra/Wiener series approximation (Marmarelis & Marmarelis 1978, Victor & Purpura 1998, Victor & Shapley 1979). Given enough terms in the expansion, the series is guaranteed to approximate any arbitrary function well. However, in practice, this approximation can extend only to linear and quadratic terms. Thus, in the

Wiener approach, only quadratic functions of the stimulus can be modeled. Unfortunately, neural responses often contain much sharper nonlinearities than can be described by a quadratic function. For example, in a simple threshold model where the neuron produces a spike only when the stimulus value exceeds a certain value, the quadratic function is too smooth to approximate the nonlinearity well.

A very elegant, simple, yet agile way to capture sharp nonlinearities in the neural response is provided by the so-called linear-nonlinear (LN) model (de Boer & Kuyper 1968, Meister & Berry 1999, Victor & Shapley 1980). In statistics, this model is also known as the generalized linear model (Weisberg & Welsh 1994). Whereas the linear model requires that the firing rate of a neuron depend on the stimulus similarity to the RF (computed as the stimulus projection onto the RF, a purely linear operation), the LN model allows the firing rate to be an arbitrary nonlinear function of the stimulus projection on the RF (Figure 1). This function is often referred to as the nonlinear gain function to emphasize that the gain between the firing rate and the stimulus projection onto the RF now depends on this projection. In the LN model, the RF is often referred to as the filter or the relevant stimulus dimension.

The LN model has many computational advantages. For example, its linear component---the relevant dimension representing the RF---can be found using the linear techniques described above if the stimulus ensemble is circularly symmetric (Chichilnisky 2001, de Boer & Kuyper 1968) or, for certain nonlinearities, in the more general case of finite energy band-limited (Lazar & Slutskiy 2012, Victor & Knight 2003, Victor et al. 2006). The circularly symmetric stimulus includes the case of the white-noise Gaussian stimulus without correlations. The linear techniques also work for determining the linear part of the LN model if stimuli are correlated but are Gaussian (Ringach et al. 2002, Sharpee et al. 2004) or, for models with specific nonlinearities. In this case, the RF is computed as the decorrelated STA according to Equation 1 (Theunissen et al. 2000, 2001). If necessary, regularization may also be used. However, as in the case of the linear model, doing so may introduce biases into the RF estimates (Sharpee et al. 2008). The mathematical proof that the linear component of the LN model can be estimated with linear techniques in the case of Gaussian stimuli relies on the special property of Gaussian stimuli: Here, the average of any number of Gaussian variables is equal to the sum of products of pairwise averages between the variables (de Boer & Kuyper 1968). In other words, by specifying the stimulus covariance matrix, researchers can fully explain the statistics of stimulus correlations in terms of pairwise correlations between different stimulus values.

In a more general case, where correlations of higher than second order cannot be predicted from the knowledge of pairwise stimulus correlations, the linear techniques do not provide unbiased estimates of the RFs of the LN model. In effect, the procedure for estimating the linear and the nonlinear parts can no longer be done independently of one another, as was done for the Gaussian stimuli. Importantly, for sensory neuroscience, natural stimuli in visual, auditory, or olfactory modalities exhibit strong non-Gaussian correlations that extend beyond the second order (Ruderman 1997, Simoncelli 2003, Singh & Theunissen 2003, Vickers et al. 2001). The presence of strong higher-order correlations in the stimulus ensemble is thought to be driven by the fact that natural stimuli are composed of objects. By contrast, the presence of Gaussian second-order correlations typically yields “cloud-like” stimulus patterns that are devoid of edges and object boundaries (Field 1987). Examples of such stimuli are provided in Figure 3*b*.

Maximally Informative Dimensions

To characterize the feature selectivity with natural stimuli and other stimuli with non-Gaussian correlations, the presence of higher-order stimulus correlations in such stimulus

ensembles must be taken into account. One approach is to evaluate the relevance of different stimulus dimensions for eliciting the neural response using measures that do not rely exclusively on the first- and second-order moments of the stimulus distributions. The Kullback-Leibler (KL) distance between probability distributions provides such a measurement (Cover & Thomas 1991). When the two distributions are the same, this distance is zero. In addition, when the KL distance is computed between two probability distributions, one of which reflects all stimuli $P(\vec{s})$ and the other stimuli that elicited a spike $P(\vec{s} \mid spike)$, it corresponds to the mutual information between stimuli and the neural spike rate (Brenner et al. 2000):

$$I(\vec{v}) = \int d\vec{s}\, P(\vec{s}\,|spike) \log_2 \left[\frac{P(\vec{s}\,|spike)}{P(\vec{s})} \right].$$

If we look only at the probability distributions along a single dimension or a set of stimulus dimensions, we obtain the mutual Shannon information between these dimensions and the neural spike rate (Adelman et al. 2003, Sharpee et al. 2004). If the spike probability does not depend on this stimulus dimension, then these two probability distributions (one computed across all stimuli and the other for stimuli that elicited a spike) will be the same, indicating that these dimensions carry zero information about the neural spikes. By contrast, the dimensions that capture all the information in the neural response correspond to those upon which the decision to spike was based. Therefore, the RF of the LN model may be found by searching for the maximally informative dimension about the neural response (Sharpee et al. 2004). Because the KL distance upon which the mutual information is based is sensitive to any deviations between the probability distributions regardless of whether these deviations are described by differences in the first-, second-, or higher-order moments of the distribution, this procedure can be used with different kinds of stimuli, including natural stimuli (Figure 3*c*).

The proof that RF of the LN model corresponds to the maximally informative dimension about the neural response (I would prefer to restore “is based on”) the so-called data-processing inequality (Cover & Thomas 1991). This inequality states that adding any extra processing of inputs can only decrease the amount of information about the output of a system. Thus, if we take stimulus components along the same dimensions as were considered in the process of generating the neural spikes, then no extra processing steps are added. Otherwise, if we take stimulus components along dimensions that do not exactly correspond to the dimensions that elicited the neural spikes, then the output information will be reduced. Finding the RF of the LN model is equivalent to the maximum likelihood fitting of the LN model (Kouh & Sharpee 2009). Thus, although other distance measures can be used (Paninski 2003, Sharpee 2007) to compare changes in the probability distributions between all presented stimuli and stimuli that elicited a spike, information maximization yields the smallest variance in RF estimates versus other unbiased methods.

Multidimensional Feature Selectivity

The LN model discussed above describes the neural responses as being triggered according to the degree to which a single stimulus feature is present in the stimulus. However, responses of many types of sensory neurons exhibit a variety of contextual effects wherein their responses to the primary stimulus feature are modulated by the presence of other stimulus features in the stimulus. Examples include crossorientation suppression (Carandini et al. 1998, Priebe & Ferster 2006) and contrast-invariant orientation tuning for V1 cells (Troyer et al. 1998). Marr (1982) has argued that to distinguish a faint edge of a given orientation from a brighter edge of a nearby orientations requires that an orientation-selective neuron be suppressed by the presence of edges orthogonal to its preferred

orientation. Another example is the feature selectivity of complex cells in the primary visual cortex whose responses indicate the presence of an edge while allowing for some degree of position invariance. This property can be modeled with several visual features that correspond to Gabor patterns with different spatial phases (Adelson & Bergen 1985). To account for these aspects of neural coding, the traditional LN model is generalized such that the spike probability is a nonlinear function of several stimulus components (Rust et al. 2005). Similar to the case of a one-dimensional LN model, the nonlinearity can take an arbitrary shape, but now with respect to several stimulus components (Figure 4).

Spike-Triggered Covariance

How can we determine these relevant stimulus features from the neural responses? The STA is one of these features, and it is computed by analyzing the change in the mean between the stimulus distribution conditional on a spike and the distribution of all stimuli that were presented in the experiment (Figure 2). Analyzing the change in the variance between these stimulus distributions allows one to find all the relevant dimensions in cases where the stimuli are described by a Gaussian distribution (Bialek & de Ruyter van Steveninck 2005, de Ruyter van Steveninck & Bialek 1988, Schwartz et al. 2006) or by a circularly symmetric distribution (Samengo & Gollisch 2012). The intuition behind this procedure is that any dimension along which the variance is different from those expected a priori is associated with neural spikes. Unlike the change in the mean that defines a single dimension, the variance can differ along many dimensions.

In mathematical terms, the spike-triggered covariance method consists of two steps. The first step is to compute a difference between the covariance matrix of all stimuli and that of the stimuli that elicited a spike:

$$\Delta C = C - C^{spike}.$$

The second step is to diagonalize this matrix to find dimensions along which the variance is significantly different from zero. The variance is encoded in the eigenvalues of matrix ΔC. The eigenvectors that correspond to the significant eigenvalues describe the dimensions along which the variance is significantly different between the ensemble of stimuli that elicited a spike and that of all stimuli. To determine the significance, these two steps are repeated using shuffled spike trains that contain as many spikes as the recorded spike trains. Another way of breaking the correlations between stimuli and spikes that preserves all the structure in the neural spike trains is to shift forward these spike trains relative to the stimuli (Bialek & de Ruyter van Steveninck 2005). The spike-triggered covariance can, in principle, be used in the presence of stimulus correlations (Bialek & de Ruyter van Steveninck 2005, Schwartz et al. 2006). To obtain the relevant dimensions in this case, the eigenvectors of matrix ΔC must be divided by the stimulus covariance C_{ij} in a manner analogous to Equation 1. However, preliminary evidence indicates that this procedure can present more difficulties to execute in practice than arise when removing the effect of correlations from the STA (Aljadeff et al, 2013).

Maximally Informative Subspaces

Multiple relevant stimulus dimensions can also be found by maximizing the mutual information (Sharpee et al. 2004). Several dimensions that together account for a maximal amount of information in the neural response may be obtained by computing the KL distance between the probability distribution along these dimensions for all presented stimuli and the distribution for the stimuli that elicited a spike. For example, a pair of maximally informative dimensions may be found by maximizing

$$I(\vec{v}_1, \vec{v}_2) = \int dx_1 dx_2 P_{\vec{v}_1,\vec{v}_2}(x_1, x_2|spike) log_2 \left[\frac{P_{\vec{v}_1,\vec{v}_2}(x_1,x_2|spike)}{P_{\vec{v}_1,\vec{v}_2}(x_1,x_2)} \right]. \quad 2$$

This equation yields the KL distance between the probability distribution $P_{\vec{v}_1\vec{v}_2}(x_1, x_2)$ of stimulus projections x_1 and x_2, respectively, along dimensions $\vec{v}_1$ and $\vec{v}_2$, and the probability distribution $P_{\vec{v}_1\vec{v}_2}(x_1, x_2 \mid spike)$ of these projections across the stimuli that elicited a spike. Accordingly, a two-dimensional probability distribution must be sampled to jointly characterize two dimensions in terms of the amount of information about the neural responses that they provide.

In principle, one can find N maximally informative dimensions by sampling the stimulus probability distribution across N dimensions, although doing so is difficult to achieve in practice for more than three or four dimensions. Qualitatively, the number of samples that are available to map out the spike-conditional distribution $P_{\vec{v}_1\vec{v}_2}(x_1, x_2 \mid spike)$ is related to the number of recorded spikes. This is often the limiting factor in the computation of information. Note that the stimulus distribution $P_{\vec{v}_1\vec{v}_2}(x_1, x_2)$ is easier to sample, and in some cases, an analytic expression may be available. These constraints, known as the curse of dimensionality (Belleman 1961), limit the number of dimensions that can be simultaneously characterized according to the mutual information (Rowekamp & Sharpee 2011) as well as other divergence measures (Paninski 2003), which include the percentage of explained variance.

In some cases, it is possible to bypass the curse of dimensionality by searching for the relevant dimensions in an iterative fashion, for example, by first finding the relevant dimension, then finding the second relevant dimension in the subspace orthogonal to the first, etc. (Rapela et al. 2010, Rapela et al. 2006, Rowekamp & Sharpee 2011). Typically, this procedure works well when stimuli are not correlated. However, with uncorrelated stimuli, the relevant dimensions may also be found using spike-triggered covariance, which is a much simpler procedure. When stimuli are correlated, sequential searching can find the first relevant dimension accurately if stimulus correlations are Gaussian. The sequential search for secondary relevant dimensions is complicated by the presence of stimulus correlations: The sequential search can return a dimension that accounts for a large amount of information, not because it is relevant to the neural response, but because stimulus components along this dimensions are strongly correlated with the primary stimulus dimension (Rowekamp & Sharpee 2011).

QUADRATIC NONLINEAR MODELS

Maximally Informative Quadratic Models

A different approach for finding multiple relevant stimulus dimensions from the neural responses to natural stimuli is to modify the structure of the LN model. The model considered above allows for an arbitrary nonlinear function of a few stimulus components. A recently proposed alternative is to describe the spike probability as an arbitrary nonlinear function of a quadratic form of stimuli (Fitzgerald et al. 2011a, Rajan & Bialek 2012):

$$P(spike|\vec{s}) = F(s_i v_i + s_i J_{ij} s_j), \quad 3$$

where the sum over repeated indices is implied and F is an arbitrary nonlinear function. The parameters of the model are v_i, which is analogous to the RF, and J_{ij}. By analogy with the LN model, this model may be termed the quadratic-nonlinear (QN) model. The structure of the QN model is motivated by the need to capture such properties of sensory neurons as divisive normalization and contrast gain control (Carandini et al. 1997) where responses of

one neuron are normalized by a squared output of the responses of other neurons in the circuit. As discussed by Rajan & Bialek (2012), the QN model is also well matched both to the properties of complex cells (Adelson & Bergen 1985) in the primary visual cortex and to nonphase locked auditory neurons (Hudspeth & Corey 1977).

The QN model is also congruent with some types of the LN model. For example, the matrix J can have a low dimensional structure. In such cases, the neuronal response will be described as a quadratic function of a small number of stimulus components, as in the standard LN model. However, the LN model can, in principle, describe arbitrary interactions between the relevant dimensions, whereas these interactions are limited in the QN model (Equation 3) to sums and differences between the squares of relevant stimulus components. Nevertheless, the quadratic model provides a way forward to determine multiple relevant stimulus dimensions from the neural responses to natural stimuli. At the same time, by incorporating an arbitrary nonlinearity, the QN model represents an advance over the Wiener approach where in practice only a quadratic form of the stimulus can be estimated.

How can we estimate the parameters of the QN model from the neural responses to natural stimuli? As in the LN model, this can be done by finding its parameters---the values of the linear term v and the quadratic matrix J---that account for the maximal amount of information in the neural response (Fitzgerald et al. 2011a, Rajan & Bialek 2012). All the arguments for self-consistency of estimators carry over from the LN model because the QN model can be reformulated as an LN model with respect to the expanded stimulus $\{s_i; s_i s_j\}$, where indices i and j go from 1 to D (the dimensionality of the original stimulus). Even though the resulting expanded stimulus will have a very large dimensionality, analysis of model neurons suggests that the procedure remains feasible (Fitzgerald et al. 2011a). As in the spike-triggered covariance methods, diagonalizing the matrix J of the QN model will yield the stimulus dimensions that are relevant for spiking. Finding the parameters of the QN or LN models by maximizing information is equivalent to a maximum likelihood estimation, at least for a Poisson model of spiking (Kouh & Sharpee 2009). If the nonlinearity F in the QN model is constrained to be an exponential, then prior assumptions can be incorporated via maximum likelihood optimization onto the smoothness of the relevant stimulus features or the sparseness of their values (Park & Pillow 2011a). Minimal quadratic models: a Maximum Noise Entropy approach. The approaches described above rely on finding a suitable model structure to describe the neural circuit and then fitting parameters of these models according to the chosen criterion, such as information maximization, maximum likelihood estimation, or percent explained variance. For complex and hierarchical circuits, the appropriate model structure can be difficult to determine. An alternative to finding a suitable model structure to fit the neural responses is to construct what is known as a minimal model. The goal is to construct a model that is consistent with a given set of measurements of the neural responses and stimuli, but that is otherwise as unconstrained as possible. This approach is theoretically similar to the maximum entropy principle (Jaynes 1957) and in machine learning is known as the conditional Markov random fields (Lafferty et al. 2001).

According to the maximum entropy principle, when many distributions may be consistent with a given set of measurements, the distribution that has the maximal entropy (least constrained) should be chosen to obtain the least-biased model. Such a choice often yields models with the best predictive power on a novel set of data (Jaynes 2003). Recent studies show that this approach is fruitful in characterizing the responses of neural populations (Schneidman et al. 2006, Shlens et al. 2006). In the current discussion, the aim is to adopt this principle to build minimal models of input/output functions. Thus, the focus is on input/output functions for single neurons, although extensions to multiple neurons are certainly possible (Globerson et al. 2009, Granot-Atedge et al. 2012).

To build a minimal model of the neural input/output function, the chosen model should yield the highest entropy of the neural response for a given stimulus, averaged over all stimuli. The corresponding quantity is known as the noise entropy (Brenner et al. 2000, Strong et al. 1998). For binary responses (at sufficiently small time resolution, all neural responses are binary if patterns of spikes in time are not considered), the maximum noise entropy model has three appealing properties. First, this model is analytically solvable and the response function has a simple structure: It is a logistic function whose argument is a sum of stimulus parameters whose correlations with the neural response represent the constraints that the model needs to satisfy (Fitzgerald et al. 2011b). For example, the minimal model that is consistent with the measurements of the STA and spike-triggered covariance is

$$P(spike|\vec{s}) = 1/\left(1+exp\left(c+s_i v_i+s_i J_{ij} s_j\right)\right).$$

Here, parameters of the model are c, v_i, and J_{ij}. These parameters are adjusted such that the STA and spike-triggered covariance predicted by the model matche those measured experimentally. The relevant stimulus dimensions can be found by diagonalizing the matrix J, just as in the QN model.

Second, although similar, the main difference between the minimal quadratic model and the QN model is that the nonlinearity is a logistic function in the former, but can take an arbitrary form in the latter. Thus, the parameters of a minimal model can be found through a convex optimization that is not plagued by local minima. This practical advantage of minimal models makes it possible to find their parameters even with high dimensional stimuli. For example, in the analysis of simulated neurons with six relevant dimensions, the relevant stimulus dimensions from a minimal quadratic model were a slightly better match to the model dimensions than the dimensions from a maximally informative QN model (Fitzgerald et al. 2011a). However, in cases where only one or two relevant dimensions are needed, the maximally informative LN model yields a better match to model dimensions than either the maximally informative QN model or the minimal quadratic model.

Third, minimal models often make it possible to quantify, in information-theoretic terms, the relative importance of different constraints (Figure 5). For example, matching the minimal model to the experimental data in terms of the mean spike rate (which can be considered a zeroth order constraint because it does not involve stimuli) fully determines the overall entropy of the neural responses if they are binary or Poisson. With the entropy of the neural response fixed, maximizing the noise entropy is equivalent to seeking the model that provides the least amount of information while satisfying the necessary constraints (Globerson et al. 2009). As pointed out by Globerson et al. (2009), the information captured by the model that provides the minimal amount of information while satisfying a given set of constraints is a direct measure of the information content of these constraints (Figure 3). In general, minimally informative models are not analytically solvable, but in cases where they coincide with models computed by maximum noise entropy, the analytical solution is provided by the logistic function mentioned above. In addition to the binary and Poisson responses, it is possible to find a minimally informative model by maximizing noise entropy for Gaussian neural responses through the addition of the mean and variance of the spike rate to the set of constraints. Again, this is because the mean and variance are sufficient to specify the entropy of the neural response across a set of stimuli. Finally, the emergence of the logistic function as the least-constrained model that represents the necessary constraints between inputs and outputs suggests a possible functional explanation for the ubiquity of logistic input/output functions in systems biology, ranging from transcription control to neural gain functions (Clemens et al. 2012, Fairhall et al. 2006, Sharpee et al. 2011, Tyson et

al. 2003). It can also explain nonmonotonic gain functions, which are especially common in the auditory system wherein the neurons encode not just the mean of the relevant stimulus feature, but also its variance (Figure 6*b*).

OUTLOOK

Ultimately, computational methods for characterizing neural feature selectivity will be able to relate the neural responses to the underlying neural circuitry even in cases where many stages of nonlinear processing separate stimuli from the recorded neural responses. The progress that have been made in the field over the past ten years goes a long way toward this goal by recovering the multiple stimulus features that are relevant to the responses of high-level sensory neurons. However, much remains to be done. In particular the stimulus features upon which the responses of a given neuron are triggered usually have specific relationships to each other, as indicated by their co-occurrence in the natural environment. However, orthogonal representations of these features often make it difficult to deduce these relationships. For example, when the neural responses are triggered by the same image feature that is centered at different positions in the visual field, the features obtained by spike-triggered covariance make it difficult to guess the computational relationship between the features (Rust et al. 2005). Recent methods are beginning to characterize the neural feature selectivity by taking invariance directly into account (Eickenberg et al. 2012, Vintch et al. 2012). The resulting models can economically account for the observed neural responses as triggered by a small number of image features, while allowing for their different positioning within the visual field. In other sensory systems, relevant invariance properties are either more difficult to parametrize than position invariance or are not known. One possible way to discover types of invariance present in the responses of a particular neuron is to find such linear combinations of the relevant stimulus features that make it possible to describe the observed responses as logical OR operation with respect to a set of inputs (Kaardal et al. 2013). Given a set of relevant stimulus features, finding the appropriate linear combinations of features are numerically and computationally easy to perform, although the methods so far have not been tested in higher level sensory areas. The hope is that by building on combinations of the available computational tools for receptive fields identification future works will be able not only to characterize the neural feature selectivity in the presence of complex and in some cases yet unknown types of invariance on high level sensory neurons in vision and other sensory modalities.

Acknowledgments

The author thanks Johnatan Aljadeff, William Bialek, Michael Eickenberg, Jeffrey D. Fitzgerald, Minjoon Kouh, Kenneth D. Miller, Michael P. Stryker, Ryan Rowekamp, and Adrian Wanner for many helpful discussions. In addition, I thank Johnatan Aljadeff and Brian Sharpee for comments on the manuscript and Jeffrey Fitzgerald for help with some of the figures. This research was supported by grant R01EY019493 from the National Institutes of Health, grant 0712852 from the National Science Foundation, the Alfred P. Sloan Research Fellowship, Searle Funds, the McKnight Scholarship, the Ray Thomas Edwards Career Development Award in Biomedical Sciences, and the W.M. Keck Foundation Research Excellence Award. Additional resources were provided by the Center for Theoretical Biological Physics (NSF PHY-0822283).

LITERATURE CITED

Adelman TL, Bialek W, Olberg RM. The information content of receptive fields. Neuron. 2003; 40:823–33. [PubMed: 14622585]

Adelson EH, Bergen JR. Spatiotemporal energy models for the perception of motion. J Opt Soc Am A. 1985; 2:284–99. [PubMed: 3973762]

Ahrens MB, Linden JF, Sahani M. Nonlinearities and contextual influences in auditory cortical responses modeled with multilinear spectrotemporal methods. J Neurosci. 2008; 28:1929–42. [PubMed: 18287509]

Anzai A, Peng X, Van Essen DC. Neurons in monkey visual area V2 encode combinations of orientations. Nat Neurosci. 2007; 10:1313–21. [PubMed: 17873872]

Bakin JS, Nakayama K, Gilbert CD. Visual responses in monkey areas V1 and V2 to three-dimensional surface configurations. J Neurosci. 2000; 20:8188–98. [PubMed: 11050142]

Belleman, R. Adaptive Processes: A Guided Tour. Princeton, NJ: Princeton Univ. Press; 1961.

Bialek, W.; de Ruyter van Steveninck, RR. Features and dimensions: motion estimation in fly vision. 2005. http://arxiv.org/abs/q-bio/0505003

Brenner N, Strong SP, Koberle R, Bialek W, de Ruyter van Steveninck RR. Synergy in a neural code. Neural Comput. 2000; 12:1531–52. [PubMed: 10935917]

Carandini M, Heeger DJ, Movshon JA. Linearity and normalization in simple cells of the macaque primary visual cortex. J Neurosci. 1997; 17:8621–44. [PubMed: 9334433]

Carandini M, Movshon JA, Ferster D. Pattern adaptation and cross-orientation interactions in the primary visual cortex. Neuropharmacology. 1998; 37:501–11. [PubMed: 9704991]

Chichilnisky EJ. A simple white noise analysis of neuronal light responses. Network. 2001; 12:199–213. [PubMed: 11405422]

Christianson GB, Sahani M, Linden JF. The consequences of response nonlinearities for interpretation of spectrotemporal receptive fields. J Neurosci. 2008; 28:446–55. [PubMed: 18184787]

Clemens J, Wohlgemuth S, Ronacher B. Nonlinear computations underlying temporal and population sparseness in the auditory system of the grasshopper. J Neurosci. 2012; 32:10053–62. [PubMed: 22815519]

Cover, TM.; Thomas, JA. Elements of Information Theory. New York: Wiley-Interscience; 1991.

de Boer E, Kuyper P. Triggered correlation. IEEE Trans Biomed Eng. 1968; 15:169–79. [PubMed: 5667803]

de Ruyter van Steveninck RR, Bialek W. Real-time performance of a movement-sensitive neuron in the blowfly visual system: coding and information transfer in short spike sequences. Proc R Soc Lond Ser B. 1988; 234:379–414.

Desimone R, Schein SJ. Visual properties of neurons in area V4 of the macaque: sensitivity to stimulus form. J Neurophysiol. 1987; 57:835–68. [PubMed: 3559704]

Eickenberg M, Rowekamp RJ, Kouh M, Sharpee TO. Characterizing responses of translation-invariant neurons: maximally informative invariant dimensions. Neural Comput. 2012; 24:2384–421. [PubMed: 22734487]

Fairhall AL, Burlingame CA, Narasimhan R, Harris RA, Puchalla JL, Berry MJ 2nd. Selectivity for multiple stimulus features in retinal ganglion cells. J Neurophysiol. 2006; 96:2724–38. [PubMed: 16914609]

Felsen G, Dan Y. A natural approach to studying vision. Nat Neurosci. 2005; 8:1643–46. [PubMed: 16306891]

Field DJ. Relations between the statistics of natural images and the response properties of cortical cells. J Opt Soc Am A. 1987; 4:2379–94. [PubMed: 3430225]

Fitzgerald JD, Rowekamp RJ, Sincich LC, Sharpee TO. Second order dimensionality reduction using minimum and maximum mutual information models. PLoS Comput Biol. 2011a; 7:e1002249. [PubMed: 22046122]

Fitzgerald JD, Sincich LC, Sharpee TO. Minimal models of multidimensional computations. PLoS Comput Biol. 2011b; 7:e1001111. [PubMed: 21455284]

Gilbert CD, Wiesel TN. The influence of contextual stimuli on the orientation selectivity of cells in primary visual cortex of the cat. Vis Res. 1990; 30:1689–701. [PubMed: 2288084]

Globerson A, Stark E, Vaadia E, Tishby N. The minimum information principle and its application to neural code analysis. Proc Natl Acad Sci USA. 2009; 106:3490–95. [PubMed: 19218435]

Granot-Atedge, E.; Tkacik, G.; Segev, R.; Schneidman, E. Stimulus-dependent maximum entropy models of neural population codes. 2012. http://arxiv.org/abs/1205.6438

Hartline HK. The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. Am J Physiol. 1938; 121:400–15.

Hubel DH, Wiesel TN. Receptive fields and functional architecture of monkey striate cortex. J Physiol. 1968; 195:215–43. [PubMed: 4966457]

Hudspeth AJ, Corey DP. Sensitivity, polarity, and conductance change in the response of vertebrate hair cells to controlled mechanical stimuli. Proc Natl Acad Sci USA. 1977; 74:2407–11. [PubMed: 329282]

Huys QJ, Paninski L. Smoothing of, and parameter estimation from, noisy biophysical recordings. PLoS Comput Biol. 2009; 5:e1000379. [PubMed: 19424506]

Jaynes ET. Information theory and statistical mechanics. Phys Rev. 1957; 106:620–30.

Jaynes, ET. Probability Theory: The Logic of Science. Cambridge, UK: Cambridge Univ. Press; 2003.

Kaardal J, Fitzgerald JD, Berry MJ II, Sharpee TO. Identifying functional bases for multidimensional neural computations. Neural Computation. in press.

Kouh M, Sharpee TO. Estimating linear-nonlinear models using Renyi divergences. Network. 2009; 20:49–68. [PubMed: 19568981]

Kuffler SW. Discharge patterns and functional organization of mammalian retina. J Neurophysiol. 1953; 16:37–68. [PubMed: 13035466]

Kunsberg B, Zucker SW. Shape-from-shading and cortical computation: a new formulation. J Vis. 2012; 12:233.

Lafferty, J.; McCallum, A.; Pereira, FCN. Conditional random fields: probabilistic models for segmenting and labeling sequence data. Proc. Int. Conf. Mach. Learn., 18th; Williamstown. June 28–July 1; San Francisco, CA: Morgan Kaufmann; 2001. p. 282-89.

Lazar AA, Slutskiy YB. “Channel Identification Machines,” Computational Intelligence and Neuroscience 2012. 2012:Article ID 209590. 20 pages. 10.1155/2012/209590

Lewi J, Butera R, Paninski L. Sequential optimal design of neurophysiology experiments. Neural Comput. 2009; 21:619–87. [PubMed: 18928364]

Li A, Zaidi Q. Three-dimensional shape from non-homogeneous textures: carved and stretched surfaces. J Vis. 2004; 4:860–78. [PubMed: 15595891]

Marmarelis, VZ.; Marmarelis, PZ. Analysis of Physiological Systems The White Noise Approach. Chacon: Lavoisier; 1978.

Marr, D. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. New York: WH Freeman & Co; 1982.

McManus JN, Li W, Gilbert CD. Adaptive shape processing in primary visual cortex. Proc Natl Acad Sci USA. 2011; 108:9739–46. [PubMed: 21571645]

Meister M, Berry MJ. The neural code of the retina. Neuron. 1999; 22:435–50. [PubMed: 10197525]

Nothdurft HC, Gallant JL, Van Essen DC. Response modulation by texture surround in primate area V1: correlates of “popout” under anesthesia. Vis Neurosci. 1999; 16:15–34. [PubMed: 10022475]

Paninski L. Convergence properties of three spike-triggered analysis techniques. Network. 2003; 14:437–64. [PubMed: 12938766]

Paninski L, Pillow J, Lewi J. Statistical models for neural encoding, decoding, and optimal stimulus design. Prog Brain Res. 2007; 165:493–507. [PubMed: 17925266]

Park IMM, Pillow JW. Bayesian spike-triggered covariance analysis. Adv Neural Inf Process Syst. 2011a; 24:1692–700.

Park M, Pillow JW. Receptive field inference with localized priors. PLoS Comput Biol. 2011b; 7:e1002219. [PubMed: 22046110]

Pasupathy A, Connor CE. Responses to contour features in macaque area V4. J Neurophysiol. 1999; 82:2490–502. [PubMed: 10561421]

Press, WH.; Teukolsky, SA.; Vetterling, WT.; Flannery, BP. Numerical Recipes in C: The Art of Scientific Computing. Cambridge, UK: Cambridge Univ. Press; 1992.

Priebe NJ, Ferster D. Mechanisms underlying cross-orientation suppression in cat visual cortex. Nat Neurosci. 2006; 9:552–61. [PubMed: 16520737]

Rajan, K.; Bialek, W. Maximally informative “stimulus energies” in the analysis of neural responses to natural signals. 2012. http://arxiv.org/abs/1201.0321

Rapela J, Felsen G, Touryan J, Mendel JM, Grzywacz NM. ePPR: a new strategy for the *characterization of sensory cells from input/output data*. Network. 2010; 21:35–90. [PubMed: 20735338]

Rapela J, Mendel JM, Grzywacz NM. Estimating nonlinear receptive fields from natural images. J Vis. 2006; 6:441–74. [PubMed: 16889480]

Rieke, F.; Warland, D.; de Ruyter van Steveninck, R.; Bialek, WB. Spikes: Exploring the Neural Code. Cambridge, MA: MIT Press; 1997.

Ringach DL, Hawken MJ, Shapley R. Receptive field structure of neurons in monkey primary visual cortex revealed by stimulation with natural image sequences. J Vis. 2002; 2:12–24. [PubMed: 12678594]

Rowekamp RJ, Sharpee TO. Analyzing multicomponent receptive fields from neural responses to natural stimuli. Network. 2011; 22:1–29. [PubMed: 22149668]

Ruderman DL. Origins of scaling in natural images. Vis Res. 1997; 37:3385–98. [PubMed: 9425551]

Ruderman DL, Bialek W. Statistics of natural images: scaling in the woods. Phys Rev Let. 1994; 73:814–7. [PubMed: 10057546]

Rust NC, Movshon JA. In praise of artifice. Nat Neurosci. 2005; 8:1647–50. [PubMed: 16306892]

Rust NC, Schwartz O, Movshon JA, Simoncelli EP. Spatiotemporal elements of macaque V1 receptive fields. Neuron. 2005; 46:945–56. [PubMed: 15953422]

Sahani, M.; Linden, JF. Evidence optimization techniques for estimating stimulus-response functions. In: Becker, S.; Thrun, S.; Obermayer, K., editors. Advances in Neural Information Processing Systems. Vol. 15. Cambridge, MA: MIT Press; 2003. p. 301-8.

Samengo I, Gollisch T. Spike-triggered covariance: geometric proof, symmetry properties, and extension beyond Gaussian stimuli. J Comput Neurosci. 2013; 34:137–61. [PubMed: 22798148]

Schneidman E, Berry MJ 2nd, Segev R, Bialek W. Weak pairwise correlations imply strongly correlated network states in a neural population. Nature. 2006; 440:1007–12. [PubMed: 16625187]

Schwartz O, Hsu A, Dayan P. Space and time in visual context. Nat Rev Neurosci. 2007; 8:522–35. [PubMed: 17585305]

Schwartz O, Pillow JW, Rust NC, Simoncelli EP. Spike-triggered neural characterization. J Vis. 2006; 6:484–507. [PubMed: 16889482]

Series P, Lorenceau J, Fregnac Y. The “silent” surround of V1 receptive fields: theory and experiments. J Physiol Paris. 2003; 97:453–74. [PubMed: 15242657]

Sharpee T, Rust NC, Bialek W. Analyzing neural responses to natural signals: maximally informative dimensions. Neural Comput. 2004; 16:223–50. [PubMed: 15006095]

Sharpee TO. Comparison of information and variance maximization strategies for characterizing neural feature selectivity. Stat Med. 2007; 26:4009–31. [PubMed: 17597484]

Sharpee TO, Miller KD, Stryker MP. On the importance of static nonlinearity in estimating spatiotemporal neural filters with natural stimuli. J Neurophysiol. 2008; 99:2496–509. [PubMed: 18353910]

Sharpee TO, Nagel KI, Doupe AJ. Two-dimensional adaptation in the auditory forebrain. J Neurophysiol. 2011; 106:1841–61. [PubMed: 21753019]

Sharpee TO, Victor JD. Contextual modulation of V1 receptive fields depends on their spatial symmetry. J Comput Neurosci. 2008; 26:203–18. [PubMed: 18679785]

Sherrington, CS. The Integrative Action Of The Nervous System. New York: Schribner & Sons; 1906.

Shlens J, Field GD, Gauthier JL, Grivich MI, Petrusca D, et al. The structure of multi-neuron firing patterns in primate retina. J Neurosci. 2006; 26:8254–66. [PubMed: 16899720]

Sillito AM, Jones HE. Context-dependent interactions and visual processing in V1. J Physiol-Paris. 1996; 90:205–9. [PubMed: 9116668]

Simoncelli EP. Vision and the statistics of the visual environment. Curr Opin Neurobiol. 2003; 13:144–49. [PubMed: 12744966]

Simoncelli EP, Olshausen BA. Natural image statistics and neural representation. Annu Rev Neurosci. 2001; 24:1193–216. [PubMed: 11520932]

Singh NC, Theunissen FE. Modulation spectra of natural sounds and ethological theories of auditory processing. J Acoust Soc Am. 2003; 114:3394–411. [PubMed: 14714819]

Strong SP, de Ruyter van Steveninck RR, Bialek W, Koberle R. On the application of information theory to neural spike trains. Pac Symp Biocomput. 1998; 1998:621–32. [PubMed: 9697217]

Theunissen FE, David SV, Singh NC, Hsu A, Vinje WE, Gallant JL. Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. Network. 2001; 12:289–316. [PubMed: 11563531]

Theunissen FE, Sen K, Doupe AJ. Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. J Neurosci. 2000; 20:2315–31. [PubMed: 10704507]

Troyer TW, Krukowski AE, Priebe NJ, Miller KD. Contrast-invariant orientation tuning in cat visual cortex: feedforward tuning and correlation-based intracortical connectivity. J Neurosci. 1998; 18:5908–27. [PubMed: 9671678]

Tyson JJ, Chen KC, Novak B. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. Curr Opin Cell Biol. 2003; 15:221–31. [PubMed: 12648679]

van Hateren JH, van der Schaaf A. Independent component filters of natural images compared with simple cells in primary visual cortex. Proc R Soc Lond Ser B. 1998; 265:359–66.

Vickers NJ, Christensen TA, Baker TC, Hildebrand JG. Odour-plume dynamics influence the brain's olfactory code. Nature. 2001; 410:466–70. [PubMed: 11260713]

Victor, JD.; Knight, BW. Simultaneously band and space limited functions in two dimensions, and receptive fields of visual neurons. In: Kaplan, E.; Marsden, J.; Sreenivasan, KR., editors. Springer Applied Mathematical Sciences Series. New York: Springer-Verlag; 2003. p. 375-420.

Victor JD, Mechler F, Repucci MA, Purpura KP, Sharpee TO. Responses of V1 neurons to two-dimensional Hermite functions. J Neurophysiol. 2006; 95:379–400. [PubMed: 16148274]

Victor JD, Shapley R. A method of nonlinear analysis in the frequency domain. Biophys J. 1980; 29:459–83. [PubMed: 7295867]

Victor JD, Purpura KP. Spatial phase and the temporal structure of the response to gratings in V1. J Neurophysiol. 1998; 80:554–71. [PubMed: 9705450]

Victor JD, Shapley RM. The nonlinear pathway of Y ganglion cells in the cat retina. J Gen Physiol. 1979; 74:671–89. [PubMed: 231636]

Vintch, B.; Zaharia, A.; Movshon, JA.; Simoncelli, EP. Efficient and direct estimation of a neural subunit model for sensory coding. In: Bartlett, P.; Pereira, FCN.; Burges, CJC.; Bottou, L.; Weinberger, KQ., editors. Advances Neural Information Processing Systems. Vol. 25. NIPS; La Jolla, CA: 2012. NIPS Found. http://books.nips.cc/papers/files/nips25/NIPS2012_1432.pdf

Weisberg S, Welsh AH. Adapting for the missing link. Ann Stat. 1994; 22:1674–700.

Zipser K, Lamme VA, Schiller PH. Contextual modulation in primary visual cortex. J Neurosci. 1996; 16:7376–89. [PubMed: 8929444]

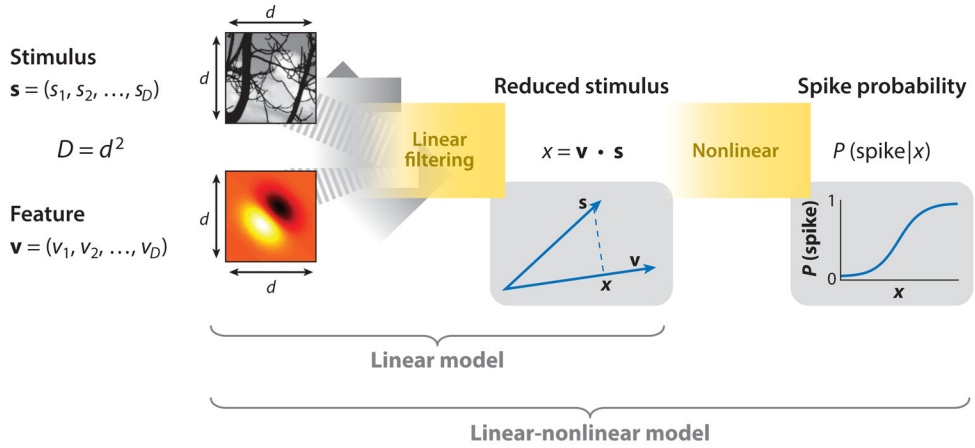

Figure 1.
Geometric interpretation of the receptive field (RF) in the context of the linear and linear-nonlinear (LN) models. An example stimulus is a natural image taken from a van Hateren data set (van Hateren & van der Schaaf 1998). The stimulus has d pixels in the horizontal and vertical dimensions, which yields a stimulus of $D = d^2$ dimensions. The RF taken to mimic properties of V1 neurons is also defined in this space. The linear model predicts the spike probability as taking a projection between the stimulus and the RF. The LN model adds a nonlinear gain function to account for such properties as rectification and saturation in the neural response.

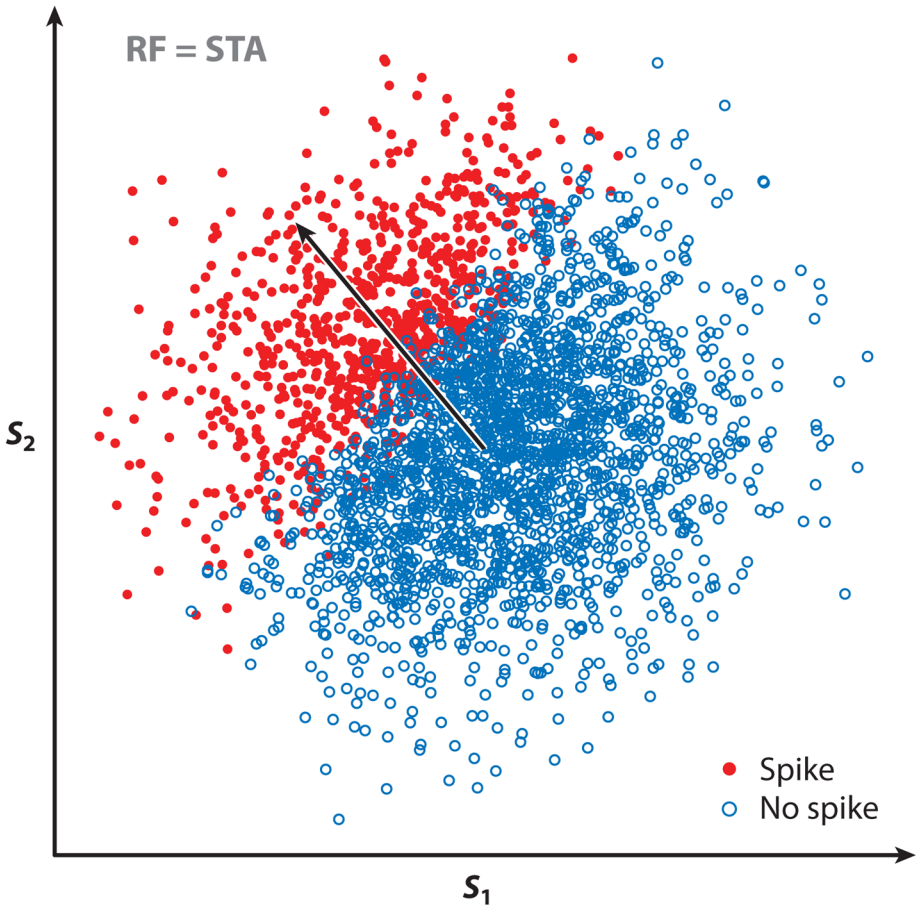

Figure 2.
Illustration of how the receptive field of a neuron can be estimated from its responses to white-noise stimuli by computing the spike-triggered average. Each dot represents a high-dimensional stimulus projected onto a plane. Each stimulus was taken from an uncorrelated, white-noise Gaussian distribution. Stimuli that elicited a spike are marked with red filled dots. Their average yields a vector, which is the relevant stimulus dimension for generating spikes from this neuron.

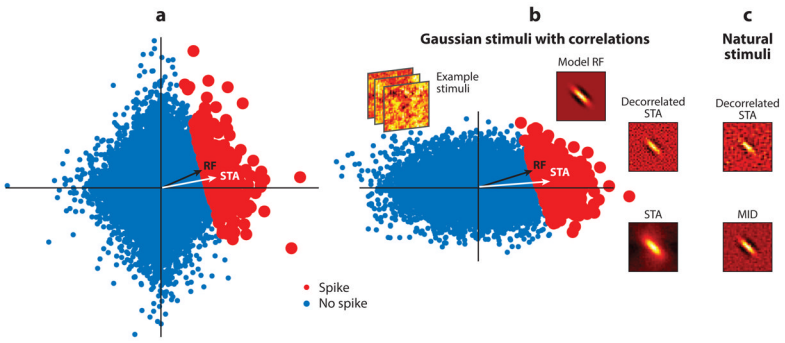

Figure 3.

Differences between the spike-triggered average (STA) and the receptive field (RF) for noncircular probability distributions. (*a*) Example of a non-Gaussian distribution of stimuli without correlations. (*b*) Example of a Gaussian distribution with correlations. In both examples, the STA does not always yield a correct estimate of the RF. However, for the Gaussian distribution, it can be corrected according to Equation 1 to yield an accurate estimate. (*c*) In the case of natural stimuli, which are non-Gaussian and have strong correlations, even with corrections for the second-order correlations, the STA does not yield a correct RF estimate. However, the RF can be estimated as a maximally informative dimension (MID).

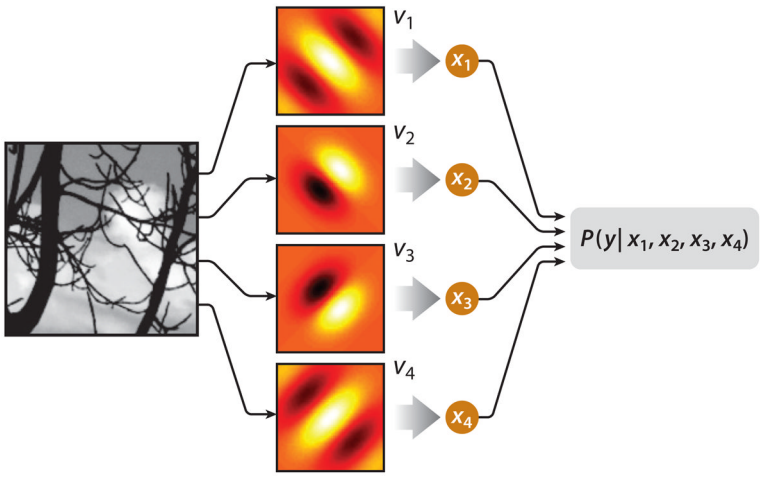

Figure 4.
Schematic of a linear-nonlinear model with multiple relevant stimulus features.

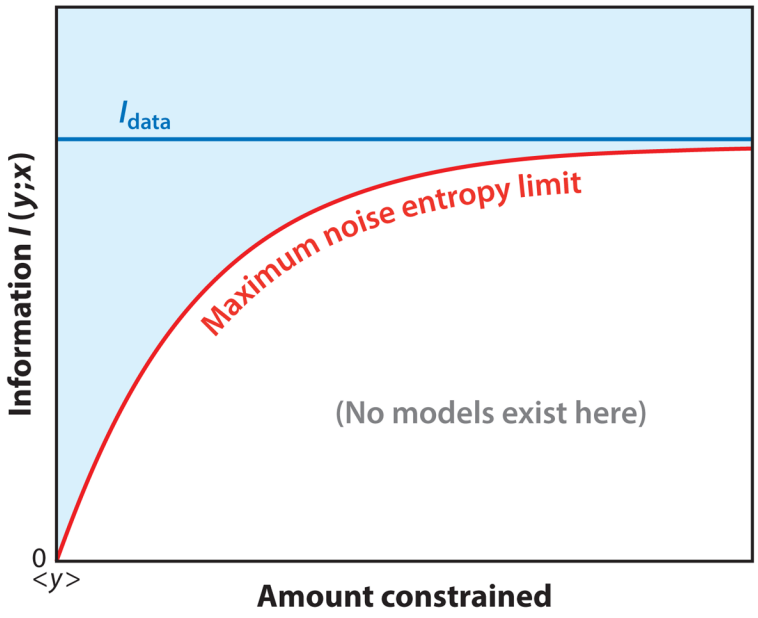

Figure 5.
Geometric intuition of information transmission in minimal models. For a given set of constraints, the maximal noise entropy model provides the smallest amount of information between the neural response $y = 0,1$ and the stimulus. The point of origin with zero information corresponds to the model where only the mean spike rate $<y>$ is constrained; this model carries no information about the stimulus. As more constraints are added, information captured by the minimal model will approach the true information in the data I_{data}. As a result, the information content of any constraint can be quantified. In contrast, a maximally informative model is used to find the value of parameters that account for the greatest amount of I_{data} in one step. Although optimization to find the maximally informative set of parameters is nonconvex, the corresponding optimization for the minimal models is convex. Given the right set of constraints, both models will converge to the same input/output function of the neuron.

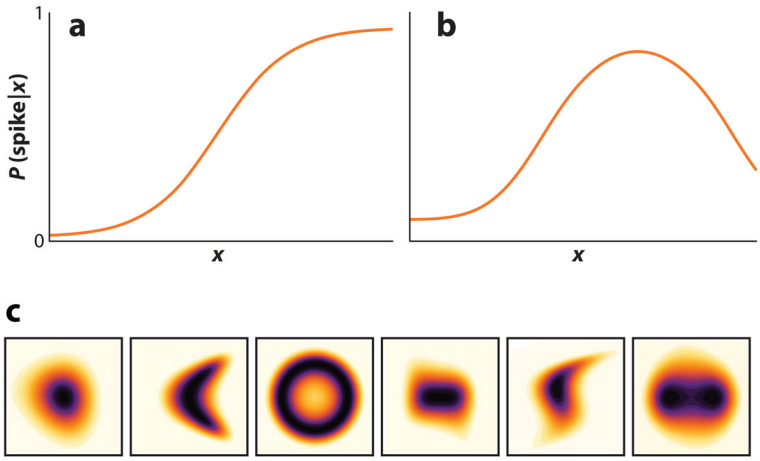

Figure 6.
Maximum noise entropy models can account for variety of nonlinear effects. (*a*) The logistic function arises when the neural response encodes stimuli linearly. (*b*) The logistic function of a quadratic argument can be observed in models where the neural response encodes both the mean and the variance of stimulus components along the relevant stimulus features. In this case, the nonlinear gain function is nonmonotonic. (*c*) Maximum noise entropy models based on two relevant stimulus features can characterize a variety of nonlinear computations when extended to moments higher than two. These include “ring,” “bimodal,” and “crescent-shaped” nonlinearities previously observed in the retina (Fairhall et al. 2006).