Modeling and Characterization of Neural Gain Control

by

Odelia Schwartz

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Center for Neural Science

New York University

September 2002

Eero P. Simoncelli

ACKNOWLEDGMENTS

I am most grateful to my advisor, Eero Simoncelli, for all the time, support, and enthusiasm he has given throughout the course of this research; and for his many inspiring ideas and computational insights. I have learned so much from him.

I am very thankful to my other committee members, Mal Semple and Tony Movshon, for discussions and advice; and to my examining committee, Carlos Brody and Dan Tranchina, for suggestions and comments about the manuscript.

I thank everyone in the Simoncelli lab for their support over the years. This work has been enriched through discussions with Javier Portilla, Jonathan Pillow, Jesus Malo, and Liam Paninski.

I am grateful to James Cavanaugh, Wyeth Bair, and Nicole Rust in the Movshon lab for sharing their data and for discussions.

Many thanks to EJ Chichilnisky for his experimental collaboration and suggestions on the spike-triggered covariance work. Thanks also to Divya Chander for data collection.

Special thanks to John Harris, my former MS advisor; he encouraged my interest in computational sensory processing, and has often given me advice throughout the years.

I am very grateful to Sam Feldman and Stu Greenstein for their help and dedication during my time at the Center for Neural Science. Thanks to friends and classmates for their help, and for making my time more enjoyable.

I thank my parents for their constant love and support.

And to Shlomi, my husband and best friend, thank you for going through this with me. Contributions include but are not limited to help with figures, love, encouragement, delicious meals, and sense of humor.

Funding during my time at the Center was provided by the Sloan-Swartz foundation and the National Eye Institute (NEI).

ABSTRACT

Sensory neurons often exhibit striking nonlinear behaviors that are not adequately described by a linear receptive field representation.

In the first part of the work, we suggest that these nonlinearities arise because sensory systems are designed to efficiently represent environmental information. We describe a form of nonlinear decomposition (specifically, divisive gain control) that is well-suited for efficient encoding of natural signals. We show that this decomposition, with parameters optimized for the statistics of a generic ensemble of natural images or sounds, can account for some nonlinear response properties of "typical" neurons in both vision (area V1) and audition (auditory nerve). This work provides theoretical justification to neural models of gain control, and explains how one might choose the parameters of the model based on efficient coding considerations.

In the second part of the work, we describe a methodology for characterizing this class of nonlinear sensory models. The characterization is based on a white noise analysis, in which a set of random stimuli are presented to a neuron and the spike-triggered ensemble (specifically, the spike-triggered covariance) is analyzed. We demonstrate the applicability of the technique to retinal ganglion cell data in monkey and salamander.

The concept of gain control appears to be ubiquitous, occurring throughout the nervous system. So the computational tools we develop here may potentially have more widespread applicability.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

We interact with stimuli in the natural world through the senses. Locating predator or prey in the jungle, or recognizing a face in a crowded subway–we are constantly making sense of a complex environment. This puts to work a plethora of neurons that process the information at various stages, resulting in perception, and ultimately, in action and storage of memory. Within this very intricate system, this thesis aims to understand: why do neurons form the particular representations that they do? And how can we characterize sensory neural representations?

Specifically, we focus on neurons at relatively early stages of processing, including the retina, the primary visual cortex (area V1), and the peripheral auditory system. To answer these types of questions, we have studied neural behaviors at the signal processing, phenomenological level. This approach typically does not reveal a particular physiological mechanism; but will often initiate complementary exploration of the processes of biochemistry, spike generation, synaptic mechanisms, circuitry, and so on.

Images in the world are initially represented in a rather pixelated form, by the

light intensities that impinge upon the photoreceptor array. We are interested in the types of representations that neurons form of their input stimuli along the sensory pathway. Perhaps the simplest representation we tend to think about, and which physiologists often measure, are receptive fields (or filters) over localized regions in space and time. For example, Hubel and Wiesel swept bars across the receptive spatial regions of neurons in V1, and found that neurons were selective (and responded with a higher firing rate) to particular orientations [67]. In this way, one can map out a receptive field of a neuron, which represents, roughly, what stimuli the neuron responds to best.

If a neuron were truly linear, then one could make a prediction about its response to any given stimulus, just by knowing its receptive field. This would simply be the projection of the stimulus onto the receptive field (of course, in practice, even if one attempts to describe only the mean firing rate of the neuron, one must at least include a rectifying nonlinearity). However, receptive field characterizations are often not sufficient to describe the complexities of sensory neural behaviors. For example, in primary visual cortex, the response of a neuron to the optimal stimulus (e.g., that which is exactly equal to the receptive field) can be modulated by superimposing a second stimulus that by itself elicits no response at all in the neuron. This is a nonlinear behavior, suggesting that the representation is governed by other stimulus directions, and not just that given by the linear receptive field. These and other observations have suggested that a model of linear neural representation is too simplistic.

In this chapter we provide a literature review, that feeds on both theoretical and experimental work in the field. We first describe neural models that have been

derived purely from theoretical principles. We show that such models have been fruitful in explaining some basic aspects of sensory representation, particularly regarding linear receptive field properties measured physiologically. But examination of physiological data reveal that sensory neurons are highly nonlinear. Indeed, physiology driven neural models have been developed that can account for some of these nonlinearities (specifically, gain control models). Finally, we describe how one can characterize individual neurons, assuming a particular form of underlying model. Specifically, spike-triggered approaches have been successful in characterizing some response properties of sensory neurons, but have not proceeded to characterize gain control models. This lays out the groundwork and motivation for the remaining chapters of the thesis.

1.1 Theory Driven Sensory Neural Models

1.1.1 History and Outline

Many factors are likely to shape sensory processing, including: the particular tasks an animal must perform; the capabilities and limitations of neurons; and the structure of the sensory stimuli that neurons receive. Here, we limit the discussion to how the structure of sensory stimuli in the world effect neural processing. This is an area that has been widely studied, and in which specific hypotheses about neural processing have been formulated.

Stimuli in our sensory environment are highly structured and non random (Figure 1.1). Of all the possible could-be images and sounds, we are exposed to only a small portion in our particular environment. Since some stimuli are more likely to

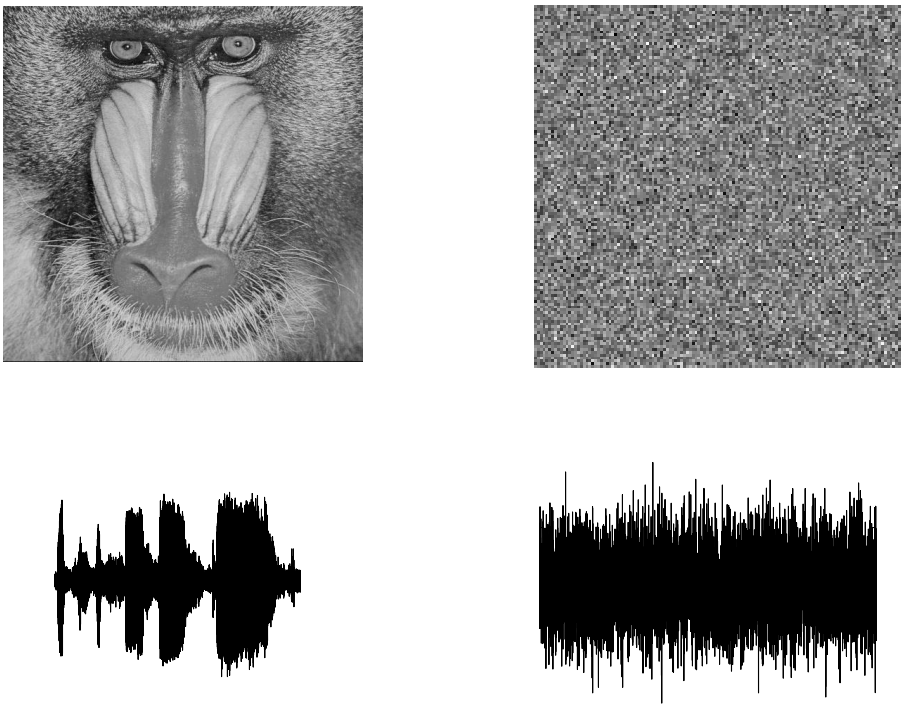

Figure 1.1: Natural image and sound vs. random white noise

occur than others, it seems reasonable that our sensory systems are most suited to process those stimuli that occur most frequently.

Early notions about sensory stimuli and the mind may actually be traced back to Greek philosophy. Aristotle held that the mind was a "tabula rasa" or clean slate; and that this slate is only filled with knowledge after sense experience. More modern versions of this idea were influenced by landmark scientific advances: with the discovery of nerve cells; with evolution emerging in the mid nineteenth century; and Shannon's information theory in the mid twentieth century.

It has been widely assumed that neurons in sensory areas of the brain are adapted, through processes of evolution and development, to the signals to which they are exposed [1] [85, 106, 161, 39, 11, 16]. Information theory provided a quantitative framework for converting these ideas into a more concrete hypothesis about neural processing. Attneave and Barlow hypothesized that the brain develops in response to the statistical properties of the natural stimuli it is exposed to – and does so in a manner that is efficient (henceforth, will be referred to as efficient coding hypothesis). We will define "efficient" shortly, from the point of view of information theory.

In the past decade or so, the efficient coding hypothesis has spurred renewed interest (see also review article [140]). This is due to recent computational advances in the statistical analysis of natural images and sounds. In addition, advances in experimental techniques are beginning to pave the way towards more direct routes for testing this hypothesis.

[1] Adaptation over shorter time scales to environmental signals is also assumed to effect neural processing, see [17, 165, 49].

In the following sections, we first describe the computational building blocks of information theory, and how they relate to forming an efficient representation. We then explain how these ideas have led to the efficient coding hypothesis. Subsequently, we describe an additional hypothesis that has emerged, sparse coding. Finally, we describe computational and experimental work that have been aimed at testing these hypotheses.

1.1.2 Information Theory

Information theory was developed to help engineers in the transmission of signals over communication devices [132, 133]. The theory is deeply rooted mathematically, and has been applied to numerous other areas that require efficient processing of signals. Hence, the theory may potentially be relevant to neural computation. Here we introduce only the most basic concepts necessary for our discussion (see [133, 38] for a more complete and detailed description).

Consider an ensemble of signals that contain symbols, such as letters of the English alphabet, light intensity values in images, sound pressure levels in sounds, and so on. We assume that these input signals are passed through channels. The outputs are typically constrained by the architecture of the channels. A goal of information theory is that the output signals through the channels form a more efficient representation of the input. To achieve this goal, Shannon introduced (among other terms) redundancy, entropy, and mutual information.

Related ideas were later developed in the framework of minimum description length, with the goal of creating a compact and efficient representation of observed signals, from which the original observed signals can be regenerated [148, 118, 170].

Additional work has also focused on efficient coding using rate-distortion criteria (see [38]), and extraction of the most "relevant" information from a signal, while forming an efficient representation [151]. These directions are beyond the scope of this thesis.

Redundancy

Signals are considered inefficient in their representation, if they are redundant. There are two forms of redundancy that can occur in signals: some symbols occur more frequently than others (in the English language 'a' occurs more often than 'q'); and, some symbols co-occur more frequently than others, i.e., they are not independent (letters "sh" are more likely to appear together than "sd"). A more efficient representation is formed by reducing the redundancies in the signal ensemble. It is important to note that efficiency is only meaningful for a particular stimulus ensemble, and what is efficient for one ensemble is not necessarily efficient for another.

Entropy and Mutual Information

To quantify the notion of redundancy reduction, we define some basic terms from information theory.

We denote X the set of input signals and Y the set of output signals (through one or more channels).

The entropy of the output is given by:

$$H(Y) = \sum_{y} \mathcal{P}(y)\, log_2 \mathcal{P}(y) \tag{1.1}$$

where $\mathcal{P}(\cdot)$ is the probability distribution.

Entropy is a measure of how informative signals in the ensemble are on average. The information for each signal y is given by $log_2\mathcal{P}(y)$ and multiplied by the probability of occurrence of the signal. Entropy is always positive, and equal to zero if and only if the outcome is certain; that is, $\mathcal{P}(y)$ is 0 for all but one of the signals y which occurs with probability 1. The $log2$ indicates that information is expressed in units of bits. Stated in coding terms, entropy is the average number of bits needed to represent values drawn from a distribution.

We first assume that the output is deterministically determined from the input. Then the representation is made less redundant and more efficient by maximizing the entropy of the output signals. So in the resulting representation the output is on average less predictable, and carries more information. But many systems, including neurons, are not deterministic: if a neuron is stimulated repeatedly with the same stimulus, the response is variable (we'll denote this variability by the term "noise"). We'd like to achieve maximal entropy of the output that's only due to the input.

To do so, we need a measure for the amount of information between the output and the input. We take a small detour to achieve this goal. First, we define the Kullback-Leibler (KL) divergence (also called relative entropy), which is a measure of similarity between any two probability distributions, P and Q:

$$D(P||Q) = \sum_k P(k) log_2 \frac{P(k)}{Q(k)} \tag{1.2}$$

KL is always positive and equal to zero if and only if Q and P are equal. This measure can also be interpreted as the average number of bits lost encoding samples

of P assuming Q as a model for the distribution.

We define the mutual information $I(X,Y)$ as a special case of KL, in which: $P = \mathcal{P}(x,y)$ and $Q = \mathcal{P}(x)\,\mathcal{P}(y)$:

$$I(X,Y) = D(\mathcal{P}(x,y)\,||\mathcal{P}(x)\,\mathcal{P}(y)) = \sum_{x,y} \mathcal{P}(x,y)\, log_2 \frac{\mathcal{P}(x,y)}{\mathcal{P}(x)\,\mathcal{P}(y)} \tag{1.3}$$

The mutual information is equal to zero if and only if $\mathcal{P}(x,y) = \mathcal{P}(x)\,\mathcal{P}(y)$; that is, by definition, if and only if the input and output are independent.

We now express the mutual information in terms of entropy. Using Bayes rule in equation 1.3 we obtain:

$$I(X,Y) = \sum_{x,y} \mathcal{P}(x,y)\, log_2 \frac{\mathcal{P}(x|y)}{\mathcal{P}(x)} \tag{1.4}$$

This is equal to:

$$I(X,Y) = \sum_{x,y} \mathcal{P}(x,y)\, log_2 \mathcal{P}(y|x) - \sum_{x,y} \mathcal{P}(x,y)\, log_2 \mathcal{P}(y) \tag{1.5}$$

The second term in the expression summed over all x is just: $-\sum_y \mathcal{P}(y)\, log_2 \mathcal{P}(y)$, which is $H(Y)$. Using Bayes rule, the first term in the expression can be written as: $-\sum_x \mathcal{P}(x) \sum_y \mathcal{P}(y|x)\, log_2(\mathcal{P}(y|x))$. But this is just the negative of the conditional entropy of the output given the input:

$$H(Y|X) = -\sum_x \mathcal{P}(x) \sum_y \mathcal{P}(y|x)\, log_2(\mathcal{P}(y|x)) \tag{1.6}$$

The inner sum indicates the entropy of the output for each input, and the outer sum indicates that this is averaged over all possible inputs. $H(Y|X)$ should be interpreted as whatever entropy is left in the output that is not due to the input. If the output is deterministically given by the input, then $H(Y|X)$ is just 0.

We now have (from the first and second terms of equation 1.5)

$$I(X,Y) = -H(Y|X) + H(Y) \tag{1.7}$$

Thus the mutual information is the entropy of the output that is due only to the input.

An efficient representation is formed by maximizing the mutual information between the input and the output. Many studies have limited their analysis to noiseless systems, and maximizing entropy, as a simplifying assumption. In general, to maximize the entropy or the mutual information, certain constraints about the system need to be imposed.

Marginal and Joint Entropies

For simplicity, we limit the discussion to entropy maximization. We'd like to formulate entropy maximization in terms of the two types of redundancies described earlier. Specifically, we differentiate between the joint and the marginal output entropies. We denote $Y = Y_1, Y_2, ..., Y_n$ the set of output signals through n channels.

It can be proven that the following inequality holds between the joint and the marginal entropies of Y_i:

$$H(Y) = H(Y_1, Y_2, ...Y_n) <= H(Y_1) + H(Y_2) + ... + H(Y_n) \tag{1.8}$$

with equality achieved if and only if:

$$\mathcal{P}\left(y_1, ...y_n\right) = \prod_1^n \mathcal{P}\left(y_i\right), \tag{1.9}$$

That is, if and only if the Y_i are independent.

Thus, maximal entropy is achieved when the Y_i are independent, and each of the marginal entropies $H(Y_i)$ are also maximal. In other words, different outputs are representing independent parts of the information, and the output through each channel transmits the maximal information possible.

Maximizing Marginal Entropy

Entropy can be increased by considering the output through a single channel (i.e., the marginal distribution). Intuitively, the output through a channel carries more information if it is as random (and unpredictable) as possible. However, the optimal distribution is also a function of the constraints on the channel, and is determined by solving a constrained optimization problem. Consider a constraint of the form $E(f(x)) = c$, where x is the response, $f(\cdot)$ is a constraint function, $E(\cdot)$ is the expected value, and c is a constant. Then it can be shown that the Maximum Entropy distribution is $P(x) \propto e^{-f(x)\lambda(c)}$, where $\lambda(c)$ is the Lagrange multiplier [72, 38, 40] (note that solving for λ for a given c is often difficult). For example, if a channel transmits data that have a constraint on the maximal value, the Maximal Entropy distribution is uniform. If a channel has a constraint of fixed mean, the Maximal Entropy distribution is exponential. If a channel has a constraint of fixed variance, the Maximal Entropy distribution is Gaussian.

Maximizing Statistical Independence

Entropy can also be increased by considering the co-occurrence of outputs through multiple channels (i.e., the joint distribution): these should be statistically independent. This can be achieved by minimizing the KL divergence between the joint

distribution and the product of marginal distributions of the output. However, this is in general difficult to achieve computationally, because characterizing the joint histogram of the output grows exponentially with the number of dimensions. Therefore, independence has been typically restricted to a particular domain that is easier to work with. A large body of research has restricted the output to be a linear function of the input. In addition, one can further make restrictions on the type of statistical dependency to be eliminated.

An algebraic method for eliminating second order correlations (i.e., decorrelation) is known as Principle Components Analysis (PCA). Decorrelation is defined as:

$$E(Y_i Y_j) = 0, i \neq j, \tag{1.10}$$

where $E(\cdot)$ is expected value. Decorrelation is a weaker constraint than independence: statistical independence implies statistical decorrelation, but the opposite is not necessarily true.

Principal component axes can be recovered using an eigenvalue analysis (see for example, [69, 140]). These axes always exists, though they need not be unique. This is partly due to repeated eigenvalues, and partly due to the fact that whitening transforms retain their properties under post-rotation. It can be shown that PCA will recover independent components if the signals are drawn from a Gaussian distribution and linearly mixed. However, in the general case, PCA does not recover independent components.

Independent Components Analysis (ICA) was devised as a method for eliminating higher order statistical dependencies in a linear transformation (see for example, [75, 32, 69, 37, 20]). Comon introduced a rigorous mathematical solution

that uses up to fourth order statistics [37]. Other techniques have been developed to incorporate higher order statistics. These typically are not quadratic and so cannot be solved using linear algebra, but can be estimated with gradient ascent or other methods (e.g., [20, 69]). In general, ICA techniques must make some assumption about the probability density function(s) of the independent components, or at least use a parameterized density that is estimated during the procedure. The underlying density is not necessarily known apriori, and wrong assumptions could lead to erroneous results. An efficient way to estimate both the parameters of the linear transformation leading to independent components, and the parameters of the underlying density, is using Estimation Maximization (EM) (e.g., [8, 40]).

We want to stress that even Independent Components Analysis does not guarantee the resulting components will in fact be independent. In particular, they will only be as independent as possible assuming the linearity constraint.

1.1.3 The Efficient Coding Hypothesis

Attneave pointed out that if we consider perception as an information-handling process, it becomes evident that much of the information received by an organism about its environment (e.g., images and sounds) is redundant in both space and in time [11]. For example, the intensity vales of two co-occurring neighboring spatial regions (or pixels) of an image tend to be similar. To an observer with knowledge of these structures, one spatial or temporal portion of a given signal may be predicted from others. Redundancy in images can also be quantified perceptually, by asking subjects to replace missing pixels in a digital image [77]. Thus, a goal of sensory processing should be to encode information in a form less redundant and more

efficient than that in which it impinges upon the receptors [11].

Barlow suggested a neural version of this hypothesis: that a goal of sensory processing should be to increase independence between neuronal responses, when exposed to natural stimuli [16]. The responses of neurons are action potentials. But in the discussion, we will often consider only the information given by the firing rate. We will also specify several studies that have used a more detailed description of the spike train (see [115, 40] for theoretical discussion).

1.1.4 Biological Relevancy of Efficient Coding

Efficient coding has been widely used in engineering, in numerous applications that require less storage space, or achieving a smaller dynamic range during the transmission of signals. Prominent examples today are the compression of sounds, images, and movies.

But to consider efficient coding and its biological relevancy, requires an understanding of what it means to be biologically efficient. Barlow in his initial work suggested that efficiency could be achieved by reducing the frequency of spikes (and therefore, reducing metabolic energy) in neurons carrying the representation. Additionally, single neurons have constraints in their representations; for example, animals encounter a wide range of stimuli in the environment, but neurons typically have a limited dynamic range, a limit to the range of inputs that they can respond to differentially. In this sense, neurons are required to be efficient in their representations. Perceptually, it has been suggested that a more independent representation may facilitate tasks such as recognition of novel patterns and associative learning (for example, it is more computationally tractable to compute a joint distribution

as the product of marginals) [6, 17]). Independent representations have also been motivated by a generative view of signal formation: if signals in the world are formed from independent sources, then ultimately, an independent representation should be able to recover these sources.

A common vein of controversy regarding efficient coding and sensory neural processing concerns the number of output channels. An important goal in signal processing is often to reduce the data dimensionality, and recent techniques have been promising in their ability to do so while preserving the signal structure (e.g., local linear embedding: [121, 149]). But in the sensory system, if there are significantly more neurons available in the output than in the input, then space is abundant for representing the information, and efficiency might not be an issue at all. Indeed, forty years after stating his efficiency hypothesis, Barlow himself argued that the increase in number of neurons from retina to cortex (from $2 * 10^6$ retinal ganglion cells to 10^9 V1 cells) is counter to the notion of efficiency [15]. However, since different neurons might serve different purposes and project to different areas of the brain, it would be difficult to quantify efficiency purely based on a neural count. Finally, redundancy reduction need not imply that the redundant information is lost and cannot be re-created; it only implies that the system uses a more efficient representation to work with.

The concept of efficient coding in the nervous system is appealing, yet controversial. Ultimately, the hypothesis needs to be tested. Approaches for doing so are described in subsequent sections.

1.1.5 Alternative Coding hypothesis: Sparse Coding

An alternative coding hypothesis that has been suggested is sparse coding: that a goal of sensory processing should be to produce a sparse representation of natural scenes [51, 52, 171]. Field interpreted this to mean that each neuron has an equal response probability across the class of images, but a low response probability for any one input signal in the ensemble. This idea was motivated by the observation that natural images are sparse: they contain regions without very much information, interspersed with regions of occasional prominent information, such as edges.

Unlike statistical independence, there is no formal mathematical definition for sparseness. But one of the most common measures used is kurtosis, in which higher kurtosis is associated with a sparser distribution. Kurtosis of Y_i (uncorrelated and variance 1) is defined as $E(Y_i^4)-3$. The kurtosis of a Gaussian is 3; a superGaussian will have kurtosis larger than 3, resulting in long tails and a large peak at 0. Field observed that natural images through typical visual filters are highly kurtotic [52], and Attias obtained similar results for natural sounds through typical auditory filters [9] (but see Baddeley, who has argued that high kurtosis in and of itself does not necessarily imply anything interesting about the input stimuli [12]).

Sparseness and redundancy reduction are sometimes used synonymously. However, it's important to keep in mind that a sparse (or highly kurtotic) representation is a description of the marginal distribution, corresponding to responses of a single neuron. Information theoretic considerations of the optimal distribution through a single channel would not necessarily lead to a sparse distribution and would depend on the particular constraint (for example, for a fixed variance, the Gaussian is preferable to a superGaussian distribution). Taken to its extreme, sparseness

could be accomplished by so-called "grandmother cells", that respond only to one stimulus. But from the point of view of a single neuron, the entropy is extremely low. A sparse distribution does not (on its own) imply anything about the joint distribution of the population of neurons. It might seem that if every neuron has a low probability for being active to a given stimulus, then neurons will on average act independently; but this is not guaranteed. In particular, observing sparseness in neurons experimentally, is not an indication of whether the neural responses are independent.

Nevertheless, in the linear domain, maximizing independence and maximizing sparseness in images (while minimizing the error between the original image and its reconstruction from the components) have led to similar results [22, 103]. The distinction is subtle, because ICA on natural images works best when assuming a sparse prior for the marginal distributions; subGaussian priors do not work well (for example, the estimated independent components do not match the probability distributions of the prior). It turns out that Bell and Sejnowski's formulation (maximizing independence) and Olshausen and Field's formulation (maximizing sparseness) are very similar under a maximum likelihood framework [102]. Therefore, for a linear transform and natural images, the most independent possible components are sparse.

1.1.6 Testing the Statistical Hypotheses

Testing the Hypotheses Through Experiments

Experimentally, the efficient coding hypothesis can be tested by examining the statistical properties of neural responses under natural stimulation conditions.

A number of studies have examined the firing rates of single neurons. One of the earlier studies, by Laughlin, demonstrated an interesting link between experimental evidence and the contrasts of natural scenes in the fly environment [80]. He measured experimentally the function relating contrast to membrane potential, and showed that this function approximately transforms the distribution of contrasts in natural scenes to a uniform distribution. This is consistent with entropy maximization with a constraint on the maximal firing rate. Baddeley et al [13] demonstrated that the firing rates of spiking neurons in primary visual cortex are exponentially distributed when stimulated with natural scenes. This is consistent with entropy maximization with a constraint on the mean firing rate Additionally, this may be viewed as a "sparse" distribution. Vinje and Gallant claimed that the responses of primary visual cortex neurons to natural scenes are sparse. They used a non-parametric statistic of the neural activity fraction to a natural movie for each individual neuron they recorded from [159]. They did not express a particular form of marginal distribution, such as Gaussian or exponential.

Few studies have also looked at information maximization of spike trains (as opposed to firing rates) through single neurons. Rieke et al. showed that the rate of information transmission of spike trains in bullfrog primary auditory afferents is higher for naturalistic sounds than for broadband Gaussian white noise [114]; and similarly, Attias showed that the rate of information transmission of spike trains in cat auditory midbrain neurons is higher for naturalistic than for non-naturalistic stimuli [10].

Gawne et al. tried to measure the mutual information between responses of neurons in inferior temporal cortex and primary visual cortex [57, 56]. They re-

ported around 20 percent mutual information for complex cells in V1, and less in inferior temporal cortex. But the stimuli were artificial and not constructed to resemble natural images. Therefore, they do not provide a test of the efficient coding hypothesis. Chechik et al. examined redundancy reduction in the auditory system between groups of neurons using criteria of mutual information [55]. They recorded from cat using bird vocalizations and demonstrated that redundancy is reduced when proceeding from peripheral to more central areas. The analysis in these studies was performed on multiple neurons, but the recordings were done non-simultaneously (single neurons) using a fixed set of stimuli.

Ultimately, one would like to test for statistical independence to naturalistic stimuli by recording simultaneously from pairs (or groups) of neural responses. Since such multi-cellular experiments are now possible, it is our hope that future experimental work will offer more concrete answers to the efficient coding hypothesis.

Testing the Hypotheses Through Modeling

The ideas of efficient representation have led many researchers to look for a means of "deriving" a model of sensory processing purely from a statistical characterization of natural signals. The properties of the resulting model are then compared to properties of biological sensory neurons.

A wealth of studies have focused on redundancy reduction via reducing statistical dependencies or optimizing for sparseness. In these studies, the decomposition of natural signals have typically been restricted to the linear domain. The goal then is to decompose natural signals into linear components, that are optimal under some

statistical criteria: decorrelation, independence, or sparseness (Figure 1.2). The resulting linear projection functions are then compared qualitatively to receptive fields or filters in vision and audition. Interestingly, these directions have been fruitful in explaining basic properties of early sensory neurons.

A number of researchers have attempted to use PCA concepts to derive linear receptive fields similar to those determined from physiological measurements [101, 127]. The principal components decomposition is, however, not unique. Therefore, these early attempts required additional constraints, such as spatial locality and/or symmetry, in order to achieve functions approximating cortical receptive fields.

More recently, ICA and sparseness optimization techniques have yielded linear receptive fields localized in orientation, spatial frequency and spatial position [102, 22]. These are matched to qualities of receptive fields of simple cells in primary visual cortex [156], first discovered in the pioneering studies of Hubel and Wiesel [67]. ICA has been applied to numerous other findings in vision. ICA produces complex cell properties, by extending the procedure to operate on subspaces [68]. Additionally, ICA on a temporal sequences of image patches has yielded spatio-temporal receptive fields similar to primary visual cortex direction selective cells [156]. ICA on natural sounds has typically yielded localized bandpass filters qualitatively similar to the peripheral auditory system [21, 82].

Other approaches have applied linear techniques, assuming a linear system with additive Gaussian white noise. Atick and colleagues have shown that eliminating second order statistics (decorrelation) in natural scenes using this model, results in receptive fields with spatial frequency tuning similar to the retina [7, 6]. Specifically, for high intensities, the linear kernel is bandpass; but for low intensities, in which

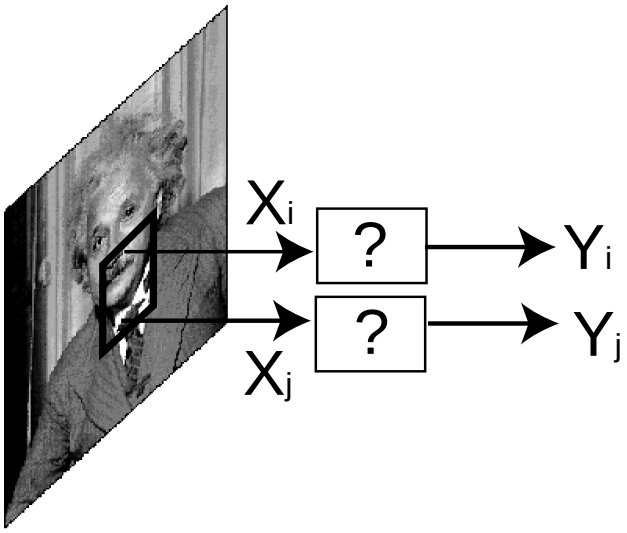

Figure 1.2. Restriction of decomposition of natural signals into linear components. Find projection functions (question marks) that maximize some statistical criteria: decorrelation, independence, or sparseness. These projection functions are often compared to linear filters or receptive fields of neurons. For simplicity, we have only drawn two such projection functions. X_1 and X_2 are input samples from a signal. For natural images, many patches of image are taken as input samples. Here we have drawn just one patch in an image. Here the number of pixels in a patch determines the number of input signals, the number of projection functions, the number of pixels in each projection function, and the number of output signals. More generally, the number of outputs may differ from the number of inputs. For natural sounds, patches are replaced by portions of the one dimensional signal.

the noise dominates, the linear kernel is low-pass. Similar approaches have been applied to explain temporal processing in the LGN [48].

Linear models provide an important starting point for linking between statistical regularities in natural signals and neuronal processing. However, we suspect that the statistical properties of natural signals are too complex to expect a linear transformation to result in an independent set of components. For example, even if one assumes that the objects that constitute a visual scene are drawn independently, the most common combination rule for image formation is *occlusion*, which is nonlinear. In addition, sensory neurons exhibit highly nonlinear behaviors (see section below).

Some studies have addressed nonlinear processing at the level of the retina. Ruderman and Bialek [122] discussed division by local contrast as a means of achieving maximal information transmission (optimizing for a marginal Gaussian distribution). This is optimal under the assumption that the neural response has a fixed variance. This methodology does not introduce criteria of statistical independence. Balboa et al. have explained nonlinearities in the retina using a different type of statistical criteria: probability of presence of an occlusion border as a function of contrast [14]. Their statistical hypothesis is not based on redundancy reduction, but rather on the notion of extracting signals that are presumed relevant for an organism.

Additional statistical frameworks have been incorporated in sensory neural modeling of cortex. Statistical models that infer the probable causes of sensory input in a hierarchical fashion have been related to the function of bottom-up and top-down neural processing in cortex [41]. Rao and Ballard have demonstrated a number of

physiological nonlinear behaviors in primary visual cortex using a predictive feedback model trained on natural images [112].

1.1.7 Open Questions in Theory Driven Models

A considerable amount of progress has been made in linking the efficient coding hypothesis to neural processing, and this link has become more quantitative over the past decade. Redundancy reduction using statistical independence criteria is appealing because it doesn't require additional constraints on the response values of a neuron. In addition, it has been argued that most of the redundancy in natural signals results from statistical dependencies rather than marginal statistics [6]. But models thus far of statistical independence have been restricted to a linear transformation. It would be interesting to check if these models in fact achieve independence, and if not, to explore nonlinear extensions. This route may help in understanding the nonlinear behaviors of sensory neurons described below.

1.2 Physiology Driven Sensory Neural Models

We first describe physiological data that have led to the development of nonlinear neural models in retina, primary visual cortex, and the peripheral auditory system. A more complete discussion about these neural systems (including linear receptive field characterizations) may be found in [90, 166, 67, 93, 92, 107, 123].

Here we particularly pay attention to three types of nonlinear behaviors: response saturation, suppression by non-optimal masks, and changes in tuning curve shape for different input levels. Neural responses in the experiments are given by

the mean firing rate.

1.2.1 Physiological Nonlinearities in Sensory Neurons

Response Saturation

In a linear system, scaling the input level should scale the responses by the same amount. We denote a "response curve" as the mean firing rate to a given stimulus, computed for different intensity levels of the input. Sensory neurons typically do not obey this form of linearity: their response curves saturate for high intensities of the input. This is evident in the retina for high illumination [134]; in primary visual cortex for high contrast [86, 3, 99, 25, 30]; and in the basilar membrane and auditory nerve for high sound pressure level [59, 63, 107, 123]. The level of saturation is often higher for an optimal stimulus than for a non-optimal stimulus [30, 70].

Response Suppression

In a linear system the response of two inputs can be predicted as the sum of the responses of each individual input. But in sensory neurons response to a stimulus can often be suppressed by a non-optimal mask, that sometimes by itself does not elicit any response.

In a classical experiment in retinal ganglion cells, a test spot is flashed on a background of changing luminance. When background luminance is high, the response to the test flash is decreased. In particular, the response curve as a function of log luminance shifts to the right [134]. Suppressive phenomenon are also documented as a function of contrast. To demonstrate this effect, an optimal bar is chosen inside

the classical receptive field (i.e., the region that elicits an excitatory response). The response to the optimal stimulus may be suppressed by a peripheral sine grating outside the classical receptive field that by itself does not elicit a response [134]).

In primary visual cortex, suppression by a non-optimal mask is abundant. The stimulus is usually chosen optimal for the cell and inside the classical receptive field. A widely documented phenomenon is known as cross-orientation suppression. When a mask stimulus of non-optimal orientation (and typically chosen as orthogonal) is superimposed on the optimal stimulus (forming a plaid), the response of a neuron is suppressed [24, 30]. Responses of a neuron are also suppressed by non-optimal frequencies [30]. In addition, suppression has been widely observed when a mask stimulus is placed in a region outside the classical receptive field of the neuron [23, 83, 81, 43, 27, 79, 34, 33]. Hubel and Wiesel already realized that extending a stimulus beyond its receptive field along its iso-oriented access may elicit suppression [67]. They thought this was a quality of a special class of neurons which they termed "hypercomplex". Today we attribute suppressive effects even to so-called "simple cells". When extending an optimal stimulus in its iso-oriented axis this suppression is often denoted "end inhibition" and in the orthogonal axis "side inhibition". More commonly today, the suppressive regions outside the classical receptive field are simply termed "surround" (not to be confused with the surround that is part of the classical receptive field and subtractive!). In contrast, the classical receptive field is often denoted "center". Suppression typically shifts the contrast response curve to the right on a log contrast scale [24, 30, 34, 33].

To further characterize these non-linearities, many studies have tested for suppression as a function of orientation, frequency, and spatial position of a mask.

When the mask is placed in the center, cells exhibit relatively broad tuning with respect to the orientation and frequency of the mask [24, 44]. Some studies have found less suppression at the preferred orientation or frequency [18, 155]. However, it is difficult to determine the extent of suppression for masks of similar orientations or frequencies as the stimulus, unless it is verified that the mask does not introduce excitation in addition to suppression [97]. One way to override this problem is to use a null-position mask, that is a phase that does not elicit any response at all in the cell. In any case, it is well established that most cells can be suppressed by a mask of orthogonal orientation or frequency greater than one octave away [30]. This suggests that suppression inside the classical receptive field results from a neuronal pool tuned to many frequencies and orientations.

In the surround, numerous studies have demonstrated that most suppression is exhibited when the mask is at the preferred orientation [83, 81, 43, 27, 79, 34, 33], or at the preferred direction of drift [81], or at the preferred frequency [81, 94]. Facilitation has also been reported [87, 83, 61, 81, 139]. Some of these facilitative effects might be explained by the use of masking stimuli that inadvertently excite the receptive field of the neuron. Facilitative effects might also be explained by disinhibition, in which a third cell inhibits a second cell, thus releasing its inhibition of the recorded cell.

Auditory nerve fibers exhibit two tone suppression, i.e., suppression by a tone of non-optimal frequency (that sometimes itself does not elicit a response). Amount of suppression is dependent on the frequency of the mask in relation to the preferred frequency of the fiber, with stronger suppression for masks that are closer in frequency [71, 63, 59]. Suppression typically shifts the rate level curves to the right

on a log (decibel) scale.

Changes in Tuning Curve Shape

We denote "tuning curve shape" as the mean response rate of a neuron as a function of some parameter of the stimulus (e.g., frequency, orientation, or radius length). In a linear system, the tuning curve shape should not change when computed for stimuli with a low intensity, versus for stimuli with a high intensity. High intensity stimuli induce a larger response, but the shape of the curves should remain constant.

This property sometimes holds in sensory neurons. In primary visual cortex, tuning curve shape as a function of orientation or spatial frequency (for a stimulus placed inside the classical receptive field) remains roughly constant when computed for different stimulus contrasts [130, 145].

But there are a number of examples in which this property is not retained. One example was recently discovered in primary visual cortex [128, 76, 33]. The experimental stimulus is a grating of optimal orientation and spatial frequency for the cell. The mean response rate is measured as a function of stimulus radius at different contrasts. The shape of the resulting radius tuning curves are different for low contrasts and for high contrasts: for higher contrasts the response peaks at a smaller radius [128, 76, 33].

Examples are also found already at the retinal level. The temporal transfer function shape of retinal ganglion cells changes as a function of contrast, and shifts towards higher temporal frequencies at high contrasts [135, 136, 137, 35].

Basilar membrane and auditory nerve fiber frequency tuning curve shape varies for different sound pressure levels. This has been known for years: it is narrow for

low sound pressure levels and wide for high sound pressure levels [120, 107].

1.2.2 Nonlinear Models Stemming from Physiological Observations

Many successful neural models do not disregard the idea of linear processing as a first stage; rather they build nonlinear behaviors on top of the linear model.

To motivate the type of models that have evolved to explain sensory nonlinearities, we first consider a model in which the mean firing rate is determined by the projection of a stimulus onto a linear filter followed by a static nonlinearity (such as a squaring rectification). Suppose we want to fit response curves as a function of stimulus intensity. Then the response to a given stimulus according to this model is proportional to its intensity (squared). This type of model could potentially explain response saturation at high intensity levels. However, it could not explain different saturation levels for optimal and non-optimal stimulus response curves (unless a different static nonlinearity is used for the optimal and non-optimal stimulus); nor could it explain rightward shifts of contrast response curves in the presence of an added mask stimulus.

To characterize these type of behaviors, it is useful to consider the Naka-Rushton (or Michaelis-Menton) equation [95, 19]. This has often been used to fit response curves as a function of intensity both in vision and audition (e.g., [169, 125, 126, 3, 47, 131]):

$$R = \frac{R_{max} K^2}{K^2 + \sigma^2} \tag{1.11}$$

where K corresponds to illumination, contrast, or sound pressure level; R_{max} is the maximal attainable response to a stimulus; and σ^2 is the semi-saturation constant

(e.g., the input level at which R reaches half its maximum value). Here the power of K is set to 2. More generally, the power can be adjusted, and determines the slope of the response curve.

Figure 1.3 plots the Naka-Rushton equation for different values of R_{max} and σ. From inspection of the figure, it is clear that experimentally observed rightward shifts of response curves (due to masks) can be fit using different values of σ; and changes in saturation level (for an optimal versus a non-optimal stimulus) can be fit with different values of R_{max}. Thus, for a single neuron, the parameters of the equation can be fit separately for each stimulus in the experiment. But rather than fitting each response curve separately, we'd like one generic model that can explain the responses of a neuron to different experimental stimuli.

The Naka-Rushton equation can be modified to potentially explain the rightward shift of response curves observed physiologically for different masks, by incorporating a gain signal (dependent on the intensity of the input signal) in the denominator. Thus, for example, a mask that alone elicits no response, may still control the gain of the linear front end. The gain acts like σ, to shift the response curve to the right. Incorporation of gain into the Naka-Rushton equation follows the discussion in Shapley and Enroth-Cugell on the Retina [134]. They formulated a gain signal that depends on the stimulus illumination over time $G = G(t, K(t))$, and that modifies σ directly [134]. Victor [158] proposed a more specific contrast gain control model of the retina in the temporal domain, in which the gain signal consists of time-delayed copies of the linear receptive field kernel.

A unifying theme in many successful neural models of sensory nonlinearities is gain control (e.g., [113, 134, 64, 84]). In general, these models assume that the linear

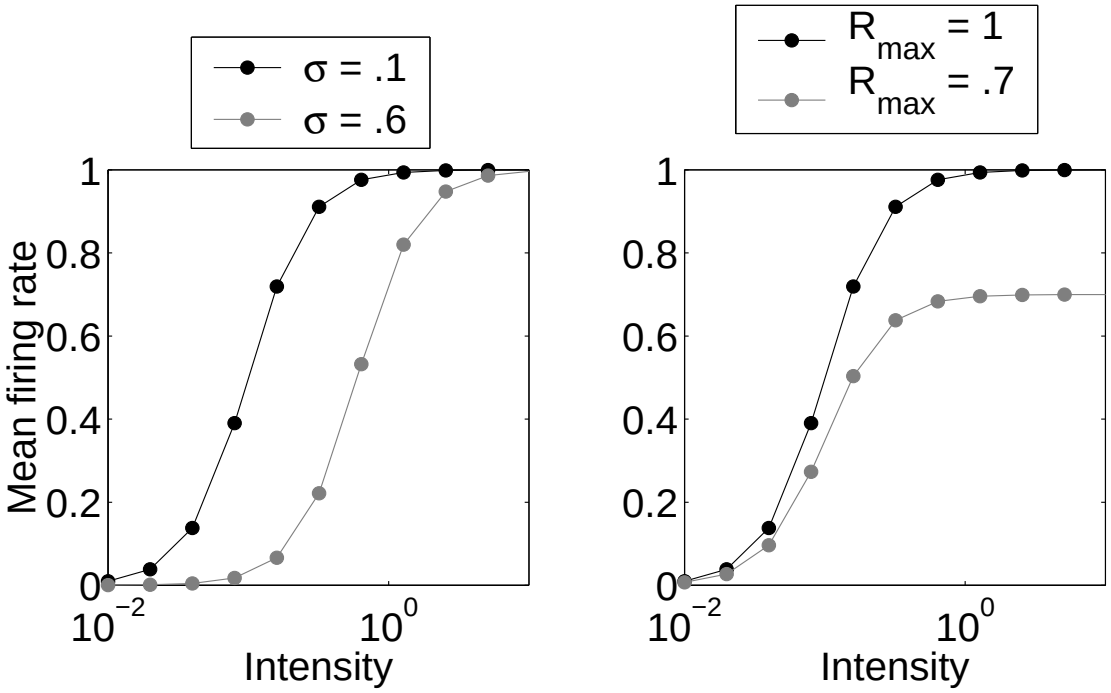

Figure 1.3. Behavior of Naka-Rushton equation. Left, The Naka-Rushton equation for a constant $R_{max} = 1$ and variable σ. Note that higher values of σ shift the response curve to the right on a log axis. Right, The Naka-Rushton equation for a constant $\sigma = .1$ and variable R_{max}. Note that lower vales of R_{max} reduce the saturation level of the response curve.

front end (often followed by a squaring or some other form of static nonlinearity) is modulated by a gain function dependent on the stimuli. Details of the model have varied, depending on the specific processing level and modality, and detail of implementation.

Nonlinearities in primary visual cortex have been modeled using a specific instantiation of a gain control mechanism, which Heeger denoted "divisive normalization" [64, 65]. In addition to describing a specific gain function (and accounting for rightward shifts), the model also accounts for different saturation levels for optimal and non-optimal stimuli. Ideas about gain control in primary visual cortex also originated from the work of [99, 100, 2, 24].

In divisive normalization, the gain is assumed to be set by a population of neurons with different tuning properties:

$$R_i = \frac{R_{max} L_i^2}{\sum_j L_j^2 + \sigma^2} \tag{1.12}$$

where L_i is the linear response of filter i, and L_j are the responses of a population of filters (including filter i). The response of each filter contributes the same weighting to the gain in the expression, $\sum_j L_j^2$. It is also assumed that the filters in the population tile the full orientation and spatial frequency space.

A mask stimulus will increase $\sum_j L_j^2$, and cause a rightward shift of the contrast response curve. Additionally, a non-optimal stimulus will reduce L_i^2 in the numerator without changing the gain in the denominator (since the filters in the population tile the full space, and their weighting is equal). As a result, the saturation level for the non-optimal stimulus will be reduced.

One of the original motivations for the divisive normalization model was its

preservation of tuning curve shape as a function of contrast, even in the face of saturation. This is matched to the observation that inside the classical receptive field orientation tuning curves stay roughly the same shape for different contrasts. We consider this model for computing an orientation tuning curve at different contrasts. The argument is essentially the same as above, for saturation level. The gain signal does not depend on orientation (due to the tiling and equal weighting); it depends only on contrast. So we write the divisive equation as proportional to one part that only depends on the linear response of the chosen filter to a particular orientation, and another part that only depends on contrast:

$$R_i = |L_i(\phi)| R_{max} \frac{K^2}{K^2 + \sigma^2} \tag{1.13}$$

where K represents contrast, and ϕ an orientation. Thus, the shape of the orientation tuning curve does not change as a function of contrast.

The divisive normalization model has successfully accounted for nonlinearities inside the classical receptive field [64, 30]. However, the model cannot explain some data in the surround. We've seen that physiological tuning curve shapes are not always preserved. In addition, the amount of rightward shift for masks in the surround is dependent on properties of the mask (such as orientation). To account for these data, we propose that a less constrained model is needed, in which the divisive weighting of each filter response in the denominator need not be equal. Cavanaugh et al. have also suggested this to explain their physiological data collected with stimuli outside the classical receptive field [34, 33].

Gain control has also been central in explaining peripheral auditory processing (e.g., [84, 59, 119]). Phenomenological auditory models have captured nonlin-

earities using a parametric set of linear filters (corresponding to neurons), whose parameters, such as bandwidth of filter, are modulated by the gain of the same filter over time as well as other filters that respond to the stimulus [84, 119]. A recent model captures frequency interactions, in a fashion that nearby frequency bands reduce the gain more than far away frequency bands [119]. This may be viewed as a weighted gain control model.

The physiological mechanisms underlying gain control is an active area of research, and potential candidates include biochemistry (e.g., in the outer hair cells [59, 173] or photoreceptors); spike generation; feedforward synaptic depression (e.g., [1]); and shunting inhibition generated by other neurons [29, 30].

1.2.3 Open Questions in Physiology Driven Models

Models stemming from physiology have made a large stride in our understanding of sensory nonlinearities. However, the choice of how to set the parameters (or weights) of the gain of different filters is often designed to fit the data. In primary visual cortex, the choice of equal weighting can explain data inside the classical receptive field, but cannot explain recent data outside the classical receptive field.

In addition, the purpose of these neural nonlinearities are still not well understood. Gain control is often explained as a mechanism for neurons to deal with their limited dynamic range. However, less clear is why this would occur more along specific dimensions (orientation, frequency, and so on) as seen in the peripheral auditory system and in primary visual cortex.

1.3 Spike-triggered (White Noise) Characterizations

White noise analysis has emerged as a powerful computational technique for characterizing the response properties of spiking neurons. Stimuli are drawn randomly from an ensemble and presented to a neuron in rapid succession, and one examines the subset that elicit action potentials. This "spike-triggered" stimulus ensemble provides information about the neuron's response characteristics. The stimuli used in these analyses are (as expected) typically white. However, extensions of this analysis have been applied to a reduced linear subspace of input stimuli [117], and to naturalistic stimuli [116, 150].

In the most widely used form of this analysis, one estimates a linear receptive field from the spike-triggered average (STA); that is, the mean stimulus that elicited a spike. In essence, this is similar to how Hubel and Wiesel characterized linear receptive fields in V1. But rather than hand-selecting the stimuli, these techniques explore a wider stimulus space; they are not prone to problems of neural adaptation to the stimulus; and they can be formulated quantitatively, assuming an underlying neural model.

DeBoer first used white noise analysis in the auditory domain to study the temporal structure of filters in the auditory nerve [45]. Jones and Palmer used white noise to recover spatial receptive fields in area V1 in the cat [74]. DeAngelis et al. recovered spatio-temporal receptive fields in V1 [43].

We illustrate the concept of white noise analysis in Figure 1.4. For simplicity, we assume that an experimenter wants to study the characteristics of a neuron as a function of time only. At each time frame, a full-field stimulus is presented to

a neuron. Its intensity is chosen from a Gaussian white noise distribution. The spike-triggered matrix is constructed as follows: each time the neuron spikes, the samples of stimuli preceding a spike (in a certain time window) are added as a new row to the matrix. Thus, the dimensions of the spike-triggered matrix are the length of the time window by the number of stimuli eliciting spikes. The average of the columns of this matrix results in the STA; this is a temporal filter, with dimensions equal to the length of the time window. More generally, to study the full spatial temporal representation in neurons, at each time frame, length by width spatial samples are chosen randomly, and the STA results in a full spatio-temporal filter.

Spike-triggered averaging should be understood in the context of an underlying neural model. Many models assume that spikes are generated using a Poisson process, and that the mean firing rate is given by a linear projection onto a kernel followed by a static nonlinearity:

$$\mathcal{P}\left(spike|\vec{s}\right) = g(\lfloor \vec{K}\vec{S} \rfloor) \tag{1.14}$$

with g a monotonic point nonlinearity, and $\vec{K}$ a linear kernel. It has been shown that under these assumptions, the STA provides an unbiased estimate of the linear kernel (a proof is provided in [36], and see also chapter 4). In addition, the monotonic point nonlinearity can be retrieved [36, 98]. Under this model, computing the STA for any stimulus intensity should retrieve the same kernel. But there are physiological examples in which the STA shape changes as a function of the stimulus intensity (e.g., [35, 128]). This suggests that a linear representation followed by a point nonlinearity is not sufficient, and brings up the need to explore nonlinear

neural representations using white noise analysis. In particular, it is interesting to ask whether a neural model with gain control could produce the type of biases in the STA observed physiologically.

It is also useful to analyze other statistics, in addition to the average. Some authors have estimated the covariance matrix associated with the spike-triggered stimulus ensemble. An eigenvector analysis of this matrix can determine the stimulus directions that account for most of the variance. Thus, this method describes a neural characterization that is not given by a single linear receptive field, but rather by some combination of filters or stimulus axes. These ideas (known as spike-triggered covariance, or STC) were first developed by de Ruyter and Bialek [42].

An example of this technique is in the characterization of the stimulus axes that make up a complex cell in V1 [152]. Complex cells have a spike-triggered average that is roughly zero. Thus, the STA is uninformative. However, STC analysis reveals two filters that are 90 degrees out of phase. Both filters have an excitatory effect on the neural response; but the specific function that combines the two filters into a neural response is not known from the analysis. In this simple case, it is easy to "guess" (and then demonstrate) how they are combined, since the sum of the squared responses of the filters produces a phase-invariant response, as expected from complex cells. Spike-triggered covariance might offer a potential path for expressing neural nonlinearities as combinations of several filters. But it has not been demonstrated if and how this may be connected with the type of nonlinearities typically observed in sensory neurons.

Another possibility for characterizing neural nonlinearities may be through the

Wiener/Volterra expansion. These constitute an approximation of a function (in this case, the neural response) as a sum of a power series [168, 160, 115]. Volterra generalized the Taylor series into an expansion, where the inputs and outputs are functionals (of time, space, etc.). As a result, each term in the expansion consists of a kernel (or filter), and kernels of higher power are denoted higher order kernels. The Wiener expansion is a reformulation of Volterra, with kernels that are easier to compute. The first order Wiener kernel is the cross correlation of the stimulus with the spike train; the second order kernel is the cross correlation of the stimulus to the second power with the spike train; and so on.

If we (or a neuron) could calculate an infinite order of Wiener kernels, then in principle any (reasonably smooth) nonlinear function could be estimated. However, large amounts of data are needed to measure high order Wiener kernels (e.g., beyond second or third order). Thus, Wiener analysis is only useful if the first few Wiener kernels do a good job at approximating the nonlinearities observed in neurons. But the type of divisive gain model discussed earlier is not well described by a low-order power series expansion (see also discussion in [115]). Marmarelis has suggested ways of describing a Wiener expansion with feedback, which could potentially explain changes in the STA as a function of stimulus level [89]. In this approach, one would need to determine the type of feedback model that is consistent with the physiological data.

The spike-triggered covariance methodology should not be confused with a second-order Wiener analysis. In such a Wiener analysis, the neural response is modeled using a polynomial series truncated to second order. In our approach, STC is used to estimate a subspace that contains the stimulus attributes that govern the

neural response, but the actual response need not be polynomial. Specifically, in chapter 4 we assume that a divisive gain control model governs the response of the neuron.

1.3.1 Open Questions in Spike-triggered (White Noise) Characterizations

White noise analysis has been fruitful in characterizing response properties of sensory neurons. But these techniques have not been used to characterize the type of nonlinear gain control behaviors observed physiologically.

1.4 The Current Work

In the first part of the thesis (chapters 2 and 3), we suggest that sensory nonlinearities arise because sensory systems are designed to efficiently represent environmental information. We describe a form of nonlinear decomposition (specifically, divisive gain control) that is well-suited for efficient encoding of natural signals. We show that this decomposition, with parameters optimized for the statistics of a generic ensemble of natural images can account for some nonlinear response properties of "typical" neurons in area V1 both inside and outside the classical receptive field. We also show few examples in which parameters optimized for an ensemble of natural sounds can account for nonlinearities observed in auditory nerve fibers.

In the second part of the thesis (chapter 4), we develop an experimental methodology for characterizing this class of nonlinear sensory models. The characterization is based on a white noise analysis, in which a set of random stimuli are presented

to a neuron and the spike-triggered ensemble (specifically, the spike-triggered covariance) is analyzed. We demonstrate the technique in simulation and on retinal ganglion cell data. We believe this methodology will also offer more general applicability for exploring nonlinearities in other neural regions and modalities.

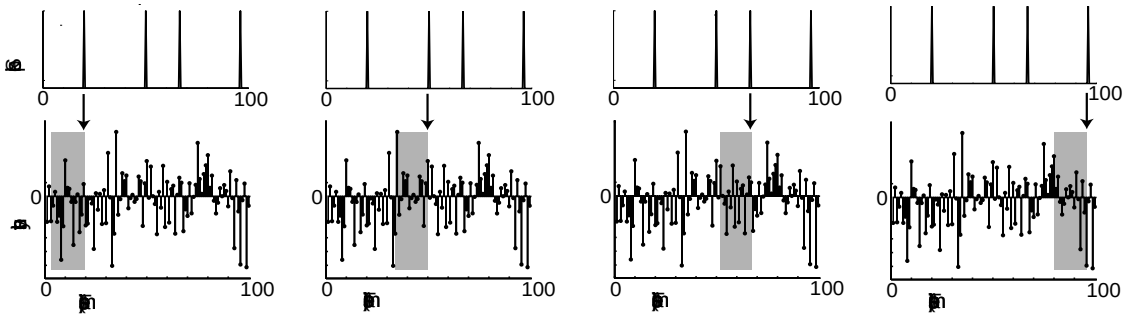

Figure 1.4. Illustration of spike-triggered analysis. Example input stimuli are temporal white noise. The spike-triggered matrix is constructed, by taking for each row the sequence of stimuli in a certain time window that preceded a spike.

CHAPTER 2

NATURAL SIGNAL STATISTICS

What kind of representations do sensory neurons form and why? In chapters 2 and 3, we suggest that a specific form of neural representation arises because sensory systems are designed to efficiently represent environmental information. In this chapter we describe a form of statistical regularity that is prevalent in typical natural signals (both images and sounds). The statistical observations set up the basis for a phenomenological neural model. The model is then tested against physiological data in Chapter 3.

Understanding the statistical properties of natural signals is of multi-disciplinary interest. From the engineering perspective, this can lead to more effective methods for compression, noise removal, and enhancement, of digital images and sounds. From the point of view of Computational Neuroscience, understanding the statistics of natural signals may provide a window into understanding the sensory brain. A longstanding hypothesis states that sensory systems are matched to the statistical properties of the signals to which they are exposed (e.g., [11, 16]). In particular, Barlow has proposed that the role of early sensory systems is to remove redundancy

in the sensory input, by generating a set of neural responses that are statistically independent.

Here we explore the statistical properties of images and sounds pre-processed with a bank of linear filters. The responses of such filters to typical natural signals exhibit a striking form of statistical dependency, in which the width of distribution of one filter grows with the response amplitude of another filter. We demonstrate this dependency empirically through standard filters that have been commonly used to model sensory receptive fields. We then show that this dependency is also present when using a set of linear filters that are optimized for statistical independence in the linear domain (i.e., ICA). We demonstrate empirically that the dependencies may be substantially reduced using an operation known as divisive normalization, in which the response of each filter is divided by a weighted sum of the squared responses of other filters.

2.1 Methods

We choose a set of representative linear filters that are qualitatively similar to neural filters derived physiologically.

The visual filters are computed using the steerable pyramid multi-scale oriented decomposition [142]. Multi-scale linear transforms have become popular for image representation. The steerable pyramid is a multi-scale representation that also includes representation of orientation. It is both rotation and translation invariant. The steerable pyramid is constructed by recursively splitting an image into subbands with directional derivative operators (4 orientations chosen here). Since the

basis functions are localized in spatial position, orientation, and spatial frequency (scale), they qualitatively resemble simple cell receptive fields in area V1. Figure 2.1 (top) illustrates example visual filters. The filters are at the second recursion level of the pyramid; they extend roughly 9 pixels along the aligned axis and 13 pixels along the non-aligned axis.

The auditory filters are gammatone, which have been commonly used to model the peripheral auditory system [73]. They are derived from Slaney's Matlab Auditory toolbox [146]. A gammatone filter is the product of a rising polynomial, a decaying exponential function, and a cosine wave. Figure 2.1 (bottom) illustrates example auditory filters in the frequency domain. Also shown is the time domain representation of one of the filters with center frequency of 2000Hz.

Natural images are obtained from a database of standard images used in image compression benchmarks (known as: boats, goldhill, einstein, baboon, etc); and from the Van Hateren database of calibrated images of outdoor scenes [156]. The images from the standard database are squared to achieve a rough calibration; typical digital images are corrected for a power-law monitor nonlinearity with an exponent of approximately 2. Most of the standard images are 512 by 512 pixels, and the Van Hateren images are cropped to 1024 by 1024 pixels. Natural sounds (speech and animal sounds) are obtained from commercial compact disks and converted to sampling frequency of 22050 Hz. The sounds are of variable length, between 6 and 15 seconds.

In the chapter we go through example image and sound statistics. Unless otherwise noted, image data are collected over the Einstein image and two Van Hateren outdoor tree scenes (altogether 2,162,688 pixels), and sound data combines two

speech sequences (altogether approximately 30 seconds, or 646,580 samples). We also show additional examples of statistics for single images and sounds.

2.2 Natural Signal Statistics Through One Linear Filter

We begin by asking what others have asked: what do responses to typical natural stimuli look like through just one linear filter? What is their distribution? This has been documented in both vision [51] and in audition [9]. We observe responses of a natural sound and natural image through their respective linear filters (figure 2.2). The response of the filter (L) to the image is computed by taking the inner product of the image with the filter at each possible location (i.e., a convolution). We thus assume spatial stationarity. The instantaneous response of the auditory filter is similarly computed by taking the inner product of the sound with the filter at each possible time.

Note in the filter response to the image that there are many gray regions corresponding to zero response, interspersed with regions of high positive (or negative) activity. In particular, the distribution of responses of the image through the linear filter is superGaussian (e.g., peaky with heavy tails) as demonstrated in figure 2.3. This reflects the fact that the filter is responding to some prominent "features" in the image, but that in most locations it does not respond very much at all. In natural sounds, we obtain an analogous trend, and the distribution of responses are also superGaussian. This has been demonstrated for numerous other typical images and sounds.

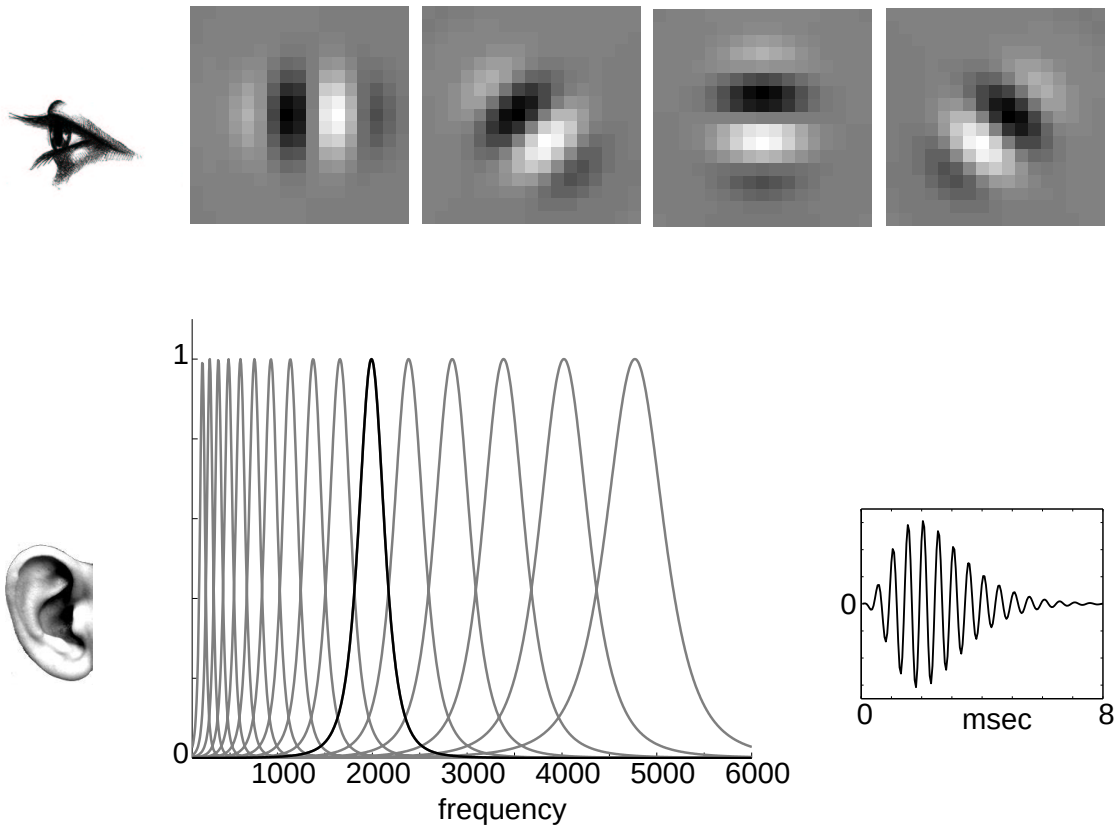

Figure 2.1. Representative linear visual and auditory filters. Top, Representative linear filters for modeling simple cell receptive fields in area V1. Filters are computed using a steerable pyramid [142] (only one scale plotted here). Bottom, Representative gammatone filters for modeling peripheral auditory processing in the frequency domain [146, 73]. The width of the filters is roughly equal on a logarithmic scale. We've also plotted the time representation of one of the filters.

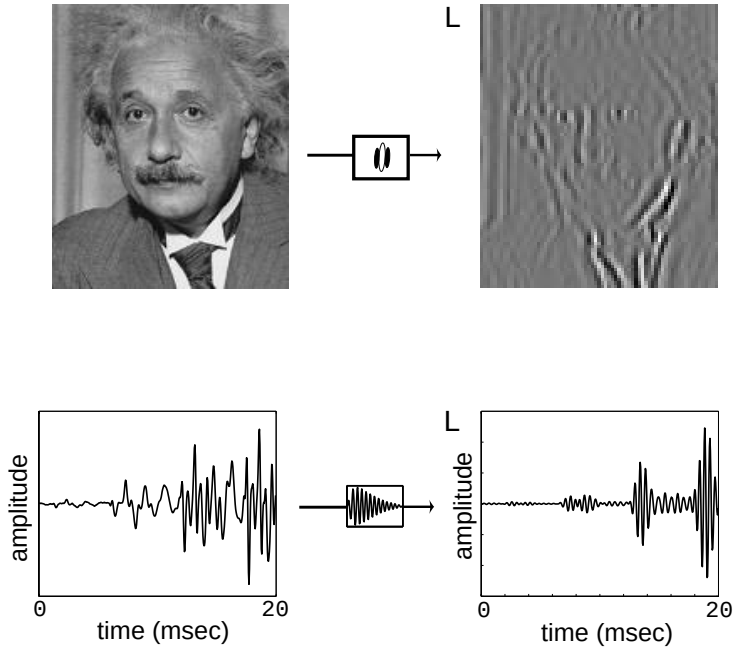

Figure 2.2. Responses of a linear filter to example image and sound. Top, Responses of a linear filter to example image. The response of the filter (L) to the image is computed by taking the inner product of the image with the filter at each possible location. We thus assume spatial stationarity. Bottom, Responses of a linear filter to example speech with sample frequency 22050 Hz (only 20 milliseconds of sound shown here). The instantaneous response of the filter (L) is computed by taking the inner product of the sound with the filter at each possible time.

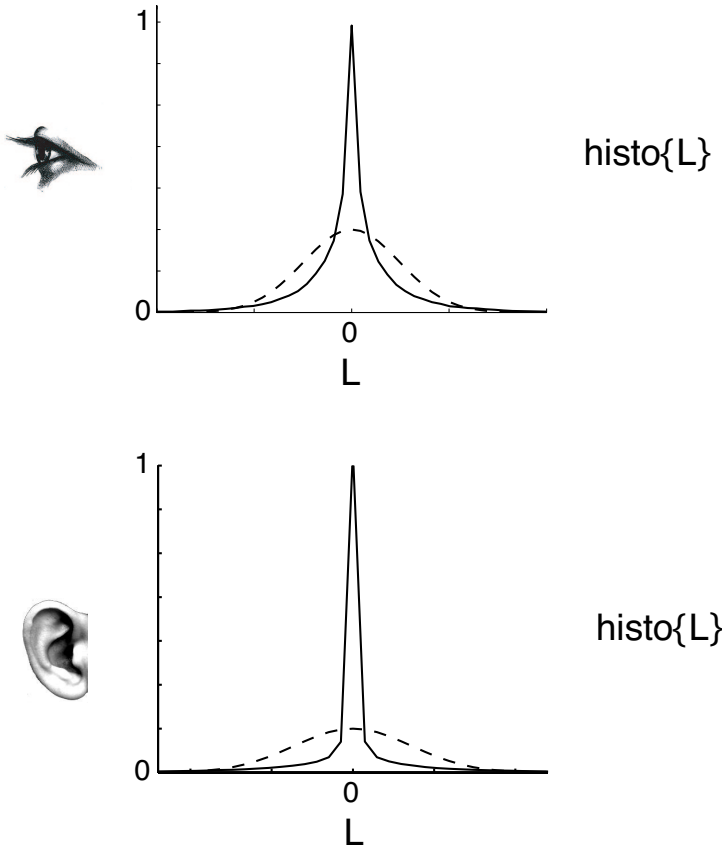

Figure 2.3. Distribution of responses of linear filter to example image and sound. Top, Solid line indicates distribution of the response of a linear filter to an image. Dashed line indicates Gaussian with same variance. Bottom, Solid line indicates the distribution of the response of a linear filter to a speech sound. Dashed line indicates Gaussian with same variance.

2.3 Statistical Dependency Through a Pair of Linear Filters

Following the independence hypothesis, a natural question to ask is whether responses of typical natural signals through a set of linear sensory filters are statistically independent. Figure 2.4 (top) shows responses to an image of a pair of oriented visual filters. The second filter is oriented 45 degrees away, and shifted up by 4 pixels relative to the first filter. We now compute the response of each of the filters to the image; we denote the responses through each filter L_1 and L_2 respectively. This can be thought of as dragging the image across both the filters (such that they are displaced by 4 pixels) and computing each respective response. This means that corresponding point in L_1 and L_2 are computed with the filters displaced by 4 pixels upon the image. Note that both filters are likely to respond simultaneously to prominent features in the image, such as around the eyes. The arrow indicates an example location corresponding to a high contrast edge, in which both filters are responding strongly (the first positive, and the second negative). These features are interspersed with regions in which both filters respond with very low activity.

Similarly, figure 2.4 (bottom) shows instantaneous responses of a pair of bandpass filters typical of the peripheral auditory system. The filters have a temporal frequency of 2000Hz and 2840Hz respectively, and their onset is at the same point in time. As in the visual case, we observe the co-occurrence of large amplitude features separated by low amplitude intervals.

These examples suggest that responses of typical sensory filters to natural signals are not statistically independent. We examine this dependency more explicitly by

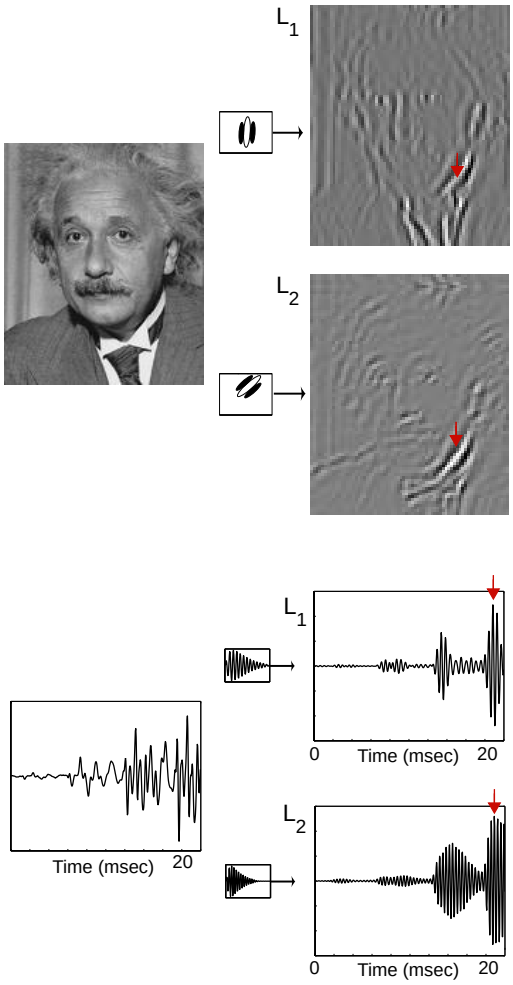

Figure 2.4. Responses of pairs of linear filters to example image and sound. Top, A natural image convolved with two filters selective for the same spatial frequency, but different orientation and spatial position; the lower filter is oriented 45 degrees away, and shifted up by 4 pixels. At a given location, when the first filter responds weakly (gray areas) the second filter will also tend to respond weakly. But when the first filter responds strongly (black or white), the second filter is more likely to respond strongly. The arrows indicate a location corresponding to a high contrast edge, in which both filters are responding strongly (the first positive, and the second negative). Bottom, A natural sound convolved with two filters tuned for different temporal frequencies (2000 and 2840 Hz center frequencies). The arrows indicate a time at which both filters are responding strongly. Note also that when the first filter responds weakly, the second also tends to respond weakly.

constructing a conditional histogram (figure 2.5). As before, we examine a natural signal through two linear filters with responses L_1 and L_2 respectively. If L_1 and L_2 are statistically independent, then knowing L_2 should provide no information about the distribution of L_1. Thus we can ask, when responses L_2 are very low (a bin centered at 0.15) what does the distribution of L_1 look like? We obtain a one dimensional conditional histogram, which is quite narrow. But when responses L_2 are quite high (a bin centered at approximately 0.85) what does the distribution of L_1 look like? Now we obtain a different one dimensional conditional histogram, which is quite wide. The conditional histograms are different, indicating that L_1 and L_2 are not statistically independent.

We can put together a more detailed image of the dependency by plotting a two dimensional conditional histogram, where the one dimensional conditional histograms represent vertical slices. The dependency has an interesting shape, it looks somewhat like a bowtie. In particular, the width of distribution of responses of one filter increases for larger response values (both positive and negative) of the other filter. Note that though responses are not independent, they are nearly decorrelated in the above examples: the expected value of L_2 given the response of L_1 is approximately zero. Thus correlation should be distinguished from the type of dependency we are describing, which depends on the width of distribution.

This form of dependency appears to be ubiquitous - it is evident in a wide variety of typical natural images and sounds (figure 2.6). The strength of the dependency may vary for individual signals. This dependency is a property of natural signals, and is *not* due purely to properties of the particular set of linear filters we chose. For example, no such dependency is observed when the input consists of white noise.

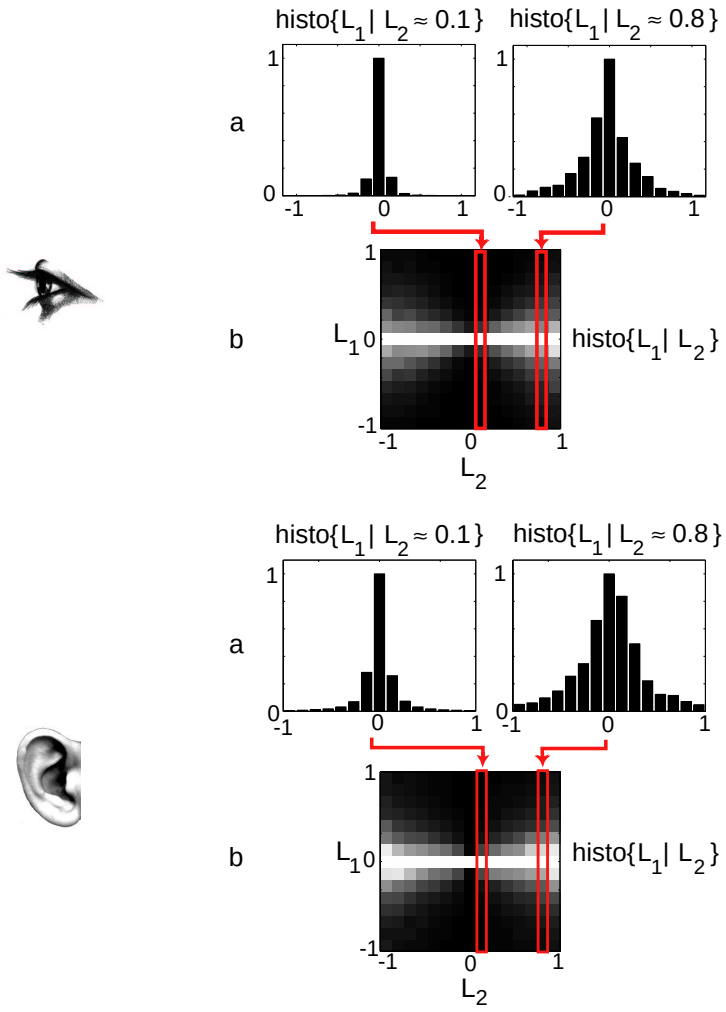

Figure 2.5. Joint statistics of typical natural images and sounds as seen through two linear filters. Top, **a**, One-dimensional histograms of the response of a vertical filter (L_2), conditioned on two different values of the response of a diagonal spatially shifted filter (L_1). Pairs of responses are gathered over all image positions. Differing widths of these histograms clearly indicate that filter responses are not statistically independent. **b**, Full two-dimensional conditional histogram. Pixel intensity is proportional to the bin counts, except that each column is independently re-scaled to fill the range of intensities. Note that L_1 and L_2 are roughly decorrelated but not statistically independent. Bottom, Same for speech and two auditory filters with center frequencies 2000 Hz and 2840 Hz respectively. Axes of the joint conditional histograms are normalized between -1 and 1.

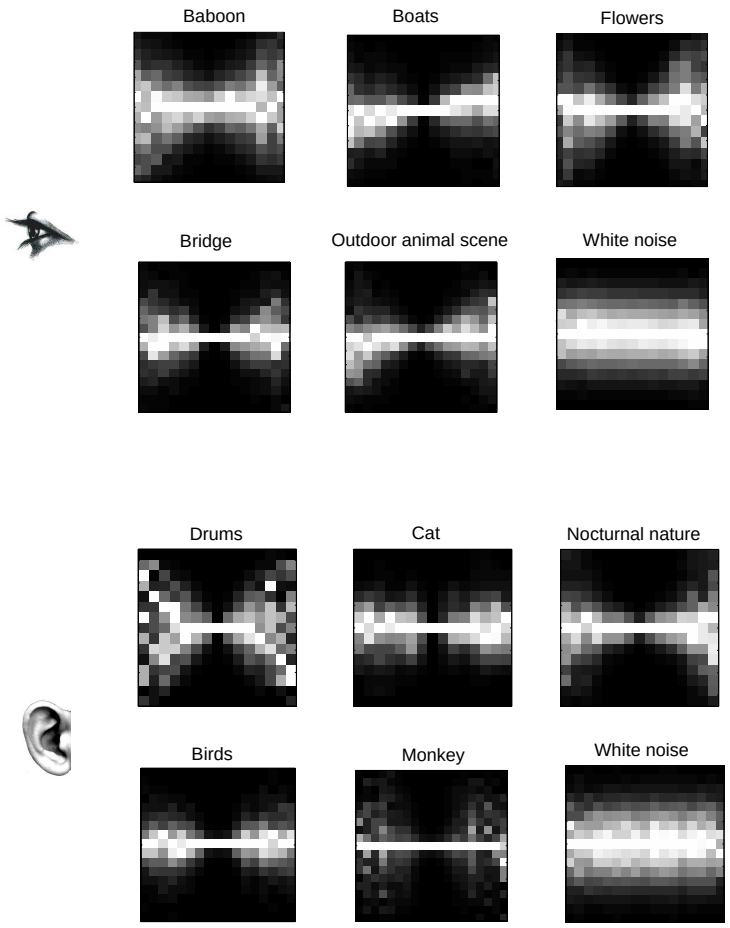

Figure 2.6. Joint statistics of different images and sounds through two linear filters. Although the strength of the dependency may vary, many natural signals exhibit a bowtie dependency, but white noise is roughly independent. Top, Two dimensional joint conditional histogram for different images and white noise. The images are (left to right, and top to bottom): baboon, boats, flowers, and bridge from the standard image database; Van Hateren outdoor scene of grass with animals; and white noise. The ordinate response is always computed with a vertical filter, and the abscissa response is computed with a diagonal filter (shifted 4 pixels). Bottom, Two dimensional joint conditional histogram for different sounds and white noise. The sounds are (left to right, and top to bottom): drums; cat; nocturnal nature; birds; monkey; and white noise. The ordinate response is always computed with a filter with center frequency 2000 Hz, and the abscissa response is computed with a filter with center frequency 2840 Hz.

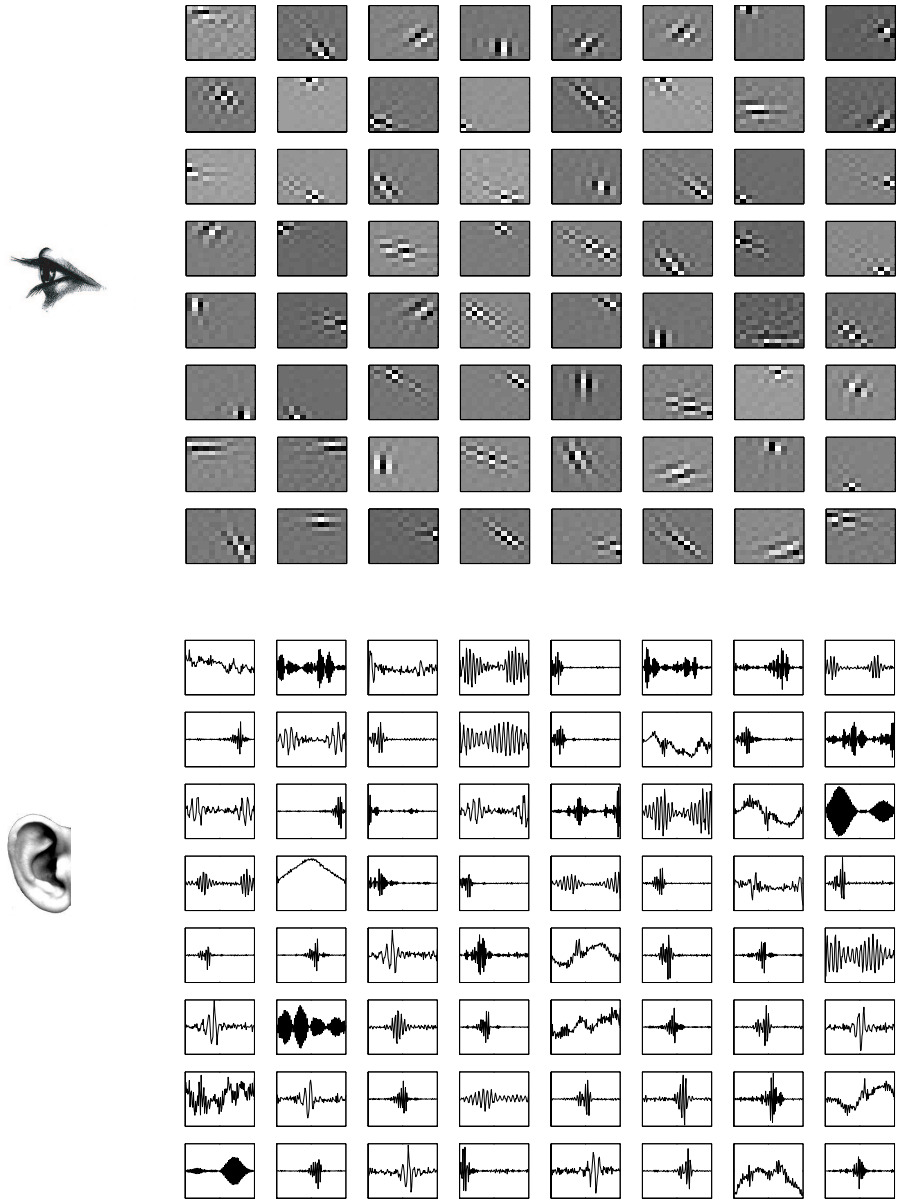

Figure 2.7. Set of filters obtained from Bell and Sejnowski's ICA procedure [20, 21, 22]. Top, ICA analysis on image with 12 by 12 patches. This yields 144 basis functions (only 64 shown here). The starting positions of the patches (20,000 altogether) were randomly selected. Bottom, ICA analysis on speech with 100-sample segments. This yields 100 basis functions (only 64 shown here). The starting times of the segments (20,000 altogether) were randomly selected. Speech signals are at 8KHz sampling frequency, to allow for a 12.5 millisecond time coverage.

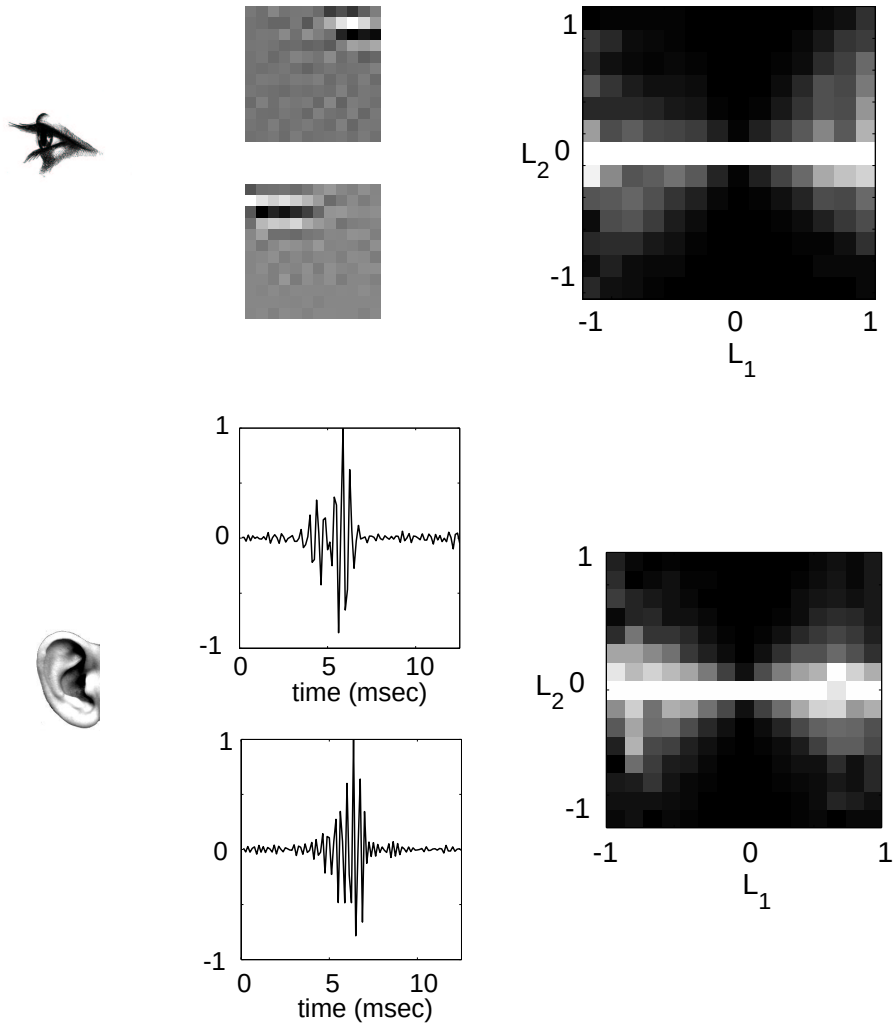

Figure 2.8. Dependency for pair of filters retrieved from ICA analysis. The joint conditional histograms of the filter responses follows the familiar bowtie shape. Top, A pair of linear filters retrieved from ICA analysis on image. ICA analysis was performed on 12 by 12 image patches. The starting positions of the patches (20,000 altogether) were randomly selected. Bottom, A pair of linear filters retrieved from ICA analysis on a speech signal. ICA analysis was performed on 100-sample segments of the signal. The starting times of the segments (20,000 altogether) were randomly selected. Speech signals are at 8KHz sampling frequency, to allow for a 12.5 millisecond time coverage.

There is still a concern that the observed dependencies might be a result of the particular linear filters we chose. Here we verify that this form of dependency occurs even when the filters are chosen from a set that is optimized for statistical independence in the linear domain. We compute an independent components analysis (ICA) on the example images and sounds, using a procedure developed by Bell and Sejnowski [20, 22]. Figure 2.7 depicts the linear receptive fields obtained from the images and sounds respectively. Consistent with what is reported in the literature, image basis functions are localized in orientation, spatial position, and spatial frequency [22, 103]; and speech basis functions are localized in phase and temporal frequency, although some are less localized and several have a bimodal shape (see [82] for a more detailed discussion on ICA for classes of natural sounds). Figure 2.8 shows the conditional histogram constructed from the responses of two ICA filters to the same signals: responses are not statistically independent, and a similar bowtie dependency is observed.

2.4 Modeling the Statistical Dependency

The strength of the dependency varies according to the specific pair of filters chosen (figure 2.9). For the visual filters, the strength of the dependency can vary as a function of relative orientation, spatial position and spatial frequency of the filter pair. For the auditory filters, the dependency is a function of the relative temporal frequency, as well as the relative timing of the two filters.

Figure 2.10 illustrates additional examples of dependencies for different filter pairs and different signals. For visual filters, a stronger dependency is often ob-

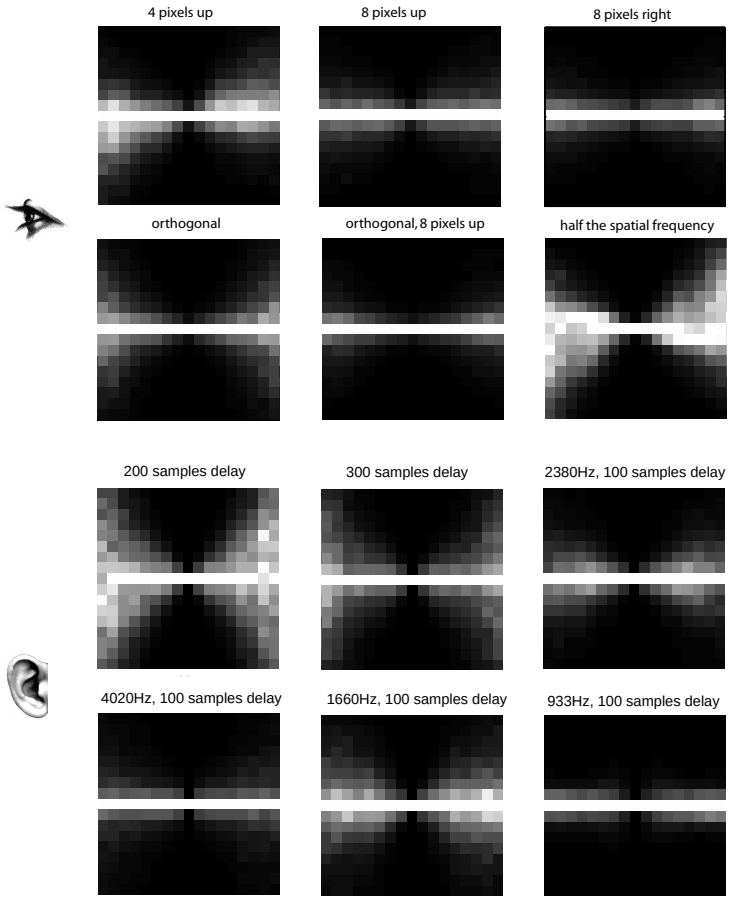

Figure 2.9. Image and sound statistics for different filter pairs. Top, two dimensional conditional histograms for different pairs of visual filters. The ordinate response is always computed with a vertical filter, and the abscissa response is computed with an aligned iso-orientation filter (4 pixels up); aligned iso-orientation filter (8 pixels up); non-aligned iso-orientation filter (4 pixels right); orthogonal filter (same position), orthogonal filter (8 pixels up); and iso-orientation filter at half the spatial frequency (same position). Bottom, two dimensional conditional histograms for different pairs of auditory filters. Temporal frequency of ordinate filter is 2000 Hz. Temporal frequency of abscissa filter are 2000 Hz (shifted 9 msec in time); 2000 Hz (shifted 13.5 msec in time), 2380 Hz (shifted 4.5 msec in time); 4020 Hz (shifted 4.5 msec in time); 1660 Hz (shifted 4.5 msec in time); and 933 Hz (shifted 4.5 msec in time);

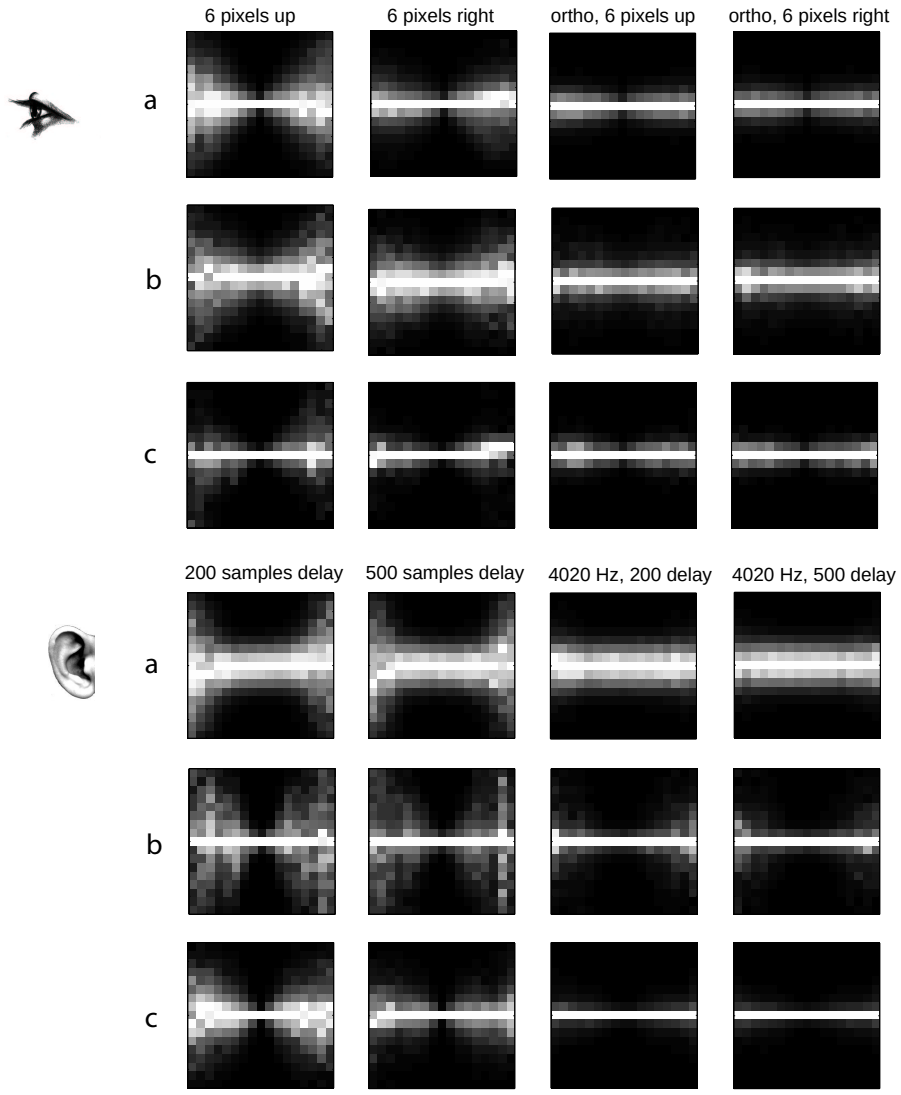

Figure 2.10. Image and sound statistics for different signals and filter pairs. Top, Ordinate response is always computed with a vertical filter, and the abscissa response is computed with an aligned iso-orientation filter (6 pixels up); non-aligned iso-orientation filter (6 pixels right); orthogonal filter (6 pixels up), orthogonal filter (6 pixels right). **a,** Van Hateren outdoor scene of straw. **b,** Baboon image. **c,** Van Hateren outdoor scene of grass and animals. Bottom, Temporal frequency of ordinate filter is 2000 Hz, and of abscissa filter is 2000 Hz (shifted 9 msec); 2000 Hz (shifted 22.5 msec), 4020 Hz (shifted 9 msec); 4020 Hz (shifted 22.5 msec). **a,** Jungle sounds. **b,** Chimps. **c,** Puppies.

served for a parallel aligned filter than for a parallel non-aligned filter. Additionally, stronger dependency is typically observed for parallel filters than for orthogonal filters. For auditory filters, a very striking dependency is typically seen when the response of a given filter is time-shifted. The dependency is often reduced for filters that are far away in temporal frequency.

More generally, the dependency appears to decrease for pairs that differ markedly in one or more attributes. However, the dependency may be present even if the two filters are non overlapping. In some cases, when there is no dependency, the joint conditional histogram becomes horizontally flat (obtaining a similar shape to the white noise examples in figure 2.6). There are also cases in which for low response values, a rather flat histogram is observed (signifying relative independence), but for higher response values (both positive and negative) the bowtie shape appears (signifying dependence). These cases can be observed, for example, in the jungle sounds in figure 2.10.

We would like to formalize a model of this dependency. Neurons undergo rectification, so rather than looking at the width of distribution dependency between the linear responses of two filters, we can look at standard deviation (for absolute value of responses), variance (for squared value of responses), and so on. The specific choice of power will not alter the statistical observations very much (see discussion). We choose to capture this fundamental dependency using variance. The choice of variance is rather ad-hoc, and based on the goal we have in mind of modeling neural data; responses of neurons in area V1 and auditory nerve typically undergo a point nonlinearity that is close to squaring [65, 124].

We formalize the conditional relationship between a given pair of linear filter

responses $\{L_1, L_2\}$ with a model in which the variance of L_2 is proportional to the squared value of L_1 plus an additive constant (e.g., [28]). For a pair of filters with strongly dependent responses, this proportion is larger; for a pair whose responses are independent, this proportion is zero and variance reduces to an additive constant. Specifically, the variance of L_1 given L_2 is computed as:

$$var(L_1|L_2) = wL_2^2 + \sigma^2 \tag{2.1}$$

where L_1 and L_2 are the linear responses of the two filters; the connection weight between the two filters is given by w; and σ^2 is an additive constant.

But we expect L_1 to depend not only on L_2, but also on other filter responses within a neighborhood. Remember the dependencies for different pairs of filters illustrated in figure 2.9. A visual filter response might be dependent on other visual filter responses at different orientations, spatial positions, and spatial frequencies. Similarly, an auditory filter response might depend on other filter responses at different temporal frequencies, and on the responses of those filters (including its own response) over time. Incorporating a neighborhood of filters is more realistic for modeling a neural system, and computationally we expect it to provide a more powerful estimate of the variance.

Thus we form a generalization of the conditional variance model. The response of one filter (L_i) is proportional to a weighted sum of the squared responses over the neighborhood and an additive constant. We model the variance dependency of the response of filter L_i given the responses of a population of filters L_j in a neighborhood N_i:

$$var(L_i|\{L_j, j \in N_i\}) = \sum_j w_{ji} L_j^2 + \sigma^2 \tag{2.2}$$

The connection weight between the chosen filter i and filter j is given by w_{ji}. The constant $\hat{\sigma}$ can be thought of as the residual variance not predicted from the neighbors. In choosing a particular neighborhood, we make a Markov assumption that the conditional probability density of a filter response given the neighborhood is independent of everything outside the neighborhood. In practice, we choose the neighborhoods to span a reasonable range of relevant parameters, such as orientation, frequency, spatial position, and so on. In a real system, $\hat{\sigma}$ could also result from internal or external noise, which we do not account for here. Also note that the value of the $\hat{\sigma}$ depends on the somewhat arbitrary scaling of the input signal. That is, doubling the input strength would lead to a doubling of $\hat{\sigma}$.

We will later go back and check empirically that the model provides a good description of the dependency. We first explain how to optimize the parameters, and achieve more independent responses.

2.5 Reducing the Statistical Dependency: Divisive normalization

If this model describes the dependency between filter responses, how can these responses be made more independent? First, we note that this type of dependency cannot be eliminated by performing another linear operation. A linear operation can only rotate or scale the axes. Given that the dependency governs the variance, the natural solution is to *divide* the squared response of each filter by its variance (as predicted from the linear combination of its neighbors), to obtain something

more univariate:

$$R_i^2 = \frac{L_i^2}{\sum_j w_{ji} L_j^2 + \sigma^2} \tag{2.3}$$

We denote this operation divisive normalization (following the terminology in Heeger et al. [64]).

We would like to choose the parameters of the model (the weights w_{ji}, and the constant σ) to optimize the independence of the normalized response to an ensemble of natural images and sounds. Such an optimization is computationally prohibitive. In order to reduce the complexity of the problem, we assume an underlying conditional distribution; we assume this form is Gaussian with zero mean:

$$\mathcal{P}\left(L_i \middle| \{L_j, j \in N_i\}\right) = \frac{1}{\sqrt{2\pi\left(\sum_j w_{ji} L_j^2 + \sigma^2\right)}} \exp\left[\frac{-L_i^2}{2\left(\sum_j w_{ji} L_j^2 + \sigma^2\right)}\right] \tag{2.4}$$

We will later also need to go back and check the Gaussian assumption. If the resulting divisively normalized response is close to Gaussian, this will mean that the assumption was reasonable. This reasoning is similar to ICA procedures; there an underlying prior is assumed on the marginal distributions, and following the procedure one can check if in fact the ICA component distributions match the assumption.

We then maximize the product of such distributions over the image/sound data at all positions/times x.

$$\hat{w}_{ji}, \hat{\sigma} = \arg\max_{w_{ji},\sigma} \prod_x \mathcal{P}\left(L_i(x) \middle| \{L_j(x), j \in N_i\}\right) \tag{2.5}$$

We solve for the optimal parameters numerically, using gradient descent. Specifically, we use matlab's fminu function and minimize the negative of equation 2.5 (by taking the derivative of the log of the expression). The weights and σ are initialized to uniform values. The weights are constrained to be positive, to avoid instability in optimization convergence and in simulations of chapter 3. We force the weights positive by squaring them within the optimization function. In chapter 3 we will discuss in detail how the resulting optimal model behaves as a function of particular parameters, such as orientation, frequency, spatial position, and so on. For now we continue to focus on the statistical story.

2.6 Testing the Model Empirically

First, we optimize the weights and the constant σ for just one linear filter, over the natural images or sounds respectively. According to the model, the weighted sum of squared responses of neighboring filters (with the optimal parameters) should be a good predictor of the conditional variance of the response of the filter (equation 2.2). For readability, we continue to plot the histograms in the linear (rather than the squared) domain. Then the square root of the weighted sum of squared responses should be a good predictor of the conditional standard deviation of the response of the filter. Figure 2.11 shows the original filter response L_1, conditioned on the square root of the weighted sum of squared neighboring responses R_1 (figure 2.11). Since R_1 has only positive values (but L_1 is both positive and negative) we observe half a bowtie. Also shown is a line indicating the conditional standard deviation of L_1. If the model is correct, this should be roughly a straight unit slope

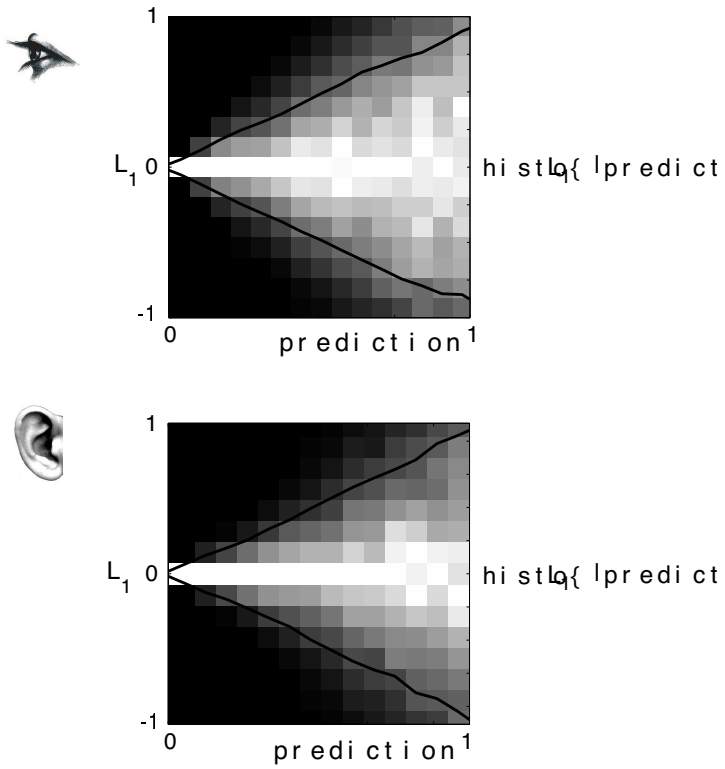

Figure 2.11. Two dimensional joint conditional histogram of filter response versus its model prediction. Top, Two dimensional joint conditional histogram of L_1 versus the square root of the variance prediction (labeled "prediction" in graphs). Variance prediction is given by the the sum of squares of weighted neighboring responses, with the optimal parameters retrieved from the Maximum Likelihood procedure. Line indicates one time the conditional standard deviation, and is roughly proportional to the value of the abscissa. L_1 is computed for a vertical filter. The pool of filters in the optimization procedure consists of 4 orientations and 24 spatial positions (displaced vertically, horizontally, and diagonally) for each orientation, and 2 spatial frequencies. Bottom, Same with L_1 computed for speech through auditory filter with center frequency 2000 Hz. The pool of filters in the optimization procedure consists of 15 temporal frequencies, and 3 time shifts.

line, except perhaps in the low intensity regime, in which σ might dominate. In practice, σ is very small and cannot be distinguished in the plots. The conditional standard deviation is close to a unit slope line, as expected.

Now our goal is to apply this procedure to each of the two filters that exhibited the variance dependency (Figure 2.5), and then demonstrate empirically that the resulting dependency (between the two normalized filter responses) is reduced. That is, for each of the two filters, the variance is estimated from the sum of weighted squared responses of its neighbors. The squared response of the filter is then divided by the variance estimate (equation 2.3). When we actually do this for each of the linear filters in figure 2.5, we observe that responses to the natural signals are more independent. We plot a joint conditional histogram of the two normalized filter responses, R_i. The resulting joint conditional histogram signifies that responses are more independent: vertical cross sections have a similar width of distribution. Also plotted is a line indicating one times the conditional standard deviation; following divisive normalization, the conditional standard deviation is roughly constant. Figure 2.13 shows additional examples of filter responses following divisive normalization. Responses are typically more independent, although there are variations for different filter pairs and signals. Note that this technique largely reduces second order variance dependencies, but does not guarantee to achieve full independence.

We can also go back and look at the marginal distribution of a filter response following divisive normalization. Recall that in our formulation of equation 2.5 we assumed that L_i given a neighborhood of filter responses is Gaussian, with the variance determined by the weighted sum of neighborhood responses. Then dividing

L_i by the square root of the variance estimate should result in a Gaussian marginal distribution. Figure 2.14 demonstrates that the normalized marginal distributions are close to Gaussian. Figure 2.15 shows additional examples of marginal distributions for different signals and filters. The fit is better in the example images than in the example sounds. But in both cases, this should be contrasted with the highly superGaussian distribution of responses through the original linear filters L_i (Figure 2.2).

2.7 Discussion

We have shown that natural signals through typical sensory filters exhibit striking statistical dependencies, in which the width of distribution of the response of one filter grows with the response amplitude of neighboring filters. We have demonstrated empirically that divisive normalization by a weighted linear combination of squared responses of neighboring filters can largely reduce this form of dependency.

The statistics we describe are common to both visual and auditory stimuli. Intuitively, we interpret this to mean that natural signals tend to have strong "features" that may be captured by several linear filters simultaneously. These type of dependencies have been documented in the vision image processing literature. Simoncelli et al. have shown that image processing techniques, such as compression, denoising, and enhancement, can benefit from taking advantage of the variance dependency [141, 28]. Statistical dependencies in vision have also been documented through a quadrature pair of filters [172]. Variations in the choice of filters, neighbors in the optimization, and specific optimization procedure have yielded similar results (see

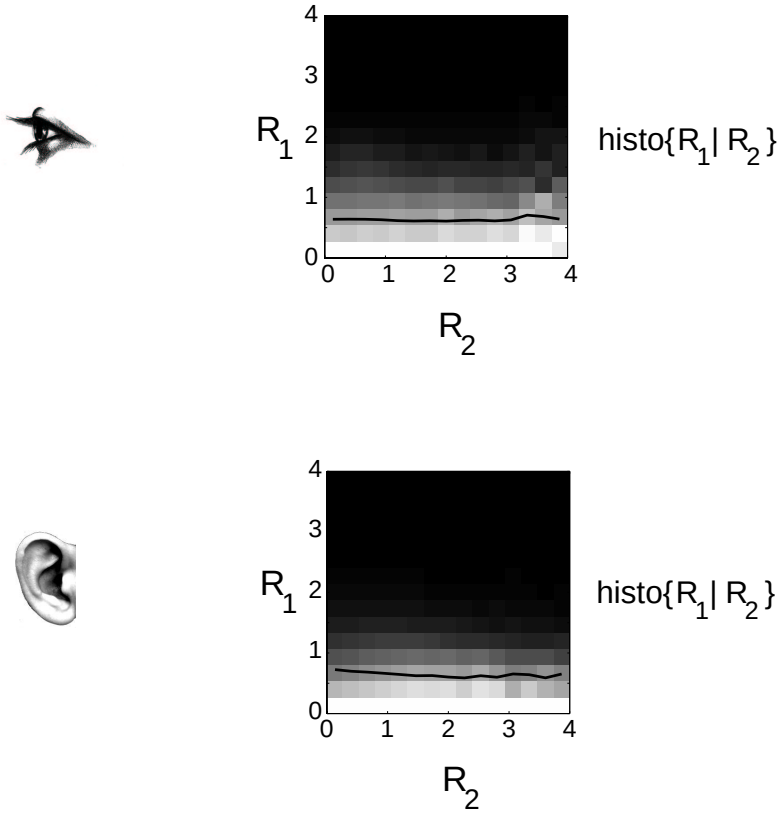

Figure 2.12. Two dimensional joint conditional histogram following divisive normalization. Each filter response (L_1 and L_2 respectively) is squared and divisively normalized by the sum of weighted squared responses of its neighbors (equation 2.3). Shown is the square root of the resulting responses (R_1 and R_2 respectively). Responses are more statistically independent. Line represents one time the conditional standard deviation. Top, R_1 and R_2 computed for a vertical filter, and a diagonal filter (shifted 4 pixels). The pool of filters in the divisive normalization procedure consists of 4 orientations and 24 spatial positions (displaced vertically, horizontally, and diagonally) for each orientation, and 2 spatial frequencies. Bottom, R_1 and R_2 computed for speech through two auditory filters with center frequencies 2000 and 2840 Hz. The pool of filters in the divisive normalization procedure consists of 15 temporal frequencies, and 3 time shifts.

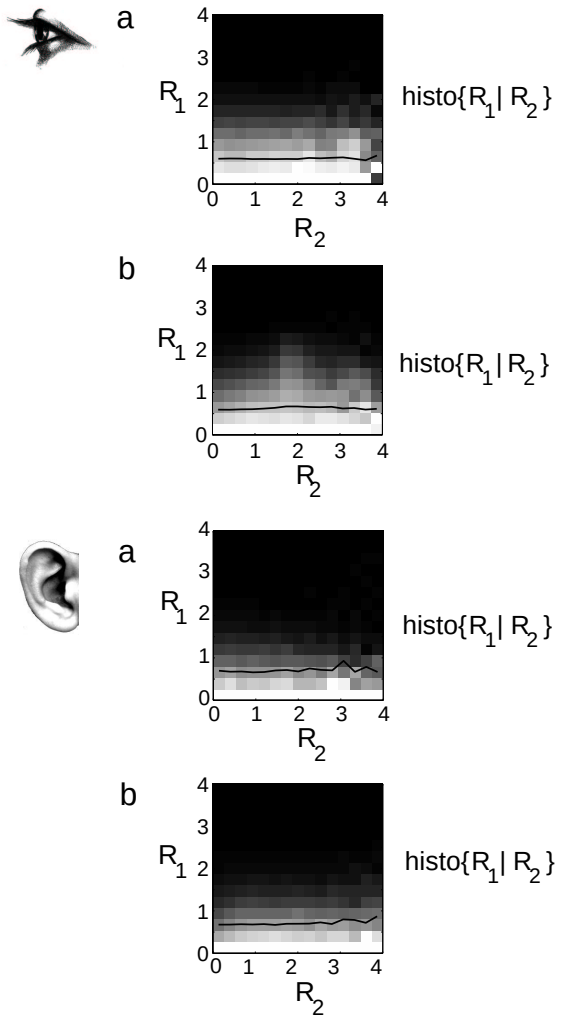

Figure 2.13. Additional examples of two dimensional joint conditional histograms following divisive normalization (as in Figure 2.12) Top, **a,** Van Hateren outdoor scene of grass and animals. Divisively normalized responses R_1 and R_2 computed for a vertical filter, and a diagonal filter (shifted up by 4 pixels). **b,** Regular example images. R_1 and R_2 computed for a vertical filter, and a vertical filter shifted up by 8 pixels. Bottom, **a,** Monkey sound. R_1 and R_2 computed for auditory filters with center frequencies 2000 and 2840 Hz. **b,** Regular speech sounds. R_1 and R_2 computed for auditory filters with center frequencies 2000Hz and 1660 Hz. Line represents one time the conditional standard deviation.

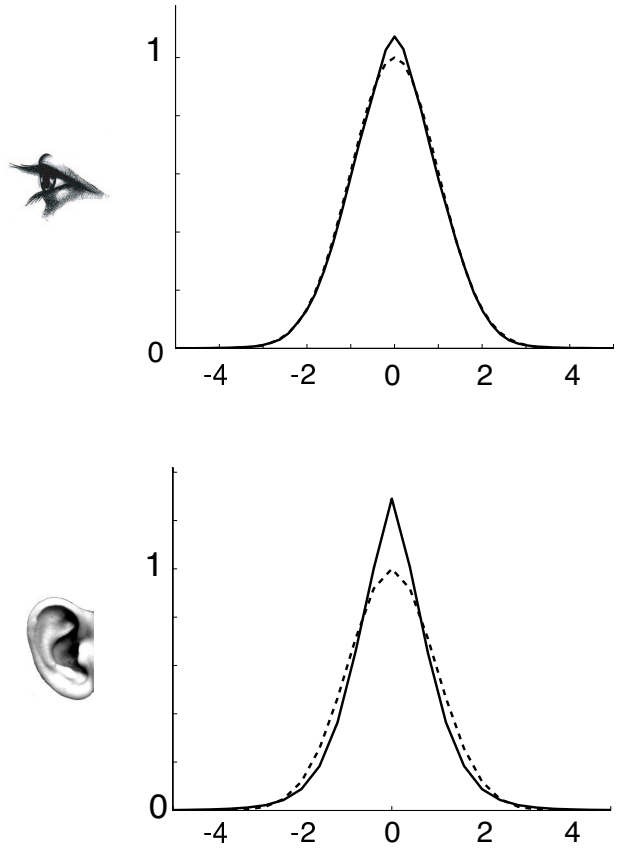

Figure 2.14. One dimensional marginal histograms following divisive normalization. R_1 values are positive due to the rectification and squaring in equation 2.3. But for readability, we compute the histogram over both positive and negative values (that is, L_1 divided by the normalization signal). Top, Normalized signal computed for images through a vertical filter. Dashed line indicates Gaussian with same variance. Bottom, Same with normalized signal for speech through auditory filter with center frequency 2000 Hz.

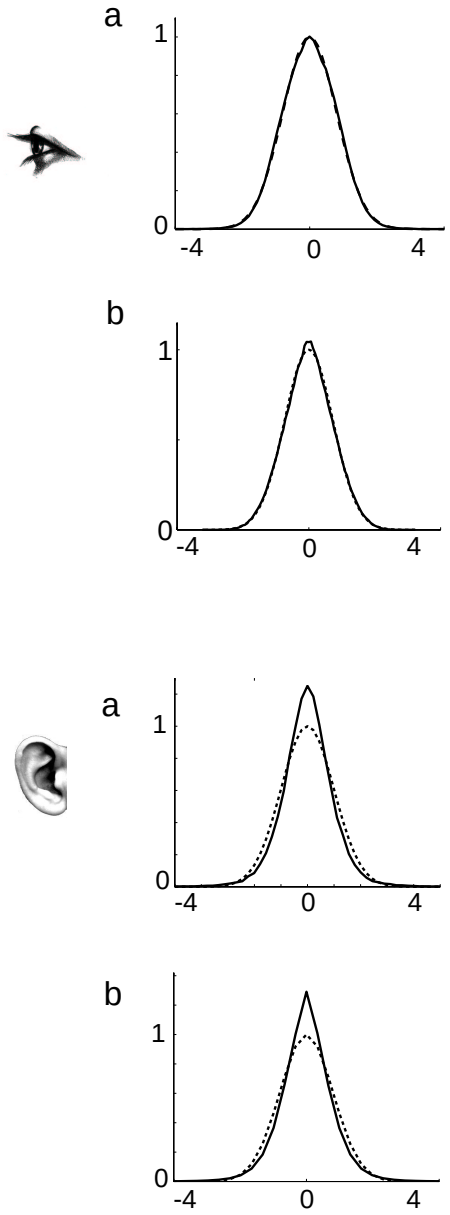

Figure 2.15. Additional examples of marginal histograms following divisive normalization (as in Figure 2.14) Top, **a,** Van Hateren outdoor scene of grass and animals. Normalization signal computed for vertical filter. **b,** Regular example images. Normalization signal computed for diagonal filter at +45 degrees. Bottom, **a,** Monkey sound. Normalization signal computed for filter with center frequency 2000 Hz. **b,** Regular speech sounds. Normalization signal computed for filter with center frequency 1660 Hz.

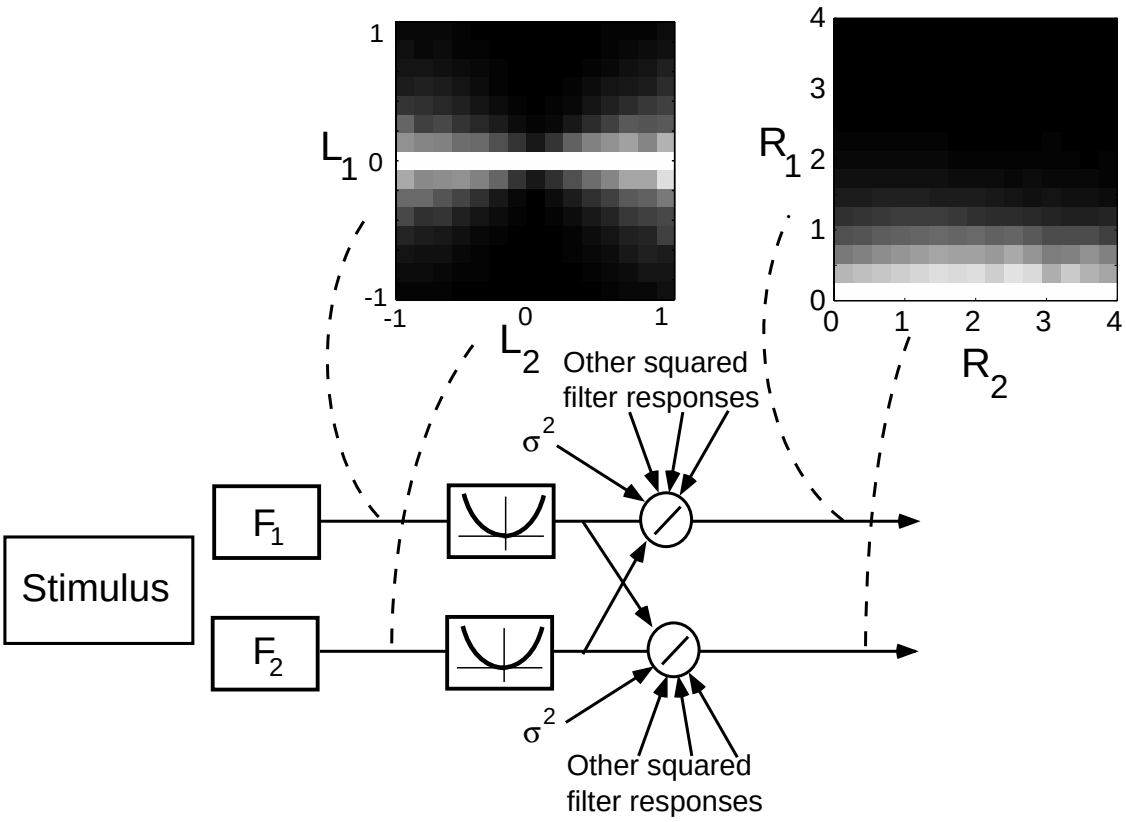

Figure 2.16. Generic normalization model for vision and audition. Each filter response (L_1 and L_2) is squared and divided by the weighted sum of squared responses of neighboring filters and an additive constant (yielding R_1^2 and R_2^2). Parameters are determined using Maximum Likelihood on typical natural signals. The conditional histogram of the original filter responses indicates that they are highly dependent; the conditional histogram of the normalized responses demonstrates that the variance of R_1 is roughly constant, independent of R_2. The diagram is a schematic representation of the computation and is not meant to specify a particular mechanism or implementation.

[28, 144, 162]). The statistical observations for natural sounds are rather preliminary. Para et al. have also described bowtie dependencies in speech [105]. Further analysis is needed to understand possible variations in the statistics across different classes of natural sounds. This line of work could potentially open up possibilities for auditory signal processing applications, and it is our hope that this may also lead to improved aids for hearing and vision. In chapter 3 we try to link between the statistical observations and neuronal processing.

The procedure we describe for reducing statistical dependencies is a step forward from ICA, which is restricted to a linear decomposition. Nevertheless, our technique also does not guarantee to capture all dependencies; we only capture a specific form of variance dependency. Additionally, we have not proven that divisive normalization is in fact optimal. Wainwright et al. have established a more formal class of image model for which division is optimal [164, 165]. Specifically, Gaussian Scale Mixture models are probability densities given by a Gaussian distribution multiplied by a random (hidden) variable. These models produce marginal distributions that are superGaussian, and variance scaling joint conditional distributions (consistent with the statistics presented in the chapter). Responses of filters are linked by the hidden scaling variables corresponding to local image structure. These hidden variables can be estimated, and a divisive normalization procedure can be used to Gaussianize and make more independent the filter responses. In the simplest case (uncorrelated Gaussian), the estimator is a sum of squares.

A limitation of the optimization procedure is that the self filter (or any highly correlated filter response) cannot be included in the neighborhood. If, for example, the self is included, then the variance estimate can be extracted exactly from the

self. The Gaussian Scale Mixture model offers more complicated computational routes that can include the self [111]. Intuitively, these procedures estimate a single hidden variable for a local neighborhood, and thus no one filter in the neighborhood can dominate the estimation. Here, in contrast, the estimate is done separately (and is different) for each filter. These recent computational routes offer potential opportunities to improve our statistical model. However, these issues are still under investigation in the engineering community, and the underlying computations are not a simple sum of squares.

As a consequence, the neighborhood filters in our optimization procedure cannot be sampled too finely along the parameters of interest. There is a tough tradeoff in our model between smoothly covering the parameters of interest, and choosing a neighborhood that is roughly uncorrelated. For example, the steerable pyramid guarantees smooth coverage of orientation, but adjacent orientation bands are not uncorrelated. We minimize the correlations by choosing only 4 orientation bands. We have previously used an orthonormal pyramid, which does not have such correlations; the model has still yielded similar results, although the diagonal orientation filters are mixed and do not provide a good model of simple cells [144]. In the gammatone filters, we choose frequency bands that are close to uncorrelated, but still reasonably cover the frequency space. Sampling in space and time is also chosen such that there is minimal correlation; this means that we do not get as smooth a coverage as one would like. An additional route would be to start off with a set of filters computed from ICA analysis. The disadvantage is that such filters tend to be blocky; they are usually computed over a fixed size of spatial patch or fixed time window (as in figure 2.7). It is also more difficult to inspect their behavior

as a function of different parameters. For our purposes here, and in chapter 3, we have chosen to work with a smoother set of filters (steerable and gammatone).

As noted earlier, we have assumed a variance estimation procedure, but more generally the conditional relationship between a filter and its neighbors need not be a power of two. Specifically, for any p, one can estimate $\sum_j w_{ji} L_j^p + \sigma^p$. We actually do not know which choice of p is a better statistical description of the dependency, from the observed pairwise conditional distributions. In particular, for a pair of filters L_1 and L_2, the expression $w_{21} L_2^p + \sigma^p$ would appear rather similar for different p, except perhaps at the low intensity regions (governed by σ). This issue could perhaps be resolved, by examining dependencies between more filters (beyond pairwise) empirically. Here we have chosen a power of 2 for convenience of modeling neural data.

Finally, we aim to link more closely between the signal statistics presented here and neuronal processing. The statistical properties described in the chapter suggest a generic phenomenological model, analogous for vision and audition (Figure 2.16). A natural signal is passed through a bank of linear filters (only 2 depicted for readability). In the divisive normalization stage, the response of each filter is squared and divided by a weighted combination of squared responses of other filters in the population plus an additive constant. The resulting responses are significantly more independent. The model is phenomenological, i.e., the simplest conceptualization of what the neurons might actually be computing.

We assume that the model parameters are determined over the course of evolution and development, when the system is exposed to many signals in the natural world. Note that the responses of the normalized filters to any arbitrary signal

need not be statistically independent. Responses are presumably independent to "typical" natural signals.

It is interesting to note that this particular model for achieving more independence might be at odds with an alternative hypothesis–that a goal of neural processing should be to achieve sparseness [51, 52, 171, 159]. As stated in the introduction, independence is a different motivating principle than sparseness. The notion of sparseness that people often have in mind is that for a given natural stimulus, only few neurons are firing. This is difficult to verify experimentally, since typically one is recording only from a single neuron, or from several neurons.

A more concrete measure for examining sparseness is based on the marginal distributions of neural responses to natural stimuli (but note that this measure does not guarantee the more intuitive notion above). Based on examination of marginal distributions, the model we propose is not sparse. The original marginal distributions are superGaussian through typical filters, whereas the marginal distributions following divisive normalization are Gaussian. In theory, each neural response could now be passed through an additional point nonlinearity to achieve a desired marginal distribution. If responses of two neurons are statistically independent, then this property will be retained following a point nonlinearity. Adding such a stage to our model would require both a good understanding of what marginal statistical criteria to use (for example, based on the constraints through the individual channels) and what marginal distributions the real neurons have when exposed to natural stimuli. Since both these issues are unresolved at the moment, we examine the model in its current simple form.

From now on, we'll consider just a single filter that interacts with a neighborhood

of filters. The single-filter scenario is analogous to a physiologist recording from a single neuron. Though we assume that all filters in the neighborhood undergo a similar divisive normalization procedure, we examine the behavior of only one filter. This is a simplification, that ignores issues of implementation and dynamics. In the next chapter we set up specific versions of the model to compare to physiology.

CHAPTER 3

NATURAL SIGNAL STATISTICS MODEL VS. PHYSIOLOGY

In this chapter, we continue to address the question of efficient coding and sensory neural representations. In chapter two we derived a phenomenological model of neural processing from the statistical properties of natural signals. Figure 3.1 depicts the resulting conceptualization of a model neuron. The response of the model neuron to an input stimulus is given by the squared response of a filter, divided by the weighted sum of squared responses of neighboring filters and an additive constant. The weights in the model are determined from the statistical properties of an ensemble of natural signals.

We assume that the model parameters are determined over the course of evolution and development, when the system is exposed to signals in the natural world. Given these parameters, the model represents an abstraction of a neuron. The response of the model neuron can be computed for any stimulus; it is simply given by the squared linear response, divided by the weighted sum of squared responses of neighboring filters plus an additive constant.

The model developed from signal statistics incorporates a form of automatic gain control known as "divisive normalization" that has been used to account for some nonlinear steady-state behaviors of neurons in area V1 (e.g., [64, 58, 30]). Divisive normalization models have been motivated by several basic properties. First, gain control allows a system with limited response range to handle a wider dynamic range of input. Divisive normalization achieves this goal, producing sigmoidal contrast-response functions similar to those seen in neurons. In addition, it seems advantageous for tuning curves in stimulus parameters such as orientation to retain their shape at different contrasts, even in the presence of response saturation [64]. Previous models have accomplished this by computing a normalization signal that is independent of parameters such as orientation (achieved with a uniformly weighted sum over the entire neural population). A consequence of this design is that the models can account for the response suppression that occurs, for example, when a grating of non-optimal orientation is superimposed on a stimulus.

Here we set up specific versions of the normalization model to compare to physiological data. We focus on vision (area V1), and also present few examples from audition (auditory nerve). We show that the resulting models, with weights determined from natural signal statistics, can qualitatively account for a number of nonlinear behaviors in neurons.

3.1 Primary Visual Cortex

We now compare the model with electrophysiological measurements from single neurons in area V1. Physiologists have distinguished between two basic types of

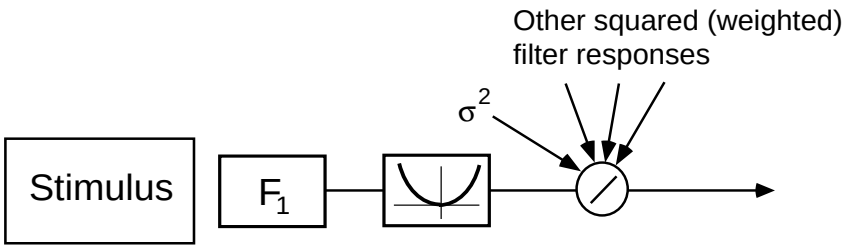

Figure 3.1. Phenomenological neural model for vision and audition. Each filter response is squared and divided by the weighted sum of squared responses of neighboring filters and an additive constant. Parameters are determined from natural signal statistics.

V1 experiments: inside the classical receptive field (center); and outside the classical receptive field (surround). The so-called classical receptive field is the spatial region that elicits an excitatory response in the neuron. We note at the outset that the spatial borders of the classical receptive field are not easily determined, and various methods can yield receptive fields of different sizes. We will discuss this in more detail later on in the chapter. The surround is considered a spatial region that by itself elicits no response in a neuron; but nevertheless a stimulus placed in this region can modulate the response of a neuron to stimuli placed inside the classical receptive field.

We first show that the statistically derived model can account for a number of nonlinear neural behaviors inside the classical receptive field, as in previous divisive normalization models [64, 65, 30]. Subsequently, we show that the model can also account for additional nonlinearities outside the classical receptive field. Physiological data inside the classical receptive field are taken from published journals, and data outside the classical receptive field were collected by Cavanaugh, Bair and Movshon ([33]).

3.1.1 Methods

Linear receptive fields are derived using the steerable pyramid multi-scale oriented decomposition [142]. The basis functions of this representation are localized in spatial position, orientation, and spatial frequency (scale). In this sense, they are qualitatively similar to simple cell receptive fields in area V1. The steerable pyramid is constructed by recursively splitting an image into subbands with directional derivative operators a given order (4 orientations chosen here).

In order to simulate an experiment, we first choose a primary filter and a set of neighboring filters that interact with the primary filter. The primary filter is the linear front end of the model neuron response, which is modulated by the responses of the neighboring filters. We make a simplifying assumptions that other filter choices behave similarly (see discussion). The primary filter is vertically oriented with peak spatial frequency of 1/8 cycles/pixel (second recursion level of the pyramid).

The filter neighborhood includes four orientations at the spatial frequency of the primary filter. At each of these orientations, displacements of neighboring spatial filters are horizontally, vertically, and diagonally subsampled at 3-pixel intervals (up to 9 pixel displacement in each direction). The farthest displacement (vertically and horizontally) extends to approximately 2 times the diameter of the primary filter. The neighborhood also includes four orientations at half the spatial frequency, but only at the center location (figure 3.2). There are a total of 103 filters in the neighborhood.

Next, we choose a natural image ensemble. The ensemble consists of 8 images from the Van Hateren database of calibrated images [156], and 4 images from a

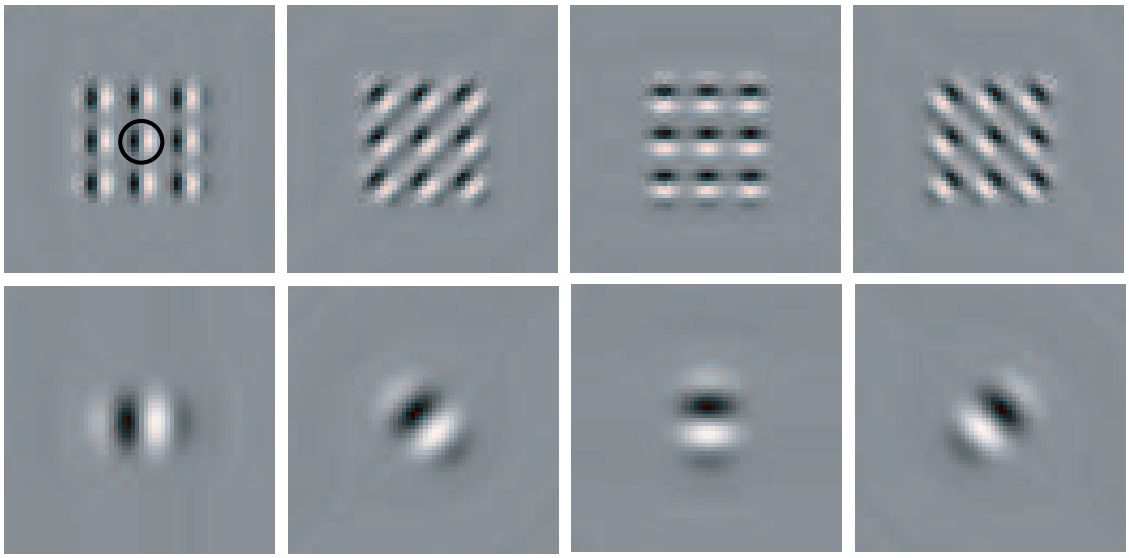

Figure 3.2. Neighborhood for vision optimization. The primary filter is circled. The figure does not show all neighboring filters, but depicts the spatial extent of the farthest filter at each orientation and spatial frequency.

database of standard images used in computer science (Figure 3.3). The images from the standard database are squared to achieve a rough calibration; typical digital images are corrected for a power-law monitor nonlinearity with an exponent of approximately 2. We crop each image to 256 by 256 pixels. We also include the symmetric left-right flipped version of each image, assuming that these are equally likely to occur.

We now run the maximum likelihood procedure described in chapter 2, with the chosen primary filter, neighborhood, and image ensemble. This results in a set of weights, and a constant σ. The weights in the model are pre-determined from the natural image ensemble and remain fixed throughout the model simulations. However, the constant σ is regarded as a free parameter to the model. As discussed in chapter 2, the value of the constant σ from the optimization depends on

Figure 3.3. Ensemble of images, from Van Hateren database of calibrated images [156], and from database widely used in computer science and neural modeling efforts.

the somewhat arbitrary scaling of the input signal. One can in theory adjust the optimal σ to match the input level of the experimental stimuli. However, we have found in practice that neurons also exhibit a range of different sensitivities. So we choose different values of σ in the simulations to match the sensitivity of the cell being modeled. The value of σ controls the contrast at which the response of the neuron reaches semi-saturation, and the contrast at which the divisive gain signal is not effective in modulating the response of the model neuron; for weak signals, σ dominates the divisive equation. Changing σ is equivalent to shifting the contrast scale.

Finally, we simulate each experiment, holding all weights of the model fixed. We present the model neuron with input, typically gratings that vary in some parameter. The response of the neuron is simulated by computing the normalized responses of the primary filter to the experimental stimuli. The response of the divisive neighborhood is assumed to be phase invariant, as if it is coming from either an ensemble of simple cells at a variety of phases, or from complex cells. Computationally, similar results are obtained by averaging the response over multiple phases (this would be similar to the summed response to a drifting grating), or by taking the sum of squared responses of two filters that are 90 degrees out of phase (e.g., quadrature pairs). Here we have chosen the latter solution, because it speeds up simulation time and does not alter the results.

3.1.2 Mapping out the Receptive Field

As a prior step to running experiments on the model, we need to map out a receptive field. We could just measure the excitatory extent of the primary filter in our model.

However, a physiologist does not have this luxury; so we also estimate the receptive field extent, as if we do not know the structure of the model. We assume a circular receptive field, as is often done in physiology (note that this is not exactly correct for our receptive fields, but it is also not exactly correct for real cells).

Ideally, we'd like to choose a method for estimating the receptive field that is most comparable with the physiological data. Unfortunately, there are no uniform standards for measuring the receptive field. In general, for experiments that probe only the classical receptive field and not the surround, the exact choice of receptive field extent has not been a major issue at all, so long as the selected region drives the cell. Experimenters aiming to separate out influences of center and surround have struggled with the problem of choosing the extents appropriately. Early experiments have typically hand-fitted the receptive field based on responses to oriented bars or lights. These procedures tend to underestimate the size of the receptive field by about a factor of two relative to more recent methods (see for example [33] for a discussion). Here we follow recent physiological procedures of expanding the diameter of an optimal sine grating until a maximal response (typically 95 percent of the maximum) is achieved [43, 33, 128]. This technique is matched to all the surround experiments of Cavanaugh et al. presented in the chapter.

The procedure is illustrated in figure 3.4a for our model neuron. When a grating of optimal orientation and spatial frequency is expanded in diameter, the mean response of the model neuron increases, until it reaches a peak. Subsequently, the response of the neuron declines and eventually tapers off to a minimal value. This increase and subsequent decrease of response is very typical of cells in area V1 (see for example, [33, 128]). Based on this procedure, the receptive field is mapped out

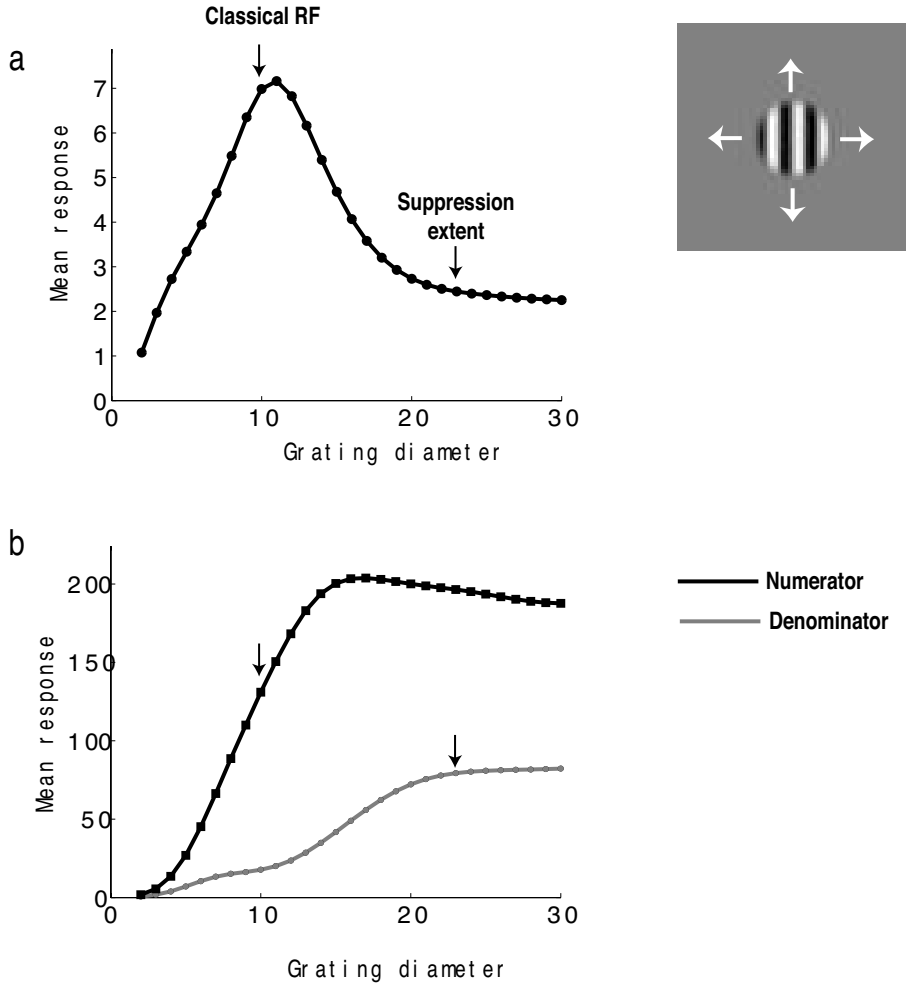

Figure 3.4. Mapping out the classical receptive field. **a**, Response of model neuron (from divisive normalization) to an expanding optimal sine wave grating. Arrows indicate 95 percent of the peak (corresponding to the estimated extent of the receptive field) and 95 percent of the minimal value (corresponding to the estimated suppressive effect). Procedure based on Cavanaugh et al. [33]. **b**, Components of the model leading to overall response in (a). Specifically, plotted are the numerator component (squared response of excitatory filter) and denominator component (sum of squared responses of the neighboring filters, with constant σ omitted in the plot). Arrows indicate same positions as in (a).

as a circular region with a diameter equal to 95 percent of the maximal value. In the model neuron, the estimated receptive field diameter spans 9 pixels (first arrow in the figure). The 95 percent minimum point following the decline in response is suggestive of the suppressive extent of the neuron [33]. In the model neuron, this point corresponds to 24 pixels (second arrow in the figure).

In physiology, it is difficult to disentangle the components that lead to this behavior. But since we know the components of the model, we can examine separately the excitatory numerator component response (coming from the true primary filter) and the divisive component (coming from the weighted squared neighboring neuron responses). We ignore for now the constant σ in the divisive component and assume it is relatively small and has minimal influence on the denominator. This is a reasonable assumption, since most physiological studies have fit the receptive field with high contrast stimuli. However, we will come back to this issue later on in the chapter.

Figure 3.4b illustrates the two components of the model. The numerator component is just the squared response of the primary filter. It continues to increase and reaches 95 percent of the peak at around a 14 pixel diameter. Note that this is approximately 1.5 times the estimated diameter. That is, we have underestimated the true spatial extent of the primary filter. The minor decrease in response following the peak is due to the structure of the linear filter, which has a small inhibitory sidelobe. The denominator component coming from the sum of squared responses of neighboring filters continues to increase farther out in space than the numerator component. The denominator reaches 95 percent of the maximum value at approximately 24 pixels. This is approximately 2.5 times the extent of the esti-

mated receptive field, and 1.7 times the extent of the true excitatory component. Note that the peak of the denominator component in the model (figure 3.4b) is also roughly equal to the minimum value of the overall model neural response (figure 3.4a). Thus, the overall model neural response gives a good indication of the suppressive extent in the model.

Now that we've mapped out a linear receptive field, we can proceed to compare model and physiology. We will point back to some of the issues raised here throughout the comparison.

3.1.3 Comparison to Physiology Inside the Classical Receptive Field

We first demonstrate behavior of the model for stimuli presented inside the classical receptive field.

Figure 3.5 shows data and model simulations of cross-orientation suppression. An optimal grating is placed inside the classical receptive field. An orthogonal mask grating is superimposed on the optimal grating (forming a plaid). Each curve in the figure indicates the response as a function of the contrast of the optimal grating, for a particular mask contrast. The sigmoidal shape of the curves results from the squaring nonlinearity and the normalization. The steepness of the model curves is determined by the fixed exponent of two, and appears to roughly match the cell.

The presence of the mask suppresses the responses to the optimal grating. Specifically, the contrast response curves are shifted to the right (on a log axis), indicative of a divisive gain change. The model exhibits similar behavior, which is a direct consequence of the statistically-chosen normalization weights. Neighboring

model filters responding to the orthogonal grating exert a divisive effect on the primary filter. This, in turn, arises from the statistical properties of natural images discussed in chapter 2.

Masking effects can also be observed at a single contrast, as a function of mask orientation. Figure 3.6 depicts the orientation tuning curve for a fixed optimal grating in the presence of a mask of varying orientation. As before, the optimal grating and the mask are additively combined to form a plaid. Also plotted is the orientation tuning curve for the model neuron without a mask (control). Both curves are normalized to a maximum value of 1. The mask exerts a mixture of excitatory and suppressive effects. In particular, the presence of the mask at orientations close to the optimal serves both to excite the primary filter, and to provide suppression through division. This results in a broadening of the orientation tuning curve, because the response is increased more near the optimal orientation, while the divisive component does not change as much as a function of orientation.

Orientation tuning in area V1 has been a widely studied topic. One of the striking features of orientation tuning is that it appears to be invariant to the contrast of the stimulus (e.g., [130, 145], and see also [50]). Figure 3.7 depicts the orientation tuning as a function of contrast for a typical V1 cell [145] versus the model. Both the cell and model orientation tuning are roughly contrast invariant.

In the context of a divisive normalization model, invariance of orientation tuning has been postulated to arise through equal weighting of the squared filter responses in the denominator [64]. If the divisive signal is equally weighted at each orientation, then the tuning of the primary filter cannot be altered. To examine the model behavior more closely, we separate out the components of numerator and

denominator (figure 3.8). The top panel depicts the same overall model response as in figure 3.7. The middle panel illustrates the two model components. The divisive denominator component is not truly flat as a function of orientation. This means that the model orientation tuning curves are only close to contrast invariance; at high contrast, the tuning is slightly sharper than at low contrast. This is possibly an artifact of the model. Specifically, the optimization procedure does not allow to include the excitatory filter or filters that are highly correlated with the excitatory filter. As a result, there is no source of self-suppression in the model (see discussion in chapter 2). Added self-suppression would help to achieve true contrast invariance. In our model, increasing the stimulus diameter would result in an added suppression at the optimal orientation. When we actually do this (bottom panel of the figure) the divisive signal is roughly constant as a function of orientation.

Experimentally, it would be interesting to understand if the contrast invariance of V1 neurons holds up across stimuli of different sizes. Though contrast invariance of orientation tuning is widely published, we believe it has only been examined for certain stimulus sizes, most probably small ones that match hand-computed receptive fields [130, 145]. Perhaps cells are most optimized for orientation invariance at particular stimulus sizes. For example, if the orientation tuning of the surround suppression is more sharply tuned than the center, then very large stimuli will have added suppression at the preferred orientation, and this could potentially result in a broader tuning curve at high contrasts. But less relative suppression at the preferred orientation for small stimuli could potentially result in a sharper tuning curve at high contrasts (as in our model). Sclar and Freeman note in their paper that "For only two cells, tuning patterns were marginally different from zero and

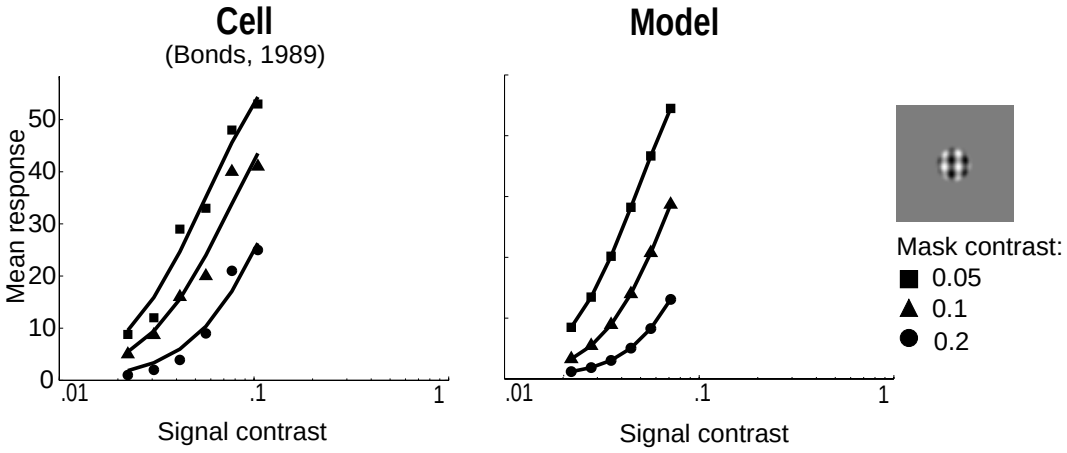

Figure 3.5. Cross-orientation suppression in area V1 [24]. Responses to optimal stimulus are suppressed by an orthogonal masking stimulus within the receptive field. This results in a rightward shift of the contrast response curve (on a log axis). Curves on cell data plot are fitted with a Naka-Rushton function, $r(c) = c^2/(ac^2 + b^2)$.

in these cases, tuning sharpness appears, counterintuitively, to decrease slightly as contrast increases" [130]. The relation between stimulus size and orientation invariance needs to be examined experimentally.

3.1.4 Mapping the Region Outside the Classical Receptive Field

As stated earlier, stimuli outside the classical receptive field are useful in experimental settings, because they elicit no response in the neuron when presented alone. Nevertheless, they can modulate the response of a neuron to stimuli placed inside the classical receptive field.

The region outside the classical receptive field is usually chosen as an annular ring surrounding the receptive field. Selecting the dimensionality of the annulus

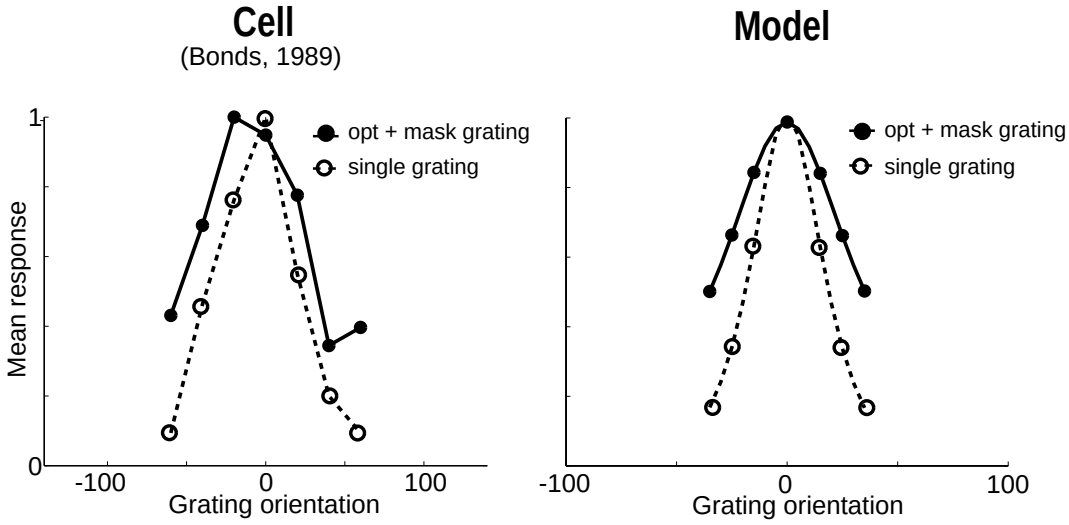

Figure 3.6. Orientation masking in area V1 inside the classical receptive field [24]. Dashed line indicates response to a single grating, as a function of orientation. Solid line indicates response to an optimal grating additively superimposed on a mask grating of variable orientation. All curves are normalized to have a maximum value of one.

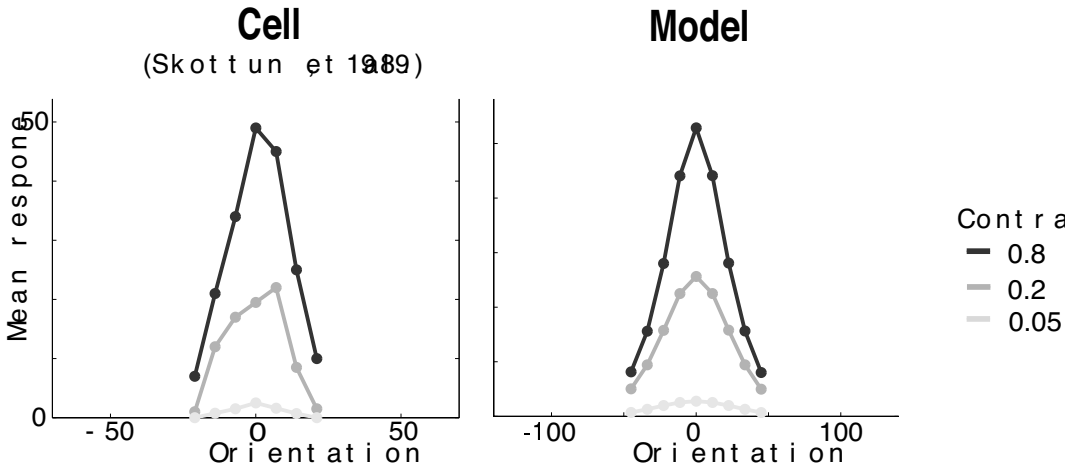

Figure 3.7. Contrast invariance of orientation tuning in area V1 [145]. Cell and model are plotted such that their bandwidth is visually equated, even though the specific bandwidth differs.

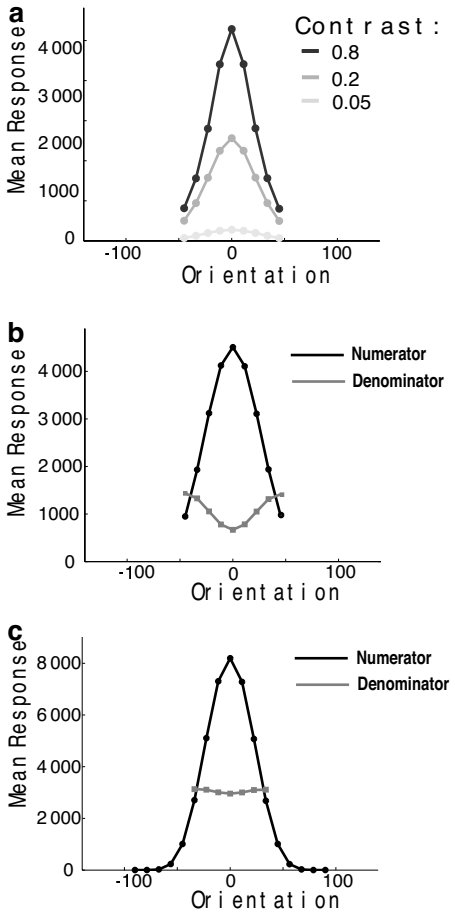

Figure 3.8. Examination of model components of orientation tuning as a function of contrast in area V1. **a**, Overall model response, as in figure 3.7. **b**, Components of the model leading to overall response in (a). Specifically, plotted are the numerator component (squared response of excitatory filter) and denominator component (sum of squared responses of the neighboring filters, with constant σ omitted in the plot). Note that denominator component is not truly constant as a function of orientation. **c**, Components of the model for stimulus that is twice the size of the estimated receptive field. Now the denominator component is roughly constant as a function of orientation, leading to true orientation invariance.

(and especially its most inner border) is again a rather tedious issue for physiologists. If one underestimates the surround, then it could actually exert an excitatory effect on the neuron. This effect could be particularly extreme if the surround inner annulus diameter is chosen equal to the diameter of the estimated receptive field. Remember in the model that the estimated receptive field extent was smaller than the true extent of the receptive field (figure 3.4). Moreover, estimated receptive fields using small bars or spots of light are even smaller. If special care is not taken to separate out the center and surround, the physiological results could be contaminated by interactions between these two components.

Only recently, physiologists have taken more scrutiny to separate out center and surround. Cavanaugh et al. have developed a method, which we follow here, for determining the surround [33]. The method is illustrated in figure 3.9a for the model neuron. A zero-luminance "hole" is placed in the center. Surrounding the hole is an annular ring at the optimal orientation and frequency for the model neuron. The diameter of the hole is expanded, which is equivalent to expanding the diameter of the inner border of the annulus. Now the 95 percent minimal value is taken to represent the diameter of the inner border of the annular surround. This amounts to a 12 pixel diameter in the model neuron. An annular ring of zero luminance is then placed between the estimated center and surround.

Figure 3.9b illustrates the two components of the model. Note that both the excitatory numerator and suppressive denominator decrease for larger proximal borders of the annulus, but the numerator decreases more rapidly to a minimal value. The 12 pixel diameter that was chosen based on the overall model neuron response is quite well matched to the minimal value of the numerator alone. We

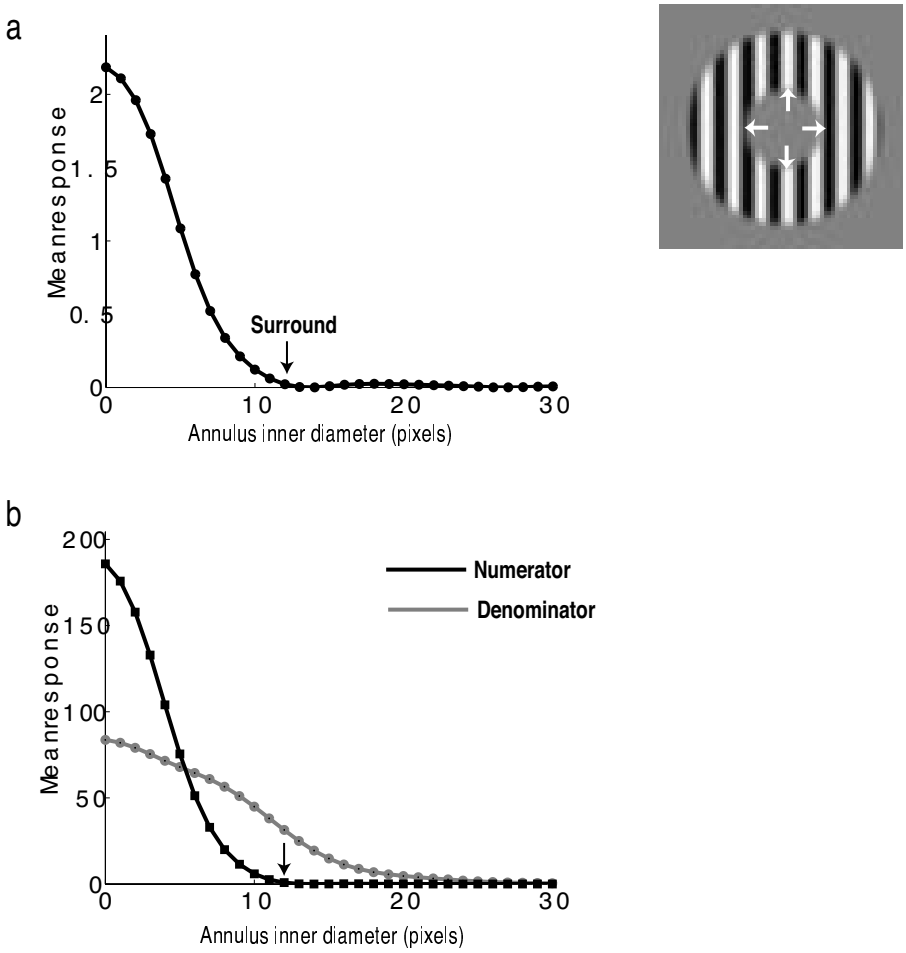

Figure 3.9. Mapping out the surround annulus inner diameter. **a**, Response of model neuron (from divisive normalization) to an optimal grating with a "hole" of zero luminance. As the hole is expanded, the proximal border of the annulus is increased. Arrow indicates 95 percent of the minimal value. Procedure based on Cavanaugh et al. [33]. **b**, Components of the model leading to overall response in (a). Specifically, plotted are the numerator component (squared response of excitatory filter) and denominator component (sum of squared responses of the neighboring filters, with constant σ omitted in the plot). Arrow indicates same position as in (a).

should point out that this value is almost equal to, but slightly below, the peak of the numerator of an expanding sine grating computed in figure 3.4b. The inconsistency between the two methods is a consequence of the squaring nonlinearity of the numerator kernel response. Thus, we have come close to but marginally underestimated the real inner border of the surround. That is, an optimal stimulus placed in the surround will elicit no response, as required. But the opposite is not true; a stimulus expanding from the center will still slightly increase its response as it encroaches onto the surround. Remember that the physiologist does not have access to the true extent of the presumed center component.

3.1.5 Comparison to Physiology Outside the Classical Receptive Field

We now compare model responses to steady-state nonlinear physiological behaviors outside the classical receptive field.

Figure 3.10 shows data from a Macaque monkey [33]. An optimal sinusoidal grating stimulus is placed inside the classical receptive field of a neuron in area V1. A mask grating is placed in the annular region surrounding the classical receptive field. Each curve in the figure indicates the response as a function of the center contrast for a particular surround contrast.

Presentation of the mask grating alone does not elicit a response from the neuron, but its presence suppresses the responses to the center grating. Specifically, the contrast response curves are shifted to the right (on a log axis), indicative of a divisive gain change. When the mask orientation is parallel to the center, this shift is much larger than when the mask orientation is orthogonal to the center (Figure 3.10).

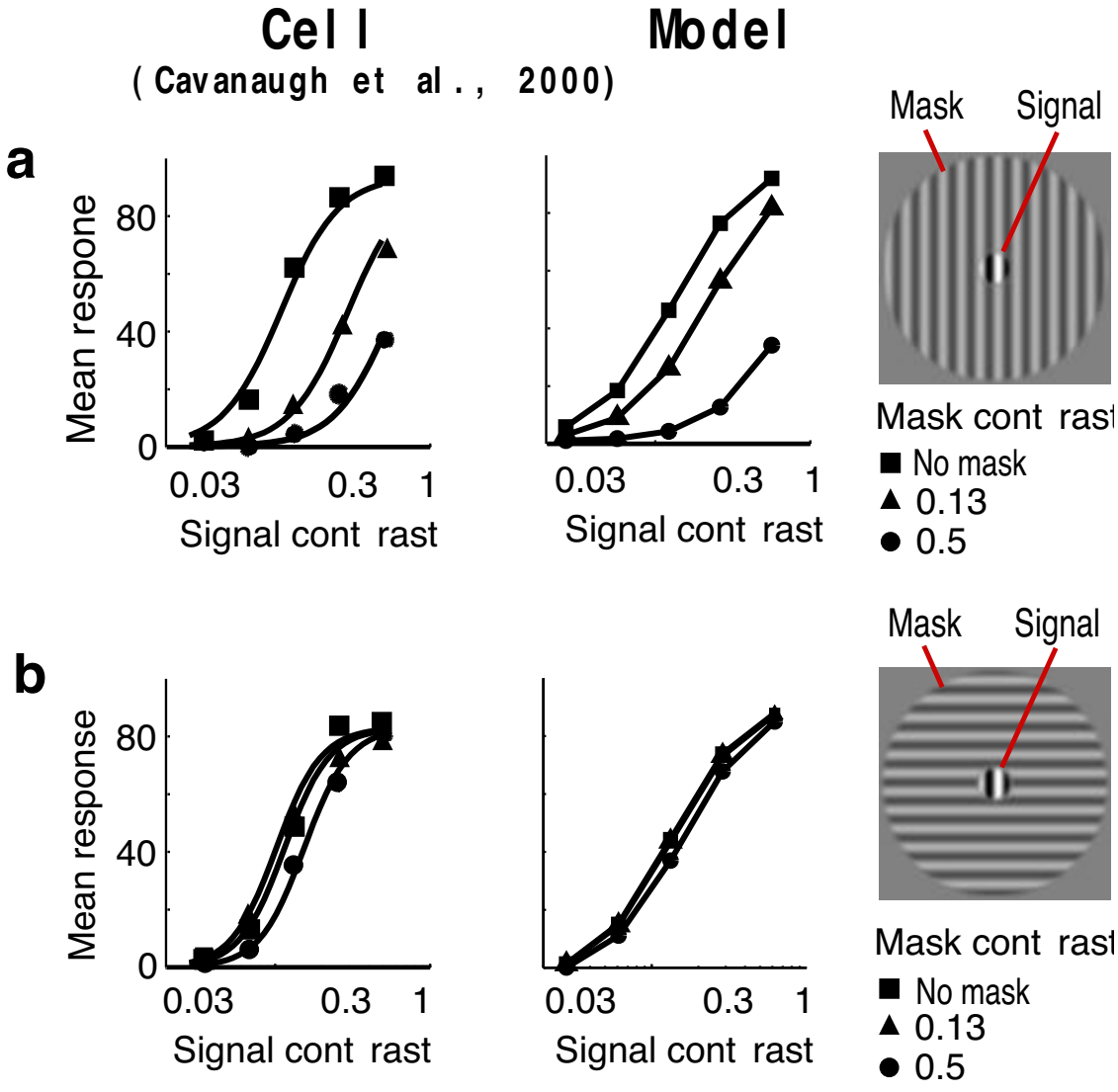

Figure 3.10. Contrast response curves for parallel versus orthogonal surround masks in area V1. cell data from [33]. **a,** Mean response rate of a V1 neuron as a function of contrast of an optimally oriented grating presented in the classical receptive field, in the presence of a surrounding parallel masking stimulus. Curves on cell data plots are fits of a Naka-Rushton equation with two free parameters [33]. **b,** Mean response rate vs. center contrast, in the presence of an orthogonal surround mask.

The model exhibits similar behavior (Figure 3.10), even though we have not fitted the model to the data. The model behavior is due to a suppressive weighting of neighboring model neurons with the same orientation preference that is stronger than that of neurons with perpendicular orientation preference (see also [144]). Note that this behavior would not be observed in previous equally weighted normalization models, because the parallel and orthogonal surround stimuli would produce the same normalization signal.

The weighting in our model is determined by the statistics of our image ensemble, and is due to the fact that adjacent regions in natural images are more likely to have similar orientations than orthogonal orientations. For example, oriented structures in images (such as edges of objects) tend to extend along smooth contours, yielding strong responses in linear filters that are separated from each other spatially, but lying along the same contour (see also [60, 138]). In other experiments in V1, it has been demonstrated that for a parallel surround, more suppression is observed for aligned patches in the surround than for non-aligned patches [33]. Our model is expected to be consistent with this result, since we obtain stronger weighting for iso-oriented filters that are aligned with the primary filter, than for non-aligned filters (also see bowties in chapter 2).

We have thus far examined surround suppression as a function of only two surround orientations, parallel and orthogonal to the optimal. Figure 3.11 explores a wider range of surround orientations at a single (high) contrast. The center orientation is held fixed at the optimal orientation of the neuron. Both in cell and in model, most suppression is observed when the surround stimulus is also placed at the preferred orientation of the neuron. The amount of maximal suppression is

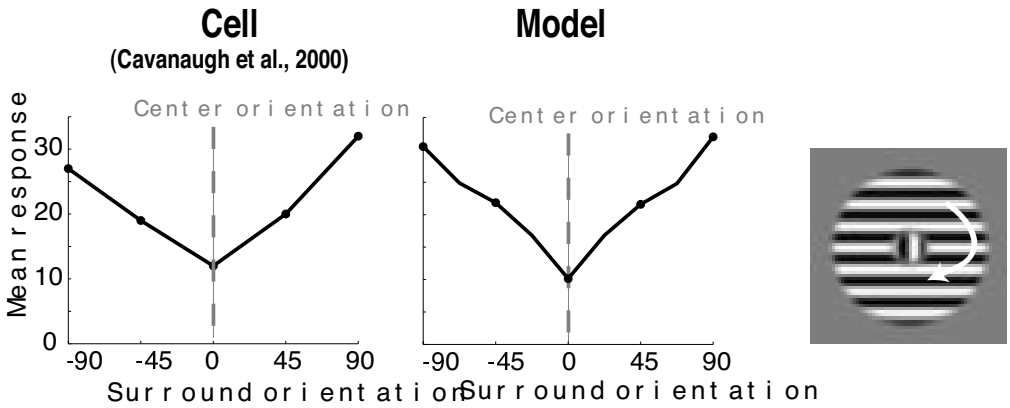

Figure 3.11. Surround suppression as a function of orientation for optimal center stimulus in area V1. Maximum suppression is obtained when the surround orientation is also at the preferred orientation of the neuron. Model neuron is plotted with twice the number of orientation points, but marked points correspond to cell. Model is scaled to have same maximal value as cell.

also very similar in model and cell.

This behavior has led experimenters to wonder whether the absolute surround orientation drives the amount of suppression, or whether the relative orientation between center and surround is pertinent. It has been suggested that the amount of surround suppression depends contextually on the orientation of the grating placed in the center, consistent with the latter hypothesis. Specifically, some experimental data show that neurons appear to shift their point of maximal surround suppression to match the orientation of a grating placed in the center [139, 33]. This seems to be counterintuitive to the type of divisive normalization model we've described, in which the point of maximal suppression ought to be fixed and determined by the weighting of the divisive pool. For example, in our model neuron, we would always expect maximal suppression for a surround stimulus with the preferred orientation

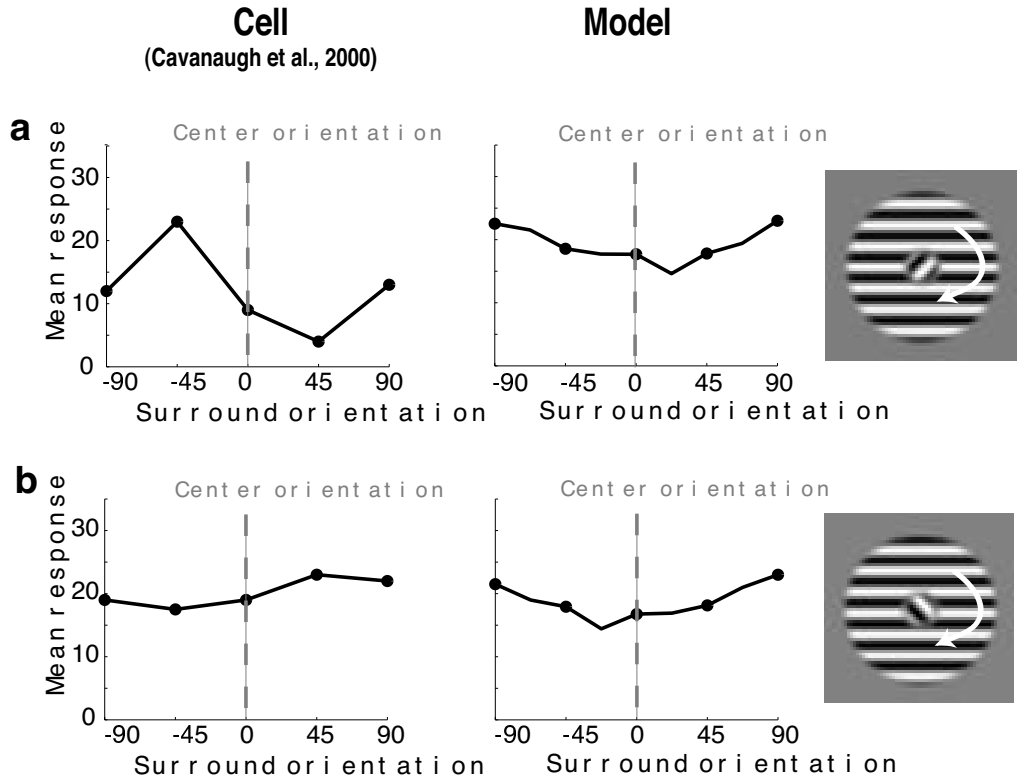

Figure 3.12. Dependence of surround suppression on center orientation in area V1. **a,** Surround orientation suppression when center orientation is fixed at +45 degrees from the preferred orientation of the neuron. Maximum suppression is obtained when the surround orientation is equal to the orientation of the center grating (for the cell) and shifted towards the orientation of the center grating (for the model). **b,** Surround orientation suppression when center orientation is fixed at -45 degrees from the preferred orientation of the neuron. Maximum suppression is obtained when the surround orientation is equal to the orientation of the center grating (for the cell) and shifted towards the orientation of the center grating (for the model). Physiological data from [33]. Model neuron is plotted with twice the number of orientation points, but marked points correspond to cell. Model is scaled to have same maximal value as cell.

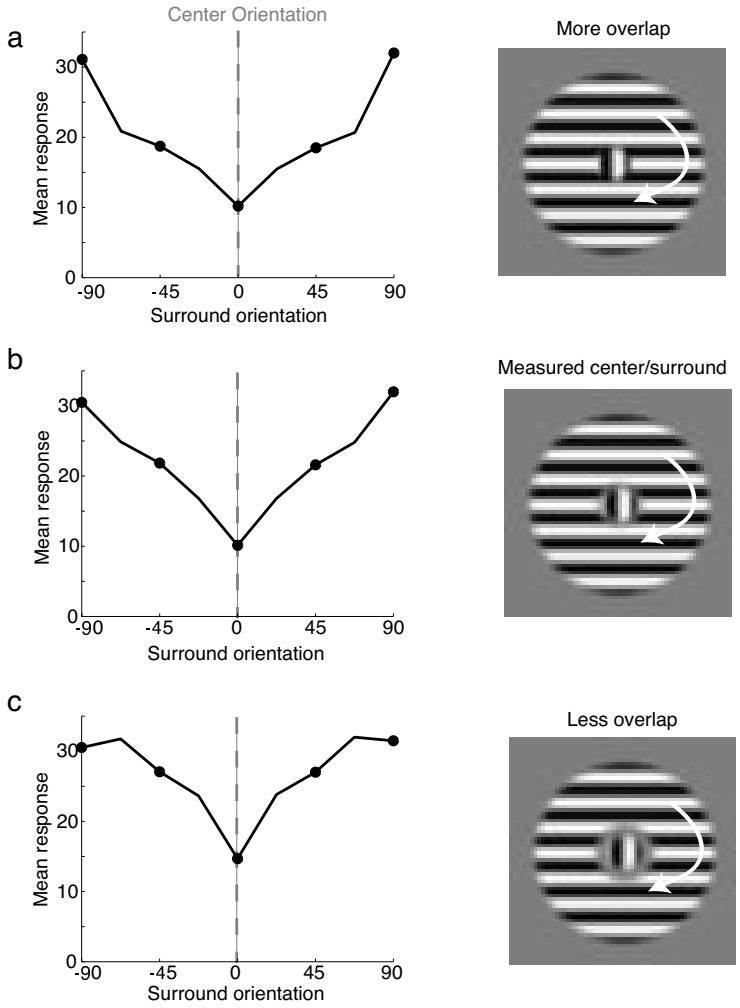

Figure 3.13. Model dependence of surround suppression for different separations of center and surround with optimal center stimulus. **a**, Center stimulus 9 pixels; surround stimulus 10.5 pixels (i.e., smaller than estimated surround). **b**, Center stimulus 9 pixels; surround stimulus 12 pixels. (i.e., equal to estimated surround). **c**, Center stimulus 9 pixels; surround stimulus 16 pixels. (i.e., larger than estimated surround). In all cases, maximal suppression is achieved when the surround grating is at the preferred orientation of the model neuron.

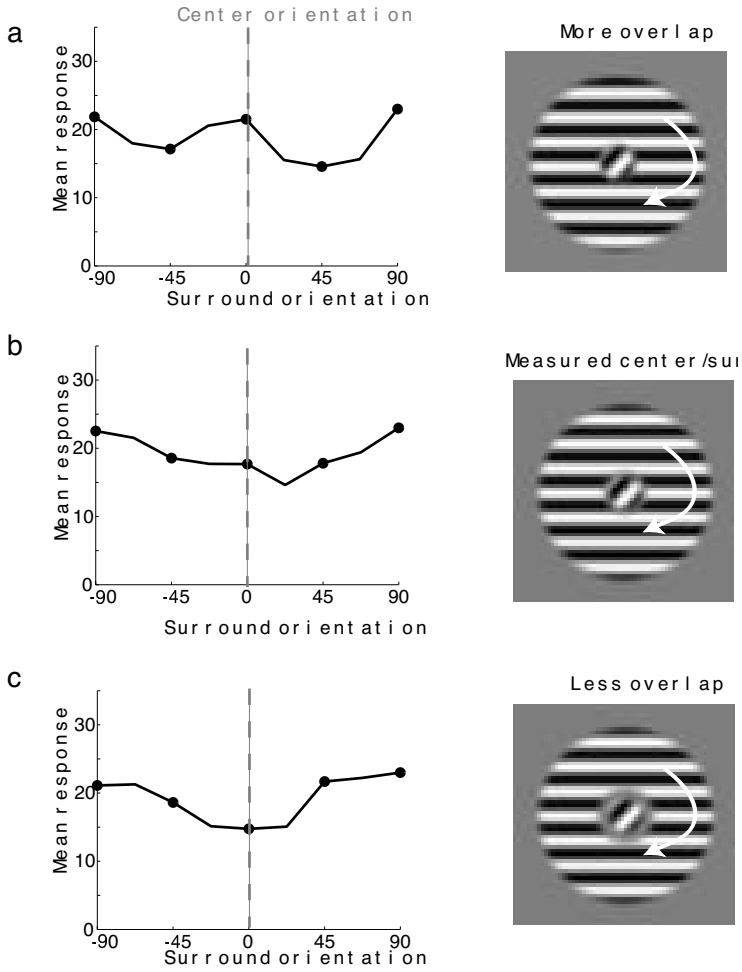

Figure 3.14. Model dependence of surround suppression for different separations of center and surround with non-optimal center stimulus. Center orientation is fixed at the +45 degrees away from the preferred orientation of the neuron. **a**, Center stimulus 9 pixels; surround stimulus 10.5 pixels (i.e., smaller than estimated surround). **b**, Center stimulus 9 pixels; surround stimulus 12 pixels. (i.e., equal to estimated surround). **c**, Center stimulus 9 pixels; surround stimulus 16 pixels. (i.e., larger than estimated surround).

of the neuron, regardless of the center stimulus orientation.

We decided to test this idea through simulation in the model. Figure 3.12a illustrates cell and model mean response rate as a function of surround orientation, for a center stimulus that is fixed at +45 degrees the optimal orientation. The point of maximal suppression now shifts towards +45 degrees in the cell. In the model, the shift is not quite as extreme; a maximal suppression is seen at +22.5 degrees. Note that we have sampled orientation at a finer scale in the model than in the experimental data to illustrate this point. Similar behavior is observed for a center stimulus that is fixed at -45 degrees the optimal orientation (figure 3.12b): the cell and model minimum is shift towards -45 degrees. There is also a broadening of tuning (relative to figure 3.11) in the model and (especially at -45 degrees) in the cell.

The model does not behave in detail like the cell. But the fact that the model also obtains a shift toward the orientation of the center stimulus is interesting. This kind of behavior could arise if the center and surround of the model are not truly separated. In fact, center and surround are not truly separated in the model. Recall that the surround was slightly underestimated in the inner annulus expansion experiment. So stimuli placed in the surround may have a minor effect on the excitatory filter of the model. In addition, stimuli placed in the center will naturally also effect the divisive normalization signal. This is expected because the suppressive pool includes filters in the center and filters at 3 pixel displacements from the center, which would still be influenced by the spatial region of the classical receptive field.

To test the hypothesis that the behavior in the model arises from incomplete

separation of center and surround, we decided to alter the separation between center and surround stimuli. That is, the center receptive field is held at the fixed estimated diameter, and the surround inner annulus diameter is modified. We either reduce (to 10.5 pixels), keep constant (12 pixels), or increase (to 16 pixels) the inner annulus diameter. Figure 3.13 depicts the behavior of the model to a center stimulus at the optimal orientation. As before, maximal suppression is obtained when the annular surround stimulus is also at the optimal orientation. This behavior is consistent for all three conditions. Figure 3.14 depicts the behavior of the model to a center stimulus shifted by +45 degrees. This time the model behaves differently in all three conditions. For smaller separation, the maximal suppression is at around +45 degrees. In contrast, for larger separation, the maximal suppression is centered approximately at the optimal orientation of the model neuron. Note that for larger separation we guarantee that the surround stimulus does not encroach on the excitatory filter. However, the center stimulus still does encroach on the suppressive filters. Analogous behavior is observed for a center shifted to -45 degrees (not shown here).

We have demonstrated that incomplete separation of center and surround could lead to changes in model behavior. This is a preliminary result; whether it has bearing to neural data requires further investigation. Perhaps biology is set up such that two stimuli that closely border one another contextually shift the orientation of maximal suppression. In any case, a divisive normalization model can produce this type of effect, at least qualitatively. Cavanaugh et al. have suggested another explanation for this behavior in the context of a divisive gain model: that changing the balance between a center and surround gain (such that the center gain

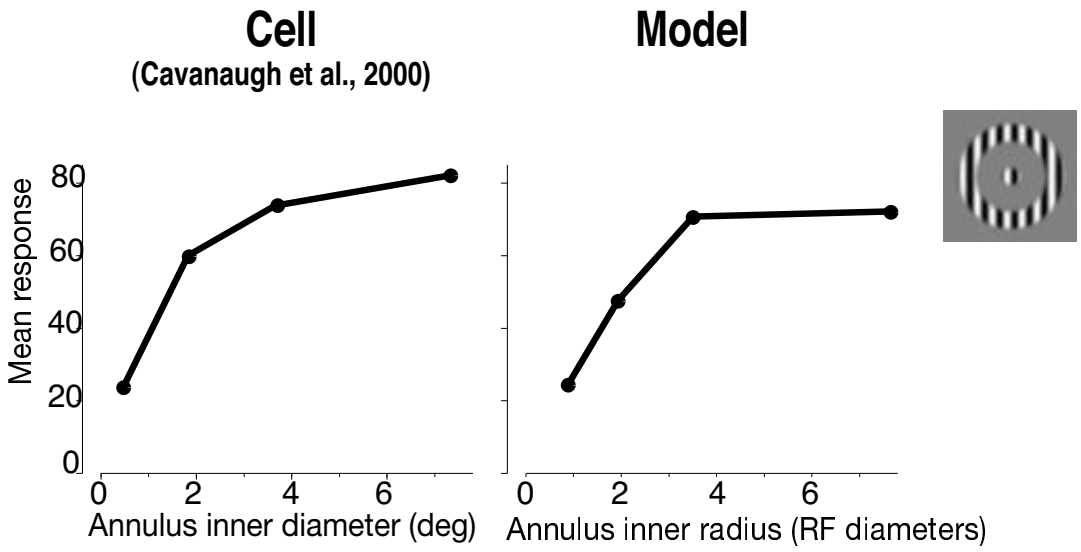

Figure 3.15. Response as a function of surround inner diameter in area V1. Physiological data from [33].

is favored) could amount to a shift in the orientation of maximal suppression [33].

We now examine the suppressive effect of the surround as a function of spatial proximity to the center. In figure 3.15 a surround mask is placed at the optimal orientation. As the surround mask is moved farther away from the receptive field, the suppressive effect is reduced both in the cell and model neuron.

As mentioned above, a motivating characteristic of normalization models has been the preservation of tuning curve shape under changes in input level. However, the shapes of physiologically measured tuning curves for some parameters exhibit dependence on input level [128, 76, 33]. Figure 3.16 shows an example of this behavior in a neuron from primary visual cortex of a Macaque monkey [33]. The graph shows the response of the cell as a function of the radius of a circular patch of sinusoidal grating, at two different contrast levels. The high-contrast responses

are generally larger than the low-contrast responses, but in addition, the shape of the curve changes. Specifically, for higher contrast, the peak response occurs at a smaller radius. The same qualitative behavior is seen in the model neuron, although weaker than in the physiological cell.

The behavior in the model is a result of two competing factors: primary filter excitation and neighboring filter suppression. At high contrasts, the suppressive signal modulates the shape of the tuning curve of the primary filter. Recall figure 3.4, in which neighboring filters that are spatially displaced relative to the primary filter reduce the excitatory response of the primary filter through division. This results in both a shift of the peak to a smaller diameter, and a continued decline in response following the peak. But at low contrasts, the tuning function is primarily determined by the excitatory numerator component, because σ dominates the divisive gain signal. So the tuning as a function of radius is more matched to the linear tuning properties of the primary filter.

3.1.6 Cell and Model Variations in area V1

We have modeled the behaviors of "typical" cells, but there are quantitative differences amongst individual cells. Figure 3.17 shows an example of variation across cells in physiological recordings in V1. In this figure, the amount of suppression for a parallel and orthogonal surround varies across the population. Amount of suppression is computed as the response of the neuron to an optimal grating inside the classical receptive field together with a surround mask (either parallel or orthogonal to the optimal orientation), divided by the response of the neuron to an optimal grating and no surround mask. This value is computed for high contrasts

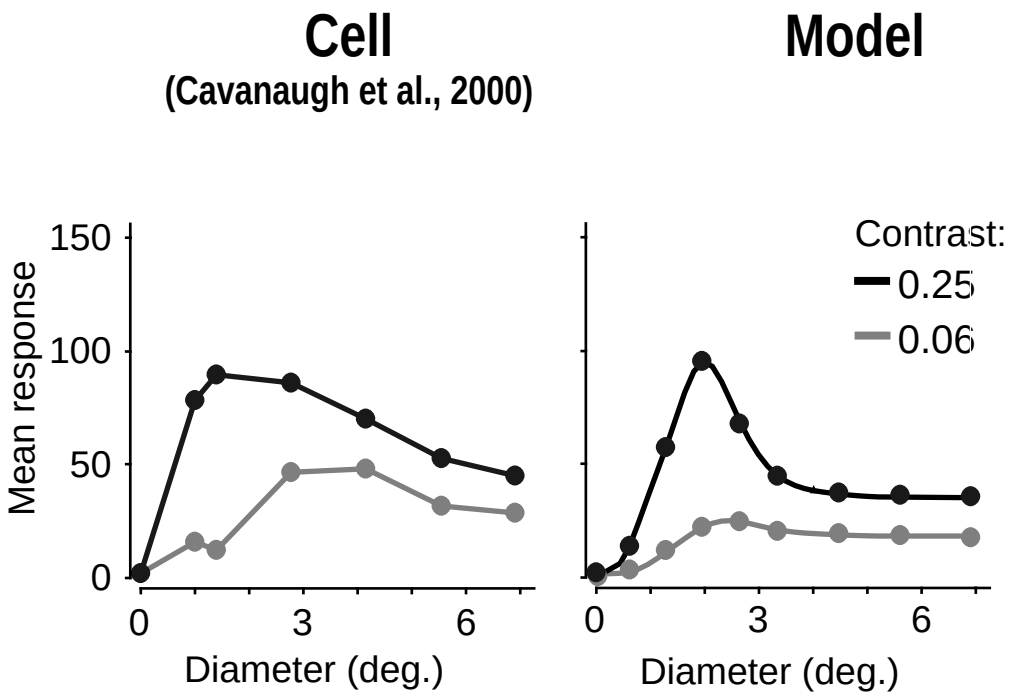

Figure 3.16. Nonlinear changes in tuning curves as a function of stimulus diameter for different input levels in area V1. Plotted is the mean response rate of a V1 neuron as a function of stimulus radius for two different contrasts [33]. The peak response radius for both cell and model is smaller for the higher contrast.

of the stimulus and mask. Each point in the graph corresponds to data from a single cell. A large proportion of the cells show substantially more suppression when the preferred orientation is in the surround than when the orthogonal orientation is in the surround. Note that some of the cells also exhibit facilitation due to the surround, although suppression is much more prominent.

Variations can also arise in the model. One source of variation in the model is observed when the divisive weights are optimized for different images. Figure 3.18 demonstrates the amount of suppression for a parallel versus orthogonal surround. Each point in the figure corresponds to a simulation with weights computed from a single image. Overall we ran simulations over 40 different images. The model follows a similar trend to the physiological data; there is more suppression when the preferred orientation is in the surround than when the orthogonal orientation is in the surround. The gray arrow in the figure points to the values obtained with the weights from the image ensemble used throughout the chapter. Figure 3.19 shows that images from the Van Hateren versus the standard image data set follow a similar distribution. White noise, in contrast, yields roughly the same amount of suppression for parallel and orthogonal surround.

The definition of "natural images" is quite loose. Some images in the set yield points away from the prominent data cloud, although these too still exhibit more suppression for the parallel than for the orthogonal surround. Shown in the figure are two such images, that appear rather texture-looking. Textures (both artificial and natural) are likely to have more variant statistics than so-called natural scenes (see also [110]).

We have illustrated one possible source of model variation, but variations in

the model may also arise from other factors, such as the choice of primary and neighborhood filters. We have chosen one primary filter in the model, and compared simulation results to neurons with different tuning properties. We have thus assumed invariance with respect to frequency and orientation. Images are roughly spatial frequency invariant, and we have verified that modifying the spatial frequency of the primary filter does not alter the qualitative results in the chapter. However, we believe that images are not orientation invariant; for example, vertical, horizontal and diagonal orientations probably do not appear with equal probability in nature. We are beginning to investigate how the results vary with orientation of the primary filter. Preliminary examination suggests that the qualitative nature of the simulations in the chapter does not change. However, the strength of the suppression, and relative contributions of aligned and non-aligned positions are likely to change. Since we also do not have a detailed account of changes in physiology as a function of neuron orientation and frequency, the assumption of invariance in the model is not unreasonable.

These two sources of variation, experimental and model, are very different; it is difficult to directly compare between them. There are numerous possible explanations for the variations that arise in physiological recordings, including cells from different animals; different layers; with different tuning properties; and (perhaps) one can also hypothesize that different cells are optimized for different subclasses of images.

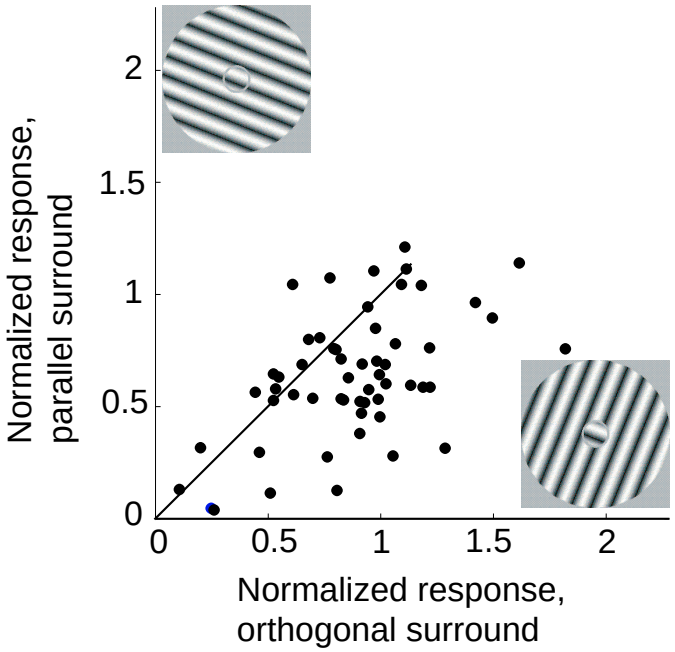

Figure 3.17. Distribution of physiological recordings from different cells, for parallel versus orthogonal surround in area V1. Each point corresponds to data from a single cell. The x axis is the response of the neuron to a center optimal grating with an orthogonal surround, divided by the response of the neuron to the center grating alone. The y axis is the response of the neuron to a center optimal grating with a parallel surround, divided by the response of the neuron to the center grating alone. Most points in the graph lie below the unit slope line, indicating that there is more suppression for parallel surround than for orthogonal surround.

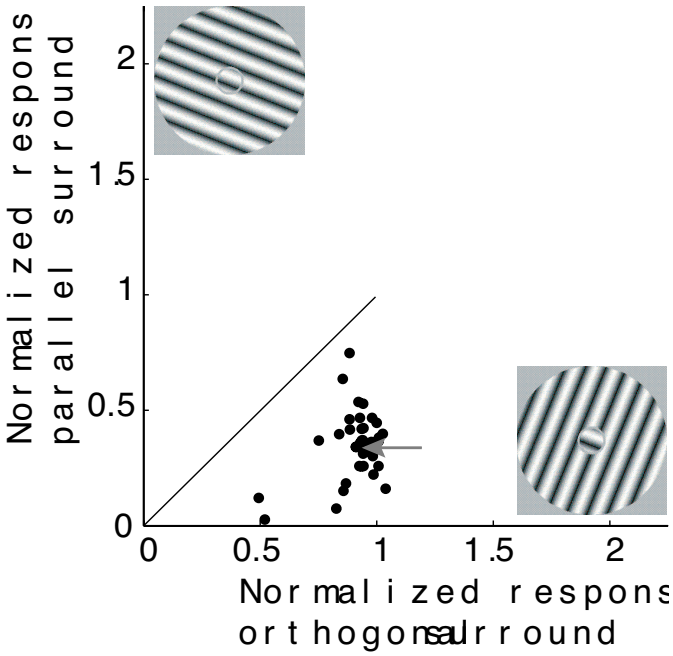

Figure 3.18. Distribution of model simulations with weights computed for different natural images, for parallel versus orthogonal surround. Each point corresponds to a simulation with weights computed for a single image. The x axis is the response of the model neuron to a center optimal grating with an orthogonal surround, divided by the response of the model neuron to the center grating alone. The y axis is the response of the model neuron to a center optimal grating with a parallel surround, divided by the response of the model neuron to the center grating alone. Most points in the graph lie below the unit slope line, indicating that there is more suppression for parallel surround than for orthogonal surround. The gray arrow points to simulation results for the image ensemble used in the chapter.

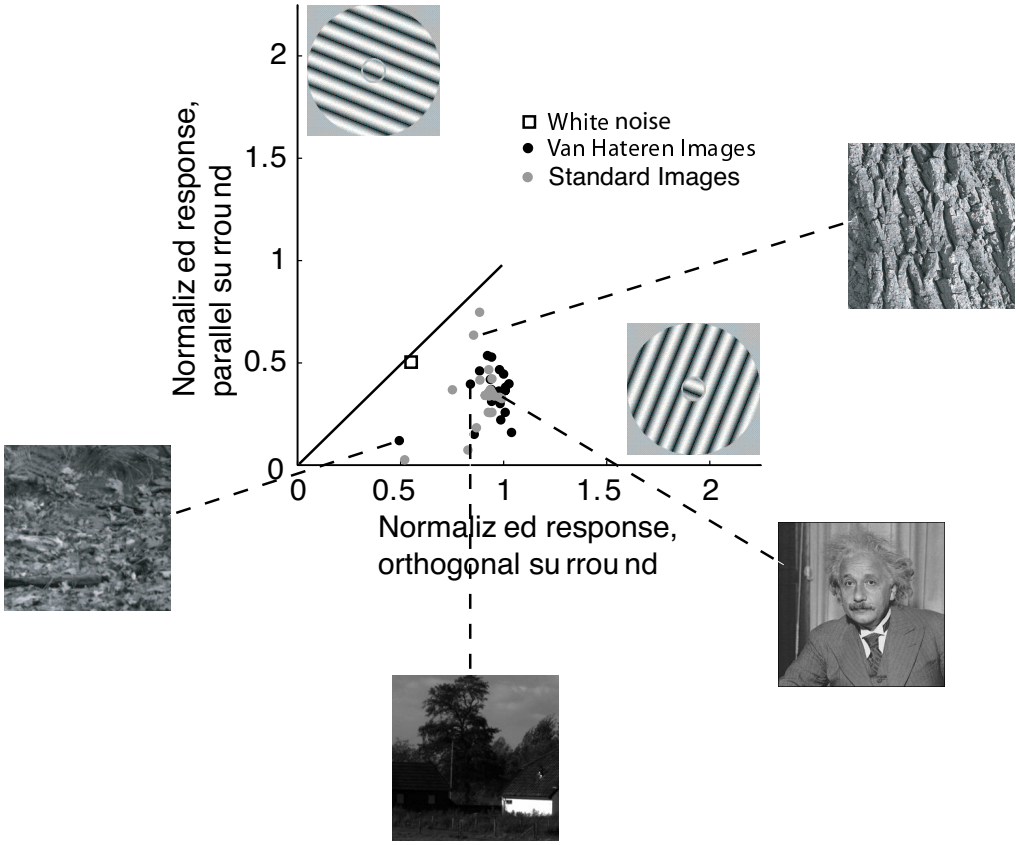

Figure 3.19. Distribution of model simulations with weights computed for different natural images, for parallel versus orthogonal surround (part 2). Same as Figure 3.18, but we plot separately simulation results from the Van Hateren image database and the standard image database. Also shown is a simulation with Gaussian white noise. Unlike the images, the white noise simulation yields roughly the same suppression for parallel and orthogonal mask gratings (lying close to the unit slope line).

3.2 Auditory Nerve

We have found interesting parallels in auditory nerve data. Though we have not studied the auditory behaviors as extensively as vision, we show below that an analogous auditory model can account for some basic nonlinearities in auditory nerve fibers.

3.2.1 Methods

For the auditory simulations we use a set of Gammatone filters as the linear front end [146]. We choose a primary filter with center frequency 2000 Hz, and a neighborhood of filters for the normalization signal: 16 filters with center frequencies 205 to 4768 Hz, and replicas of all filters temporally shifted by 100, 200, and 300 samples (figure 3.20). The time replicas should be thought of as discrete snapshots of the response of the same filter over time; in a real system, we expect the response over time to be more continuous.

The ensemble of natural sounds consists of 9 animal and speech sounds (cat, monkey, chimpanzee, and so on), each approximately 6 seconds long. The sounds are obtained from commercial compact disks and converted to sampling frequency of 22050 Hz.

We run the maximum likelihood procedure described in chapter 2, with the chosen primary filter, neighborhood, and sound ensemble. We now simulate each experiment, holding all weights of the model fixed (and hand fitting σ). The response of the neuron is simulated by computing the normalized responses of the primary filter to the experimental stimuli. As in the visual case, we assume in this

computation that the divisive signal is independent of phase. This is achieved by averaging the squared filter responses over time. We compare simulations of the model neuron to electrophysiologically auditory nerve data published in journal papers.

3.2.2 Comparison to Auditory Nerve Physiology

We concentrate both on masking experiments (known as "two-tone suppression"), and presentations of single tones at different frequencies to the neuron. These experiments are similar in flavor to some of the visual experiments, but with emphasis on temporal frequency instead of orientation.

Figure 3.21 shows data from a two-tone suppression experiment, in which the response to an optimal tone is suppressed by the presence of a second tone of non-optimal frequency. Two-tone suppression is often demonstrated by showing that the rate-level function of the optimal tone alone is shifted to the right in the presence of a non-optimal tone. In both cell and model, we obtain a larger rightward shift when the non-optimal tone is relatively close in frequency to the optimal tone (1.25 times the optimal), and almost no rightward shift when the non-optimal tone is more than two times the optimal frequency. The model behavior is due to variations in suppressive weighting across neurons tuned for adjacent frequencies, which in turn arises from the statistical properties illustrated in in chapter 2, in which stronger dependency is seen for nearby frequency bands in typical natural sounds.

Now what happens in the model if the mask frequency is even closer to the optimal? In this case, we expect the mask to elicit excitation in addition to suppression. Figure 3.22 illustrates this behavior in the model for a mask that is 1.15

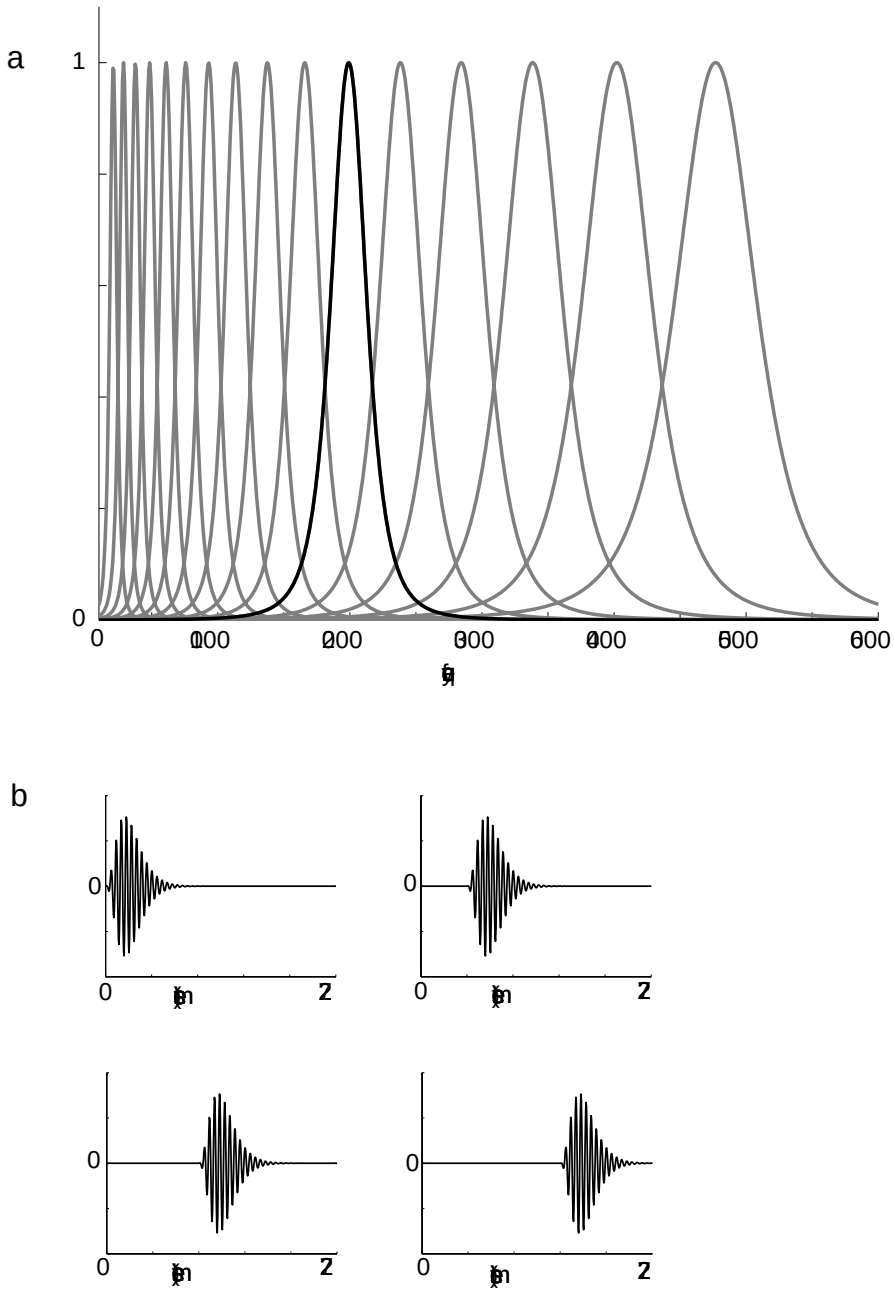

Figure 3.20. Neighborhood for auditory optimization. Primary filter is in black; neighborhood filters are in gray. Neighborhood includes 16 frequency bands and 3 time shifts of each band. **a**, Frequency tuning of all filters in the neighborhood. Note that frequency bandwidth is roughly equal on a log scale. **b**, Time representation (including 3 time shifts) of the primary filter.

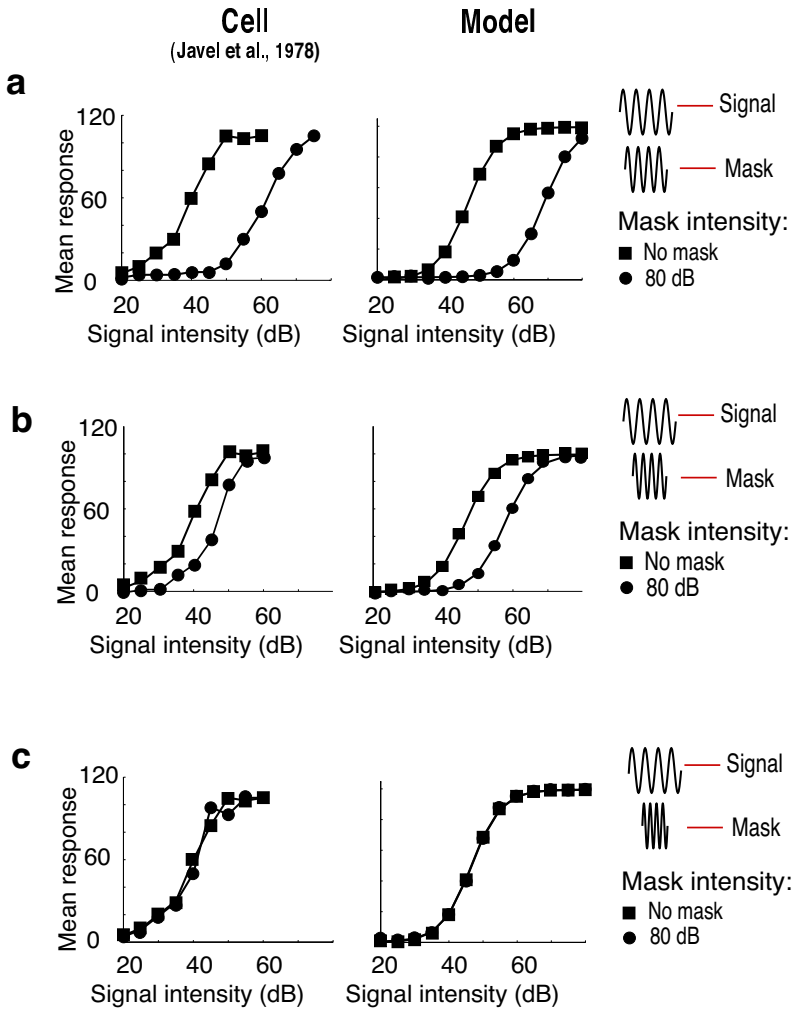

Figure 3.21. Suppression of responses to optimal stimuli by masking stimuli in auditory nerve fiber. Physiological data from [71]. **a**, Mean response rate of an auditory nerve fiber a function of sound pressure level, in the presence of a non-optimal mask at 1.25 times the optimal frequency. **b**, Mean response rate vs. sound pressure level, in the presence of a non-optimal mask at 1.56 times the optimal frequency. **c**, Mean response rate vs. sound pressure level, in the presence of a non-optimal mask at 2.08 times the optimal frequency. For all plots, maximum model response has been rescaled to match that of the cell.

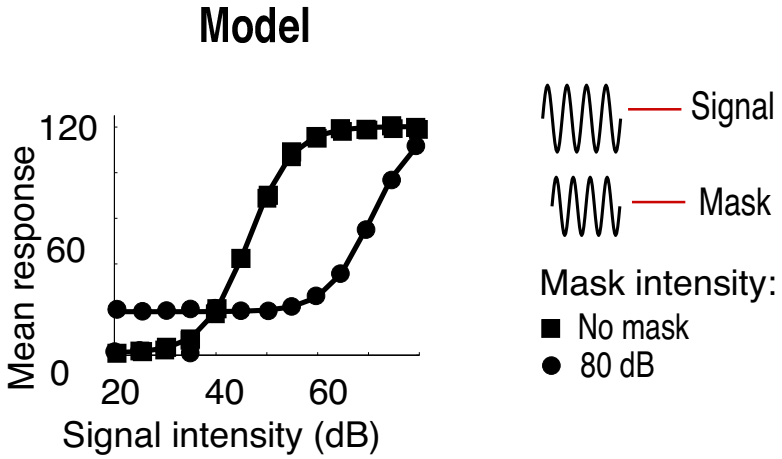

Figure 3.22. Model response for masking stimuli that exerts both excitation and suppression. Masking stimuli is at 1.15 times the optimal frequency. Follows figure 3.21 experiment, but neural data wasn't available for this mask frequency. Rate-level curve with mask is result of both excitatory and suppressive effect exerted by the mask stimulus. Similar auditory nerve fiber behaviors reported in [46].

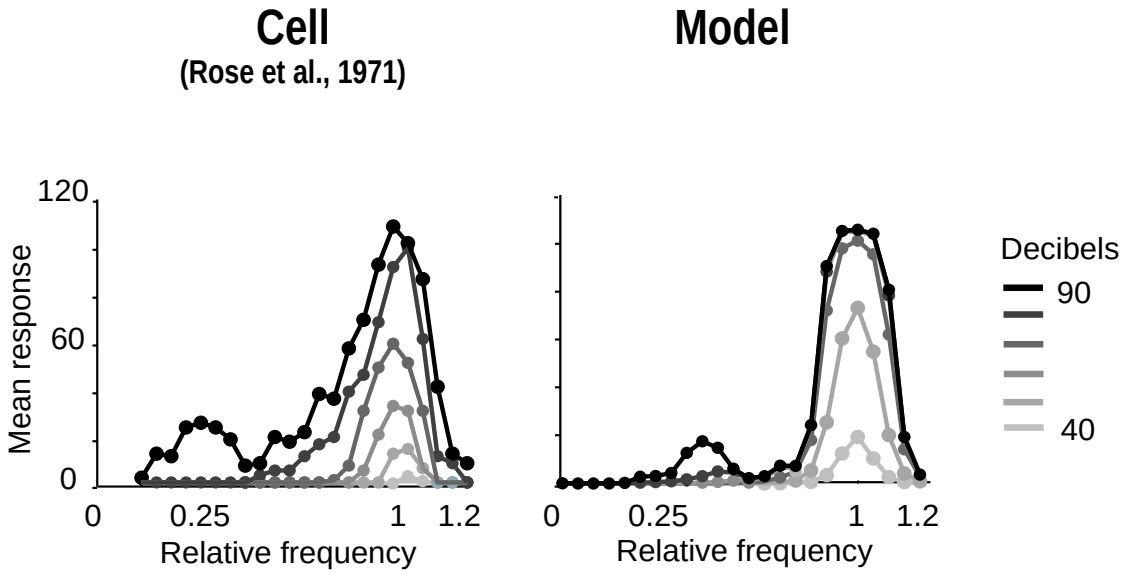

Figure 3.23. Nonlinear changes in frequency tuning curves as a function of sound pressure level in auditory nerve fiber. Physiological data from [120]. Tuning curve broadens and saturates at high levels. Maximum model response has been rescaled to match that of the cell.

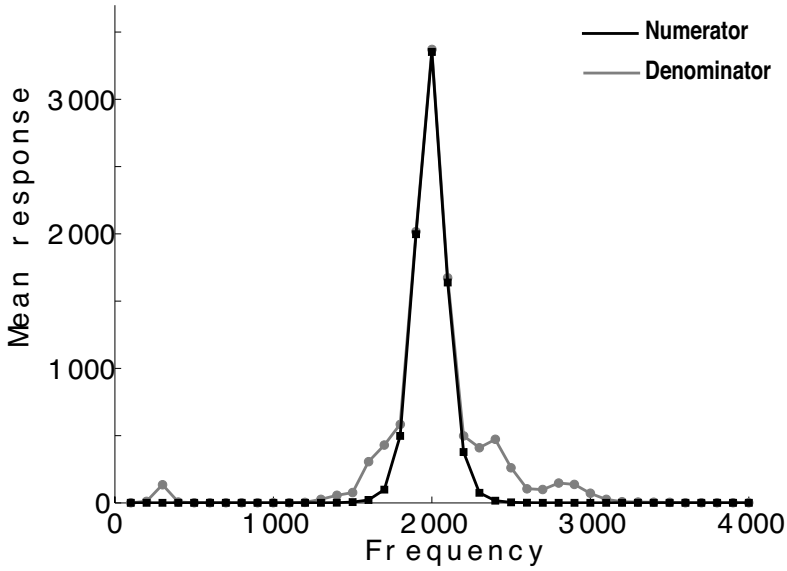

Figure 3.24. Examination of model components for nonlinear changes in frequency tuning curves as a function of sound pressure level. Shown are model components leading to overall response in figure 3.7. Specifically, plotted are the numerator component (squared response of excitatory filter) and denominator component (sum of squared responses of the neighboring filters, with constant σ omitted in the plot). Frequency axis is extended beyond frequencies available in the data.

times the optimal frequency. The rate-level curve produced with signal and mask is now significantly above zero until about 50 decibels, and then takes on the usual sigmoidal shape. This behavior occurs because the mask (which is fixed at 80 dB) is eliciting an excitatory response in the primary filter, even when the signal is eliciting a relatively low response. Thus the mask is exerting both an excitatory and a suppressive effect. In previous examples in figure 3.21, the mask primarily contributed to suppression but not to excitation. This form of excitatory masking behavior at frequencies close to the optimal has been observed in auditory nerve fibers, and denoted "Line-busy masking" by Delgutte [46].

Figure 3.23 demonstrates changes in tuning curve shape for a typical cell in the auditory nerve fiber of a squirrel monkey [120]. Responses are plotted as a function of frequency, for a number of different sound pressure levels. As the sound pressure level increases, the frequency tuning becomes broader, developing a "shoulder" and a secondary mode. Both cell and model show similar behavior, despite the fact that we have not adjusted the parameters to fit these data: all weights in the model are chosen by optimizing the independence of the responses to the ensemble of natural sounds.

The model behavior arises because the weighted normalization signal is frequency-dependent. At low input levels, this frequency dependence is inconsequential because the additive constant dominates the signal. But at high input levels, this frequency dependence will modulate the shape of the frequency tuning curve that is primarily established by the numerator kernel of the model.

To understand this behavior better, we plot the two components of the gain control model, the squared excitatory kernel response and the suppressive sum of

weighted squared responses of the denominator (figure 3.25). As expected, the suppressive response is stronger at frequencies close to the optimal than at frequencies farther away. This effectively serves to broaden the frequency tuning curve. In Figure 3.23, the high contrast secondary mode corresponds to frequency bands with minimal normalization weighting. That is, even though the excitation level is also low, for minimal values of σ, the ratio of the numerator and denominator still yields a significantly high value.

3.2.3 Cell and Model Variations

As in the visual case, there are individual differences in responses of auditory nerve fibers. One of the striking differences amongst individual responses is that auditory nerve fibers are not frequency invariant. For example, when the center frequency is relatively low, the fibers respond to high amplitudes by widening their tuning sensitivity towards higher frequencies. This is different from the example in Figure 3.23, in which the widening of tuning curve sensitivity was towards lower frequencies. In some neurons, for mid-frequencies, the widening of tuning curves is symmetric about the preferred frequency.

We simulate a new model neuron with a low center frequency (400 Hz). The neighborhood consists of 18 frequency bands from 100Hz to 3300Hz and replicas of all filters temporally shifted by 100, 200, and 300 samples. The optimization procedure is performed over the same sound ensemble as in the previous examples.

Figure 3.25 demonstrates changes in tuning curve shape for a nerve fiber of a squirrel monkey [120] with a low center frequency (200 Hz). As the sound pressure level increases, the frequency tuning of the nerve fiber becomes broader towards

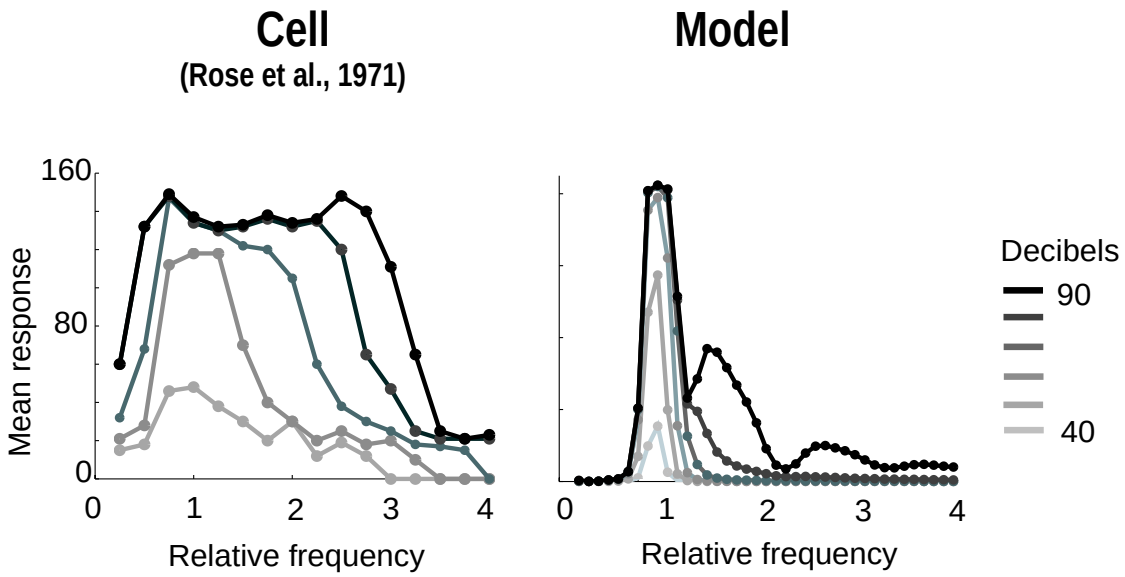

Figure 3.25. Nonlinear changes in frequency tuning curves as a function of sound pressure level low center frequency in auditory nerve fiber. Physiological data from [120]. Tuning curve broadens towards high frequencies and saturates at high levels. Maximum model response has been rescaled to match that of the cell. Model tuning also includes 40 dB, which was not measured in the cell.

high frequencies. Our model neuron does not behave the same in detail; the amount of broadening in the auditory nerve fiber is significantly larger than in the model. However, qualitatively we also obtain a broadening towards higher frequencies. The model also yields a secondary shoulder to the right, similar to Figure 3.23, but in the opposite direction. The direction of expansion in the model appears to be mostly a property of the weights–the tuning curve broadens towards frequency bands with small weighting. However, for very high or very low frequencies, ceiling effects can also come into play.

A more detailed population study is needed to quantify individual difference between nerve fibers and model neuron behavior. We think our model does not exhibit strong enough broadening of tuning curves, because the self-suppressive frequency component is underestimated. We chose discrete time shifts in the optimization, such that the time-shifted filters are roughly uncorrelated with the primary filter. This is again a limitation of the optimization procedure.

3.3 Discussion

We have described a phenomenological nonlinear model for early sensory processing, in which linear responses are squared and then divided by a gain control signal computed as a weighted sum of the squared linear responses of neighboring neurons and a constant. The form of this model is chosen to reduce the type of variance dependencies that we have observed between responses of pairs of linear receptive fields to natural signals (chapter 2). The parameters of the model (in particular, the weights used to compute the gain control signal) are chosen to maximize the

independence of responses to a particular set of signals.

We have demonstrated that the resulting model accounts qualitatively for a range of sensory nonlinearities in "typical" cells. The model can account for nonlinearities in two different modalities, suggesting a canonical neural mechanism for eliminating a type of statistical dependency prevalent in typical natural signals.

Throughout the chapter, we have examined a number of different neural behaviors. We have shown examples of saturation of response curves and rightward shifts of response curves on a logarithmic scale due to the presence of a suppressive mask (figures 3.5, 3.10, 3.21). The amount of rightward shift was determined in the model by the strength of the divisive weights corresponding to the filters that responded to the mask stimulus.

We have also demonstrated rough constancy of tuning curves (figure 3.7) versus changes in tuning curves as a function of stimulus strength (figures 3.16, 3.23). The ability of our model to account for nonlinear changes in tuning curve shape at different levels of input is particularly interesting. Such behaviors have been generally interpreted to mean that the fundamental tuning properties of cells depend on the strength of the input signal. But in our model, the fundamental tuning properties are determined by a fixed linear receptive field, and are modulated by a divisive gain control signal with its own tuning properties. Although such behaviors may appear to be artifacts, our model suggests that they occur naturally in a system that is optimized for statistical independence over natural signals.

Additionally, we have examined what appear to be more complex neural behaviors, arising in the model from interactions of excitation and suppression (figures 3.6, 3.12, 3.22). In particular, if the excitatory and suppressive components

have some overlap in the model, then these components can interact to produce rather interesting behaviors.

The model does not account for the data on a quantitative level. For example, some of the effects we describe are weaker in the model neuron than in the cell. This could arise from numerous factors, including the specific choice of optimization procedure which does not incorporate self-suppression (chapter 2 discussion); the choice of filters and filter neighborhood; the rectification power which was fixed at 2; and the stimulus ensemble which we optimized over. Some of the more technical issues (such as optimization procedure) are also current topics of research in engineering (e.g., [111, 165]). Additionally, there are quantitative variations among individual cells, as illustrated in figure 3.17. For these reasons, we have attempted to model qualitative behaviors that are rather typical in area V1 and in the auditory nerve. We have also experimented with variations of the visual model using different filters, neighborhoods, and optimization procedures, and these have yielded similar results qualitatively [144, 129, 163]. Improvements in the optimization technique together with a more detailed population study might generate more quantitative comparisons between model and neural behavior.

The concept of gain control has been used previously to explain nonlinear behaviors of neurons. A number of auditory models have incorporated explicit gain control mechanisms [84, 167, 59]. Visual models based on divisive normalization have been developed to explain nonlinear effects in cortical area V1 within the classical receptive field [64, 30]. The standard model assumes the response of each neuron is divided by an equally weighted sum of all other neurons and an additive constant. But data in the surround is not matched to an equally weighted normal-

ization model. Cavanaugh et al. have shown that their surround data can still be well fit to a divisive gain control formulation, with parameters that are allowed to vary as a function of stimulus attributes such as orientation and spatial position [33]. Our model utilizes a weighted sum for the normalization signal, and for the first time demonstrates behaviors both inside and outside the classical receptive field.

Our current model provides a functional description, and does not specify the circuitry or biophysics by which these functions are implemented. The divisive normalization computation is performed instantaneously and we have only modeled mean firing rates. Divisive gain control behavior could potentially arise through a number of mechanisms. For example, feedforward synaptic depression mechanisms have been documented and have been shown to exhibit gain control properties (e.g., [1]). Although such mechanisms may account for suppressive V1 behaviors within the classical receptive field, they seem unlikely to account for behaviors outside the classical receptive field.

It has also been proposed that divisive gain control could result from shunting inhibition driven by other neurons (e.g., [29]). In a shunting inhibition model, the linear stage feeds into a circuit composed of a resistor and capacitor in parallel. The divisive pool increases conductance of the resistor without introducing any current at rest. This effectively results in a divisive gain control operation. It has been controversial whether such increases in conductance actually occur in neurons. But recently, a number of studies have reported increases in conductance in V1 cells [26, 66, 4]. Interestingly, such increases in conductance are well tuned for orientation, and are maximal at the preferred orientation of the cell [4]. These

findings offer support to a divisive normalization model that is tuned for orientation, consistent with the predominant surround suppression observed at the preferred orientation of neurons in area V1. This type of implementation necessarily involves recursive lateral or feedback connections, thus introducing temporal dynamics (see [33] for the possible roles of feedback from higher cortical areas versus horizontal connections in determining surround behaviors). Some researchers have described recurrent models that can produce steady-state responses consistent with divisive normalization in area V1 [64, 30, 91].

In the peripheral auditory system, outer hair cells have been implicated in gain control [59]. The outer hair cells provide energy to traveling waves in the cochlear partition. Gain control arises when the outer hair cell receptor potential is pushed into its compressive regions. This in turn produces less energy into the incoming signal, resulting in suppression as a function of stimulus strength. For stimulus frequencies below the characteristic frequency, this process can occur as the traveling wave passes through the characteristic region corresponding to the outer hair cell. For stimulus frequencies above the characteristic frequency, the traveling wave cannot physically reach the characteristic location, but presumably neighboring outer hair cells can still change the gain in the outer hair cell amplification [173]. It is also believed that outer hair cells cannot account for the full extent of two-tone suppression observed in auditory nerve, suggesting the possibility that additional mechanisms are in play [59].

Some of the gain control behaviors we describe may be attributed to earlier stages of neural processing. Gain control is known to occur at the level of the retina [134, 158], although selectivity for orientation does not arise before cortical

area V1. In fact, Ruderman and Bialek [122] have suggested division by local contrast as a means of maximizing marginal entropy, thus providing a functional explanation for gain control in the retina. Our work differs conceptually in the choice of statistical criteria (independence between filters, as opposed to marginal statistics of one filter). As stated earlier, the behaviors we describe at the level of the auditory nerve are largely attributed to outer hair cells, and similar behaviors have been documented in recordings from basilar membrane.

Our model is based on a mechanism that is fundamentally suppressive, but a number of authors have reported facilitative influences in vision from outside the classical receptive field [87, 96, 61, 139]. Many of these facilitative effects might be explained by the use of masking stimuli that inadvertently excite the receptive field of the neuron (e.g., [43, 33]). We have encountered this issue throughout the chapter: if center and surround mechanism are not well separated, then stimuli placed in the surround can cause excitation of the numerator kernel in the divisive normalization formulation. This is similar to the idea that masks inside the classical receptive field with tuning close to the optimal can exert both excitation and suppression (see [97] and figure 3.6); and masks in auditory nerve data with tuning close to the optimal can exert both excitation and suppression ([46] and figure 3.22). Facilitative effects might also be explained by dis-inhibition, in which a third cell inhibits a second cell, thus releasing its inhibition of the recorded cell. As mentioned above, our current model does not utilize a recurrent implementation and thus cannot predict such effects.

There are a number of important directions for further refinement of the connection between natural signal statistics and neuronal processing. We have opti-

mized our model for a generic signal ensemble. The ensemble is small compared to the signals an animal comes across over the course of evolution and development. Furthermore, the signals are influenced by the biases of the photographer, the researcher, and so on. Sensory neurons may have evolved to represent a broad class of signals, or they might be specialized for processing particular classes of signals that are important to an animal's survival. For example, Rieke et al. have shown that auditory nerve fibers in bullfrogs are much more suitable for coding noise shaped to have a naturalistic amplitude spectrum of frog calls than broadband white noise. Additionally, the specific sensory environment in which an animal is reared may influence the neural representation. From a theoretical point of view, different classes of signals may exhibit different statistical properties. Lewicki has demonstrated that in an ICA analysis of natural sounds, the resulting linear filters vary when optimized over animal sounds, human speech, or environmental sounds [82]. A mix of the above result in filters that resemble the gammatone representation.

In addition, we have lumped evolution and development into one fixed process. But it is more likely that adjustment to signal statistics would occur over a number of associated time scales, including evolution, development, learning, and multiple adaptation processes. For example, some visual adaptation effects have been explained by adjusting model parameters according to the statistical properties of recent visual input [17, 162]. Specifically, Wainwright et al. have modeled adaptation with a divisive normalization model similar to ours. The weights are first determined generically for an ensemble of images, but these are then adjusted according to the statistics of the adapting stimuli [162]. Carandini et al. have shown that adaptation to a plaid containing both the optimally oriented grating and a per-

pendicular grating could increase the amount of cross-orientation suppression in V1 neurons [31]. Such a result is also consistent with an adaptive model, in which the increased statistical dependency in the responses of orthogonally oriented neurons would lead to an increase in the suppressive weights between them. Future work should be aimed at testing more directly how variations in the statistical properties of signals effect variations in the response properties of a neuron. This requires construction of stochastic stimuli that vary across specific statistical properties.

A more complete theory also requires an understanding of which groups of neurons are optimized for independence. A sensible assumption might be that each stage of processing in the system takes the responses of the previous stage and attempts to eliminate as much statistical redundancy as possible, within the limits of its computational capabilities. It remains to be seen how much of sensory processing can be explained using such a bottom-up criterion. Ultimately, it seems that one must also consider the specific tasks, such as detection, recognition and localization, that the organism must perform.

The relationship between the model and perception should also be explored. For example, psychophysical experiments suggest that visual detectability is enhanced along contours [53]. At first glance, this might appear to be inconsistent with our model, in which neurons that lie along contours will suppress each other. But the apparent contradiction is based on the unsubstantiated intuition that a reduction in the neural responses implies reduced detectability. Presumably, *any* difference in relative activity of neurons along the contour, as compared with the activity of neurons in other regions, could be used for contour detection. More generally, examination of the implications of our model for perception requires a method of

extracting a percept from a population of neural responses. While this has not been done for contour detection, we find it encouraging that other basic percepts have been explained in the context of a population of neurons performing gain control (e.g. detectability of a grating in the presence of a mask [54]; perceptual segregation of visual textures [62]).

Finally, the work presented in the chapter has highlighted some of the difficulties involved in disentangling the excitatory and suppressive components in neural processing. This issue is commonly faced by physiologists. In some cases, excitation and suppression appear to be overlapping in space or in time, as in the V1 experiments inside the classical receptive field and in the auditory nerve examples. The experiments outside the classical receptive field that we utilized in the chapter (and developed by [33]) were constructed to overcome this problem: there is a significantly better spatial separation of excitation and suppression. Even then, excitation and suppression are not entirely mutually exclusive. The recent experimental results illustrated in figure 3.16 suggest that it could be more correct to estimate the receptive field at low contrasts. However, contrasts that are too low may lead to noisy measurements.

More generally, it would be useful to separate out excitation and suppression, even when these components largely overlap. This calls for the difficult task of developing computational techniques for characterizing excitation and suppression in a neuron. In chapter 4, we revisit this issue in the context of a white noise methodology and gain control model.

CHAPTER 4

SPIKE-TRIGGERED COVARIANCE CHARACTERIZATION

What kind of representations do sensory neurons form and how can we characterize these representations quantitatively? To answer this question, one can design experiments with stimuli that vary along specific parameters, and propose a model that well explains the neural data. An alternative route, often referred to as white noise analysis, is to present a neuron with random input that explore a wider stimulus space, and then examine the subset of stimuli that elicit spikes. White noise methods offer a quantitative framework for analyzing this "spike-triggered" stimulus ensemble. The analysis can provide information about a neuron's response characteristics, typically assuming an underlying class of neural models. Ultimately, one would like to come up with characterizations that are generically useful for describing the response properties of neurons.

Probably the most widely used form of neural characterization is the linear receptive field. In the context of white noise analysis, one estimates an excitatory linear kernel by computing the spike-triggered average (STA); that is, the mean

stimulus that elicited a spike (e.g., [45, 74]). Under the assumption that spikes are generated by a Poisson process with instantaneous rate determined by linear projection onto a kernel followed by a static nonlinearity, the STA provides an unbiased estimate of this kernel [36].

The STA is limited to finding a single axis representing the neuron. But we would like a more general methodology for describing a subspace of axes that govern the response of a neuron. Recently a number of authors have developed an interesting extension to white noise analysis for recovering subspaces [42, 5]. For example, an excitatory subspace is recovered by estimating the covariance matrix associated with the spike-triggered stimulus ensemble (i.e., the spike-triggered covariance or STC), followed by an eigenvector analysis to determine axes of maximal variance that presumably drive the cell the most. The analysis does not specify a functional form for how the retrieved axes operate together. So although covariance is a linear operation, the axes may interact *nonlinearly*. Thus, STC offers a general framework for recovering a subspace, and potentially exploring neuronal nonlinearities within the subspace.

Sensory neurons exhibit striking nonlinear behaviors that are not explained by fundamentally linear mechanisms. For example, the response of a neuron typically saturates for large amplitude stimuli; the response to the optimal stimulus is often suppressed by the presence of a non-optimal mask (e.g., [30]); and the kernel recovered from STA analysis may change shape as a function of stimulus amplitude (e.g., [135, 35]). The third behavior signifies that STA analysis alone is often not a sufficient model for representing the response of a sensory neuron.

A variety of these nonlinear behaviors can be attributed to gain control (e.g.,

[135, 134, 84, 58, 64]), in which neural responses are suppressively modulated by a gain signal derived from the stimulus. Although the underlying mechanisms and time scales associated with such gain control are current topics of research, the basic functional properties appear to be ubiquitous, occurring throughout the nervous system.

Here we develop a white noise methodology for characterizing a neuron with gain control. We show that a set of suppressive kernels may be recovered by finding the eigenvectors of the spike-triggered covariance matrix associated with *smallest* variance. We apply the technique to electrophysiological data obtained from ganglion cells in salamander and monkey retina, and recover a set of axes that are shown to reduce responses in the neuron. Moreover, when we fit a gain control model to the data using a maximum likelihood procedure within this subspace, the model accounts for changes in the STA as a function of contrast.

4.1 Characterizing a Neuron with Gain Control

As in all white noise approaches, we assume that stimuli correspond to vectors, $\vec{s}$, in some finite-dimensional space (e.g., a neighborhood of pixels or an interval of time samples). We assume a gain control model in which the probability of a stimulus eliciting a spike grows monotonically with the halfwave-rectified projection onto an excitatory linear kernel, $\lfloor \vec{k}_0 \cdot \vec{s} \rfloor$, and is suppressively modulated by the fullwave-rectified projection onto a set of linear kernels, $|\vec{k}_n \cdot \vec{s}|$.

The method proceeds in three steps: (1) Estimating excitatory kernel with STA; (2) Estimating suppressive subspace with STC; and (3) Fitting specific gain control

model to data using maximum likelihood in the reduced subspace.

This approach can be generalized to cases in which the STA is approximately zero (such as in complex cells), but in which the STC yields excitatory axes (see extensions section). In addition, even when the STA exists, there may be more than one excitatory kernel governing the neural response. Here we focus primarily on cases in which the STA is meaningful, with the aim of understanding how the STA works in the context of a suppressive subspace. This will also be relevant for the retinal ganglion cell data in the chapter.

4.2 Estimating Excitatory Kernel with STA

First, we recover the excitatory kernel, $\vec{k}_0$. This is achieved by presenting spherically symmetric input stimuli (e.g., Gaussian white noise) to the neuron and computing the STA. Figure 4.1 provides a geometric depiction of the STA analysis. The analysis is depicted along only two dimensions. This can be interpreted as corresponding, for example, to stimulus intensity at two different times. Alternatively, we may interpret the axes to correspond to the projection of the stimuli onto two abstract kernels within a high dimensional space.

STA correctly recovers the excitatory kernel, under the assumption that each of the suppressive kernels are orthogonal (or equal) to the excitatory kernel. The proof is essentially the same as that given for recovering the kernel of a linear model followed by a monotonic nonlinearity [36]. In particular, any stimulus can be decomposed into a component in the direction of the excitatory kernel, and a component in a perpendicular direction. This can be paired with another stimu-

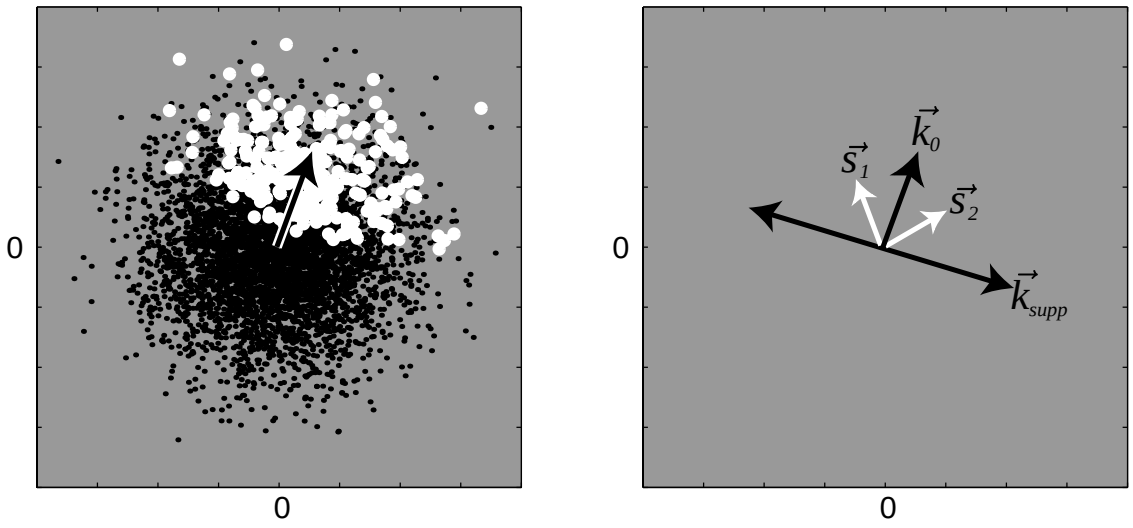

Figure 4.1. **a**, Geometric depiction of spike-triggered averaging with two-dimensional stimuli. Black points indicate raw stimuli. White points indicate stimuli eliciting a spike, and the STA (black vector), which provides an estimate of $\vec{k}_0$, corresponds to their center of mass. **b**, Based on symmetry argument, STA correctly recovers excitatory kernel if suppressive kernels are orthogonal or equal to excitatory kernel. Plotted are two stimulus vectors symmetric about the excitatory kernel (white). See main text for sketch of proof.

lus that is identical, except that its component in the perpendicular direction is negated. The two stimuli are equally likely to occur in a spherically Gaussian stimulus set (since they are equidistant from the origin), and they are equally likely to elicit a spike (since their excitatory components are equal, and their rectified perpendicular components are equal). Their vector average lies in the direction of the excitatory kernel. Thus, the STA (which is an average over all such stimuli, or all such stimulus pairs) must also lie in that direction. In a subsequent section we explain how to recover the excitatory kernel when it is not orthogonal to the suppressive kernels.

4.3 Estimating Suppressive Subspace with STC

Next, we recover the suppressive subspace, assuming the excitatory kernel is known. Consider the stimuli lying on a plane perpendicular to this kernel. These stimuli all elicit the same response in the excitatory kernel, but they may produce different amounts of suppression. Figure 4.2 illustrates the behavior in a three-dimensional stimulus space, in which one axis is the excitatory kernel, and the two-dimensional plane represents a subspace perpendicular to the excitatory kernel. Within this plane, one axis is assumed to be suppressive. The distribution of raw stimuli on the plane is spherically symmetric about the origin (because the raw stimuli are spherically symmetric, and thus remain so when projected onto the two axes of the plane). But for stimuli eliciting a spike, any distribution that deviates from spherical symmetry is potentially indicative of the response properties of the neuron. We are specifically interested in the scenario plotted in the figure, in which the

distribution of stimuli eliciting a spike is elliptical and narrower (i.e., lower variance) along one of the axes. The narrow axis corresponds to a suppressive direction: stimuli that have a large component along this direction are less likely to elicit a spike.

Stated differently, we limit the problem of finding suppressive axes, to that of finding axes in which the variance of the response of a neuron along those axes is reduced (relative to the variance of random spherical stimuli). We prove in the end of the chapter that if the response of a neuron is given by a projection onto a subspace followed by a general nonlinearity (e.g., gain control), then for a stimulus direction outside this subspace, the variance of the response of the neuron will be equal to that of the random spherical stimuli. Thus, axes that have variance below what is expected from random stimuli, are truly indicative of the response properties of the neuron.

This behavior is easily generalized from the plane in the figure to the entire stimulus space. If we assume that the suppressive axes are fixed, then we expect to see reductions in variance in the same directions for any level of numerator excitation. In practice, we do not force the stimuli to elicit constant response in the excitatory kernel, resulting in perpendicular planes at different distances from the excitatory kernel, but all with the same axes.

Given this behavior of the spike-triggered stimulus ensemble, we can recover the suppressive subspace using principal component analysis. We construct the sample covariance matrix of the stimuli eliciting a spike:

$$C = \frac{1}{N_s - 1} \sum_{\vec{s}_{\text{spike}}} \vec{s}\ \vec{s}^T, \tag{4.1}$$

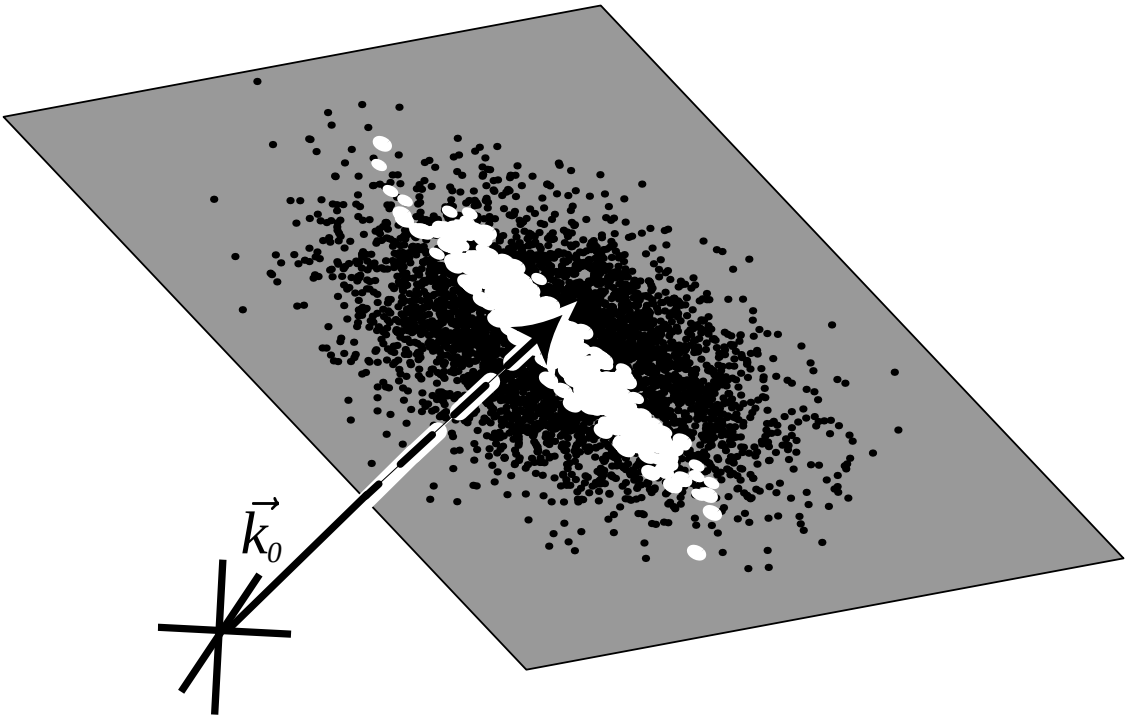

Figure 4.2. Geometric depiction of spike-triggered covariance analysis of suppressive axes. Shown are a set of stimuli lying on a plane perpendicular to the excitatory kernel, $\vec{k}_0$. Black points indicate raw stimuli. White points indicate stimuli eliciting a spike. Within the plane, stimuli eliciting a spike are concentrated in an elliptical region. The minor axis of the ellipse corresponds to a suppressive stimulus direction: stimuli with a significant component along this axis are less likely to elicit spikes. The stimulus component along the major axis of the ellipse has no influence on spiking.

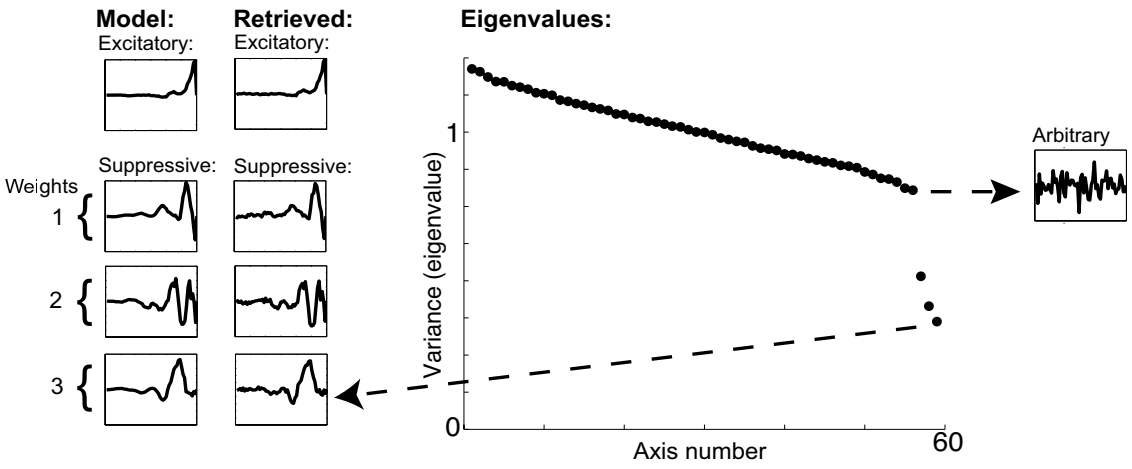

Figure 4.3. Estimation of kernels from a simulated model (equation 4.2). Left: Model kernels. Right: Sorted eigenvalues of covariance matrix of stimuli eliciting spikes (STC). Three eigenvalues fall significantly below the others. Middle: STA (excitatory kernel) and eigenvectors (suppressive kernels) associated with the lowest eigenvalues.

where N_s is the number of spikes. To ensure the estimated suppressive subspace is orthogonal to the estimated $\vec{k}_0$ (as in Figure 4.2), the stimuli $\vec{s}_{\text{spike}}$ are first projected onto the subspace perpendicular to the estimated $\vec{k}_0$. The principal axes (eigenvectors) of C that are associated with small variance (eigenvalues) correspond to directions in which the response of the neuron is modulated suppressively.

4.4 Estimating Suppressive Subspace in Simulation

We illustrate the technique on simulated data. The goal of simulation is to probe the temporal response properties of a model neuron by presenting full-field white noise stimuli at every time frame. The simulation is easily extendible to space-time subspaces (using bars for one dimension of space; or pixels for two dimensions

of space). However, the temporal example serves as a reference for most of the physiological data in the chapter.

The model neuron consists of a temporal receptive field, and a suppressive subspace made up of three orthogonal temporal kernels. The kernels are randomly chosen, and then filtered such that more emphasis is placed on recent time (i.e., to make them look more realistic). Each kernel is restricted to a 60 dimensional space. Thus, in the analysis, stimuli preceding a spike are taken over a 60-segment time window.

Spikes are generated using a Poisson process with mean rate determined by a specific form of gain control:

$$\mathcal{P}\left(spike|\vec{s}\right) = \frac{\left\lfloor \vec{k}_0 \cdot \vec{s} \right\rfloor^2}{\sum_n w_n |\vec{k}_n \cdot \vec{s}|^2 + \sigma^2}. \tag{4.2}$$

The goal of simulation is to recover excitatory kernel $\vec{k}_0$, the suppressive subspace spanned by $\vec{k}_n$, the weights w_n, and constant σ.

Figure 4.3 shows the original and estimated kernels for a model simulation with 80K input samples and 8.374K spikes. First, we note that STA recovers an accurate estimate of the excitatory kernel. Next, consider the sorted eigenvalues of C, as plotted in Figure 4.3. The majority of the eigenvalues descend gradually (the covariance matrix of the white noise source should have constant eigenvalues, but remember that those in Figure 4.3 are computed from a finite set of samples). The last three eigenvalues are significantly below the values one would obtain with randomly selected stimulus subsets. The eigenvectors associated with these lowest eigenvalues span approximately the same subspace as the suppressive kernels [1]

[1]Due to the non-uniqueness of the eigenvector decomposition, one can also get out mixtures

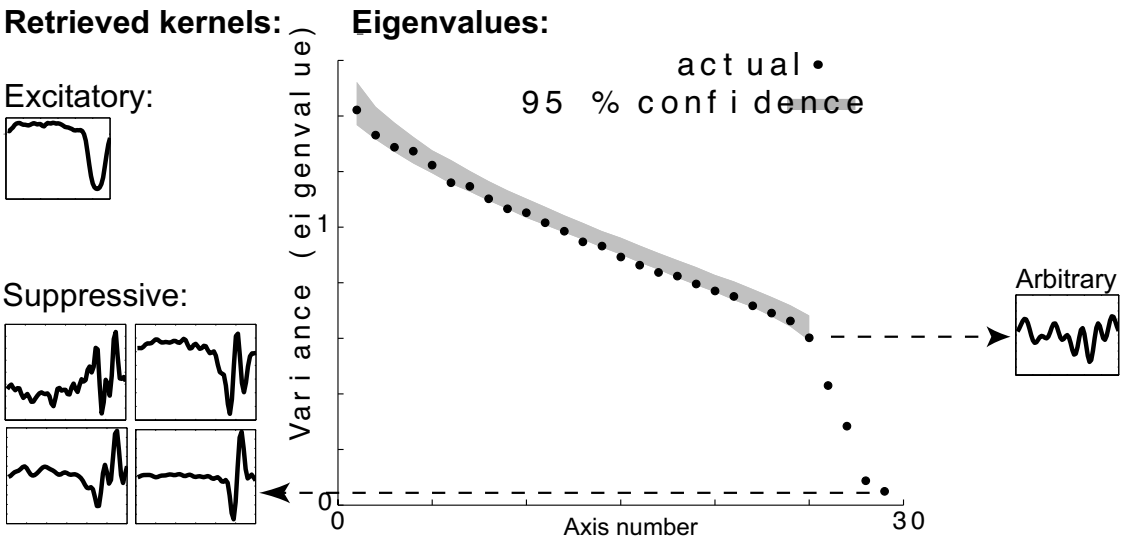

Figure 4.4. Estimation of kernels for a salamander ganglion cell. Left: Retrieved kernels from STA and STC analysis of ganglion cell data from a salamander retina (cell 1999-11-12-B6A). Right: sorted eigenvalues of the spike-triggered covariance matrix, with corresponding eigenvectors. Low eigenvalues correspond to suppressive directions, while other eigenvalues correspond to arbitrary (ignored) directions. Raw stimulus ensemble was whitened prior to analysis and low-variance axes under-represented in stimulus set. Therefore, only 30 out of 60 input dimensions are plotted here.

In contrast, eigenvectors corresponding to eigenvalues in the gradually-descending region appear arbitrary in their structure and cover the full temporal range.

4.5 Estimating Suppressive Subspace in Retinal Ganglion Cells

Retinal ganglion cells exhibit rapid [135, 158] as well as slow [35, 147, 78] gain control; slow gain control has been observed over tens of seconds, and fast gain

of the original suppressive kernels. In the example simulation, weights are uniquely chosen.

control over tens of milliseconds. We demonstrate that we can recover a rapid gain control signal by applying the method to data from retinal ganglion cells.

Data were recorded by Chander and Chichilnisky from a planar array of extracellular microelectrodes [35]. The input sequence in the experiments consists of either full-field 33Hz flickering binary white noise refreshed every two frames (salamander) or full-field 33Hz/66Hz flickering Gaussian white noise refreshed every frame (monkey). Some of the data were recorded with stimuli of a single contrast, and others with multiple contrasts. We define the stimulus vectors in the analysis over a 60-segment time window. Since the raw stimuli are finite in number (and sometimes binary) they are not spherically distributed. To correct for this, we discard low-variance axes and whiten remaining axes. In the Salamander data, half the axes are disregarded, due to the fact that the stimuli are refreshed every two frames.

We first demonstrate the technique on a salamander retinal ganglion cell. The input sequence consists of 80K time samples of full-field 33Hz flickering binary white noise (contrast 8.5%). Figure 4.4 depicts the kernels estimated from the 623 stimulus vectors eliciting spikes. Similar to the model simulation, the eigenvalues gradually fall off, but four of the eigenvalues appear to drop significantly below the rest. To make this more concrete, we test the hypothesis that the majority of the eigenvalues are consistent with those of randomly selected stimulus vectors, but that the last 4 eigenvalues fall significantly below this range. We perform a Monte Carlo simulation, drawing (with replacement) random subsets of 623 stimuli from the full set of raw stimuli. More specifically, since the position of the spikes (and possible stimulus correlations for adjacent spikes) may result in a distribution

slightly different from random, we test the hypothesis that that the majority of the eigenvalues are consistent with those of randomly selected stimulus vectors with the same inter-spike interval. This does not make any difference for the particular cell shown here because the spike rate is so low, but makes a minor difference for other examples in the chapter.

We also randomly select 4 (orthogonal) axes, representing a suppressive subspace, and project this subspace out of the set of randomly chosen stimuli. We then compute the eigenvalues of the sample covariance matrix of these stimuli. We repeat this 1000 times, and estimate a 95 percent confidence interval for each of the eigenvalues. The figure shows that the first eigenvalues lie roughly within the confidence interval, but the 4 lowest eigenvalues drop significantly below. In practice, we repeat this process in a nested fashion, assuming initially no directions are significantly suppressive, then one direction, and so on up to four directions (Figure 4.5).

These low eigenvalues correspond to eigenvectors that are concentrated in recent time (as is the estimated excitatory kernel). The remaining eigenvectors appear to be arbitrary, spanning the full temporal window. We emphasize that these kernels should not be interpreted to correspond to receptive fields of individual neurons underlying the suppressive signal, but merely provide an orthogonal basis for a suppressive subspace. The derivative-like appearance of some of the low-eigenvalue eigenvectors could potentially be explained by the fact that PCA forces the recovered axes to be orthogonal (for example, consider a time-shifted version of the same kernel forced to be orthogonal).

Figures 4.6 and 4.7 illustrate more examples of low eigenvalues and correspond-

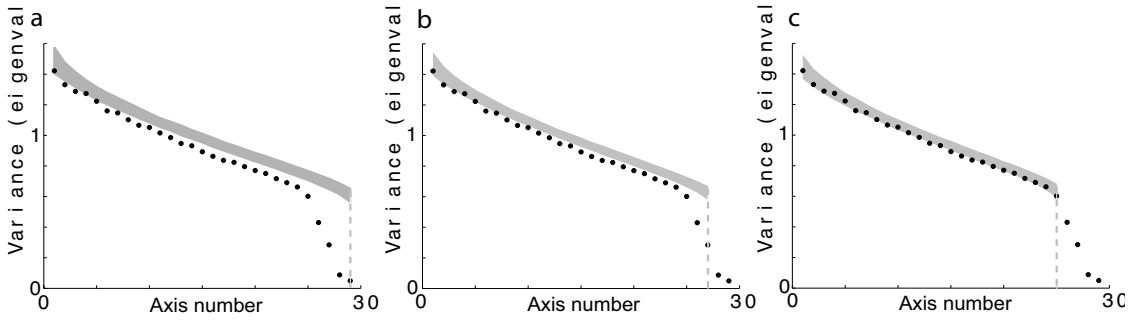

Figure 4.5. Nested Monte Carlo hypothesis testing of number of suppressive axes in ganglion cell data from a salamander retina (1999-11-12-B6A). Gray solid line corresponds to 95 percent confidence interval, assuming 0 suppressive axes (left), 2 suppressive axes (middle), and 4 suppressive axes (right). If the hypothesis is correct, eigenvalues to the left of the dashed line should lie within the confidence interval. For the assumption of 0 or 2 suppressive axes, some eigenvalues to the left of the dashed line lie below the confidence interval, indicating the hypothesis is incorrect. In contrast, for the assumption of 4 suppressive axes, eigenvalues to the left of the dashed line lie roughly within the confidence interval.

ing eigenvectors in salamander (including a summary of the original cell) and monkey cells. These differ in the number of suppressive kernels and their variance (even among the same species), and hence the presumed suppressive strength. The differences in the slopes of the gradually descending eigenvalues (and how tightly they are around 1, and the width of the confidence interval) is due to the number of spikes, which varies from cell to cell. Note that in all cells, the suppressive kernels are concentrated in recent time, sometimes extending farther in time than the STA.

We now verify that the recovered STA axis is in fact excitatory, and the kernels corresponding to the lowest eigenvalues are suppressive. Figure 4.8a shows a scatter plot of the stimuli projected onto the excitatory axis versus an arbitrary axis in the original salamander cell. Spikes are seen to occur only when the component

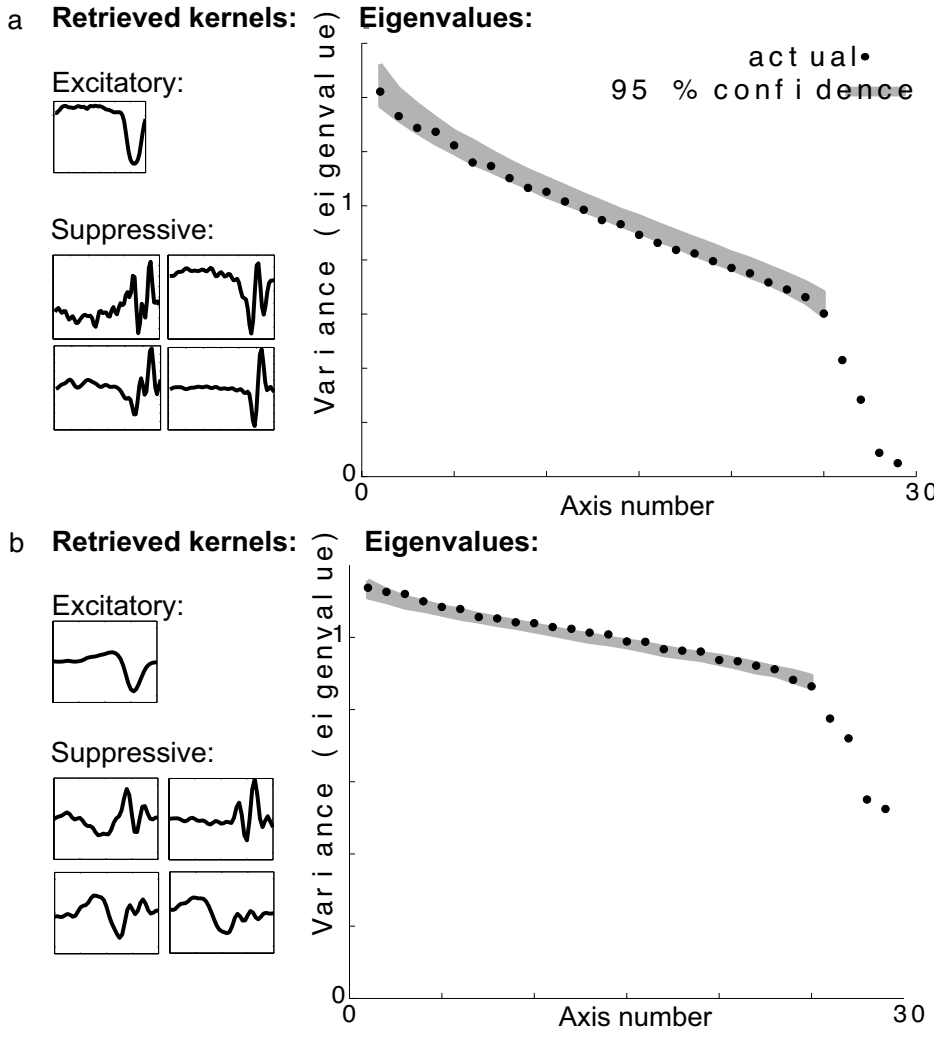

Figure 4.6. Examples of estimated kernels in salamander ganglion cells. Left: Retrieved kernels from STA and STC analysis of ganglion cell data from salamander, with binary white noise stimuli. Right: sorted eigenvalues of the spike-triggered covariance matrix, with corresponding eigenvectors. Low eigenvalues correspond to suppressive directions, while other eigenvalues correspond to arbitrary (ignored) directions. **a**, 1999-11-12-B6A (from previous figures, 623 spikes out of 80K input samples) **b**, 1999-01-19-G2A (8882 spikes/181290 input samples)

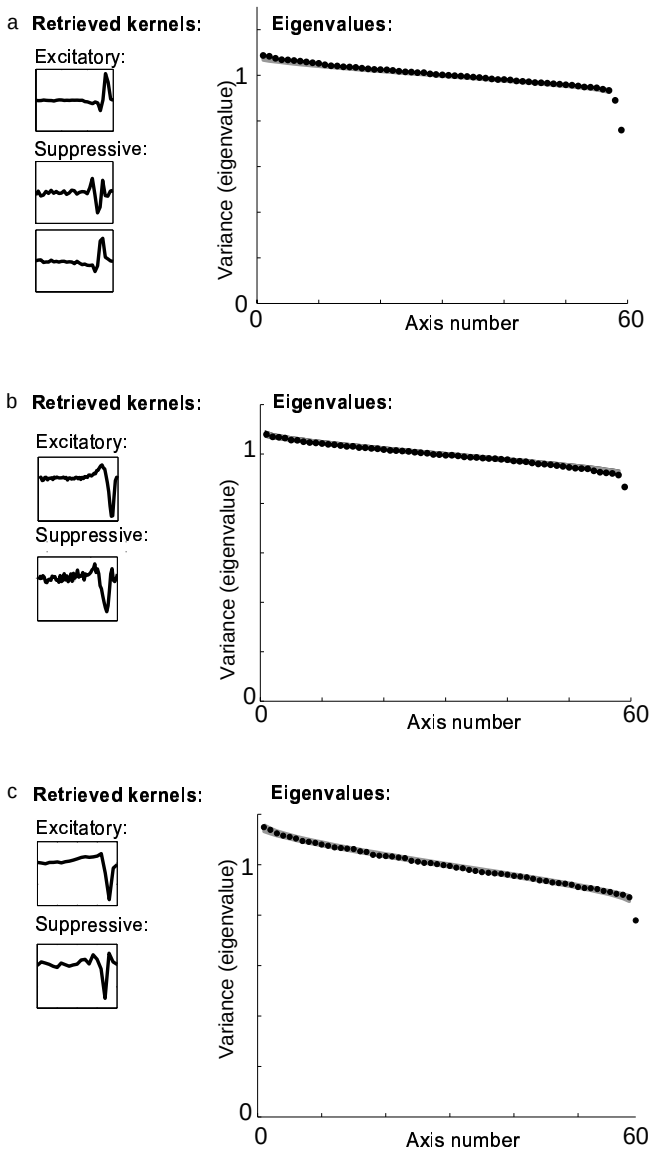

Figure 4.7. Examples of estimated kernels in monkey ganglion cells. Left: Retrieved kernels from STA and STC analysis of ganglion cell data from monkey, with Gaussian white noise stimuli. Right: sorted eigenvalues of the spike-triggered covariance matrix, with corresponding eigenvectors. Low eigenvalues correspond to suppressive directions, while other eigenvalues correspond to arbitrary (ignored) directions. **a**, 2001-09-29-1-E6A (36436 spikes/284745 input samples) **b**, 2001-09-29-0-B5A (24350 spikes/284745 input samples) **c**, 1999-01-28-B1A (17477 spikes/84135 input samples)

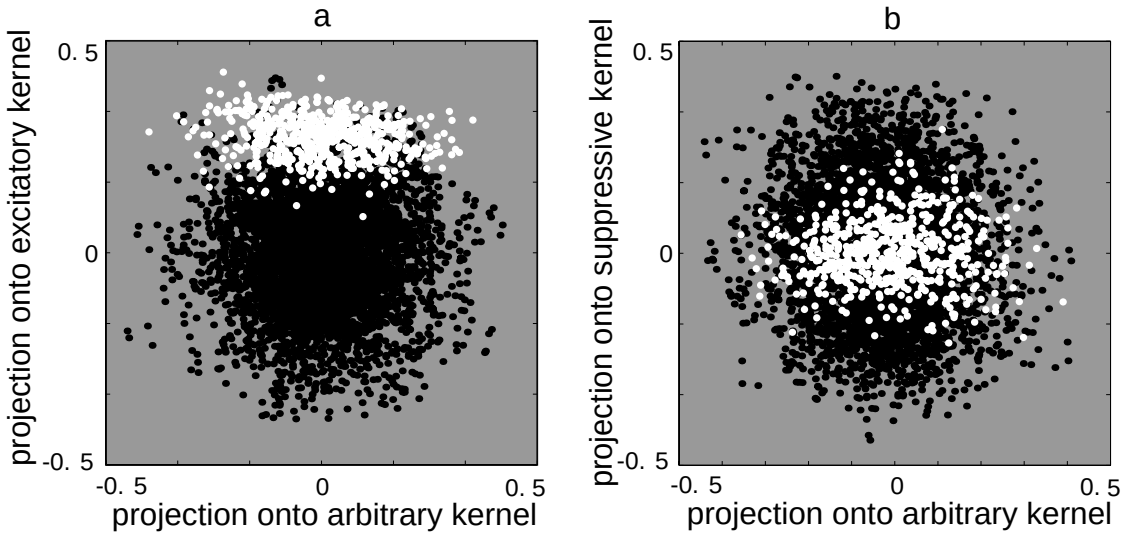

Figure 4.8. Scatter plots of stimuli projected onto estimated kernels for salamander ganglion cell (1999-11-12-B6A). Black points indicate the raw stimulus set. White points indicate stimuli eliciting a spike. **a**, Projection of stimuli onto estimated excitatory kernel vs. arbitrary kernel. **b**, Projection of stimuli onto an estimated suppressive kernel vs. arbitrary kernel.

along the excitatory axis is high, as expected. Figure 4.8b is a scatter plot of the stimuli projected onto one of the suppressive axes vs. an arbitrary (ignored) axis. The spiking stimuli lie within an ellipse, with the minor axis corresponding to the suppressive kernel. This is exactly what we would expect in a suppressive system, such as that plotted in Figure 4.2.

Figure 4.9 illustrates additional examples of scatter plots from monkey cells. The excitatory scatter plots (even within species) differ in the threshold at which the neuron spikes (also seen in salamander); and the suppressive scatter plots differ in the amount of variance reduction of the minor axis of the ellipse.

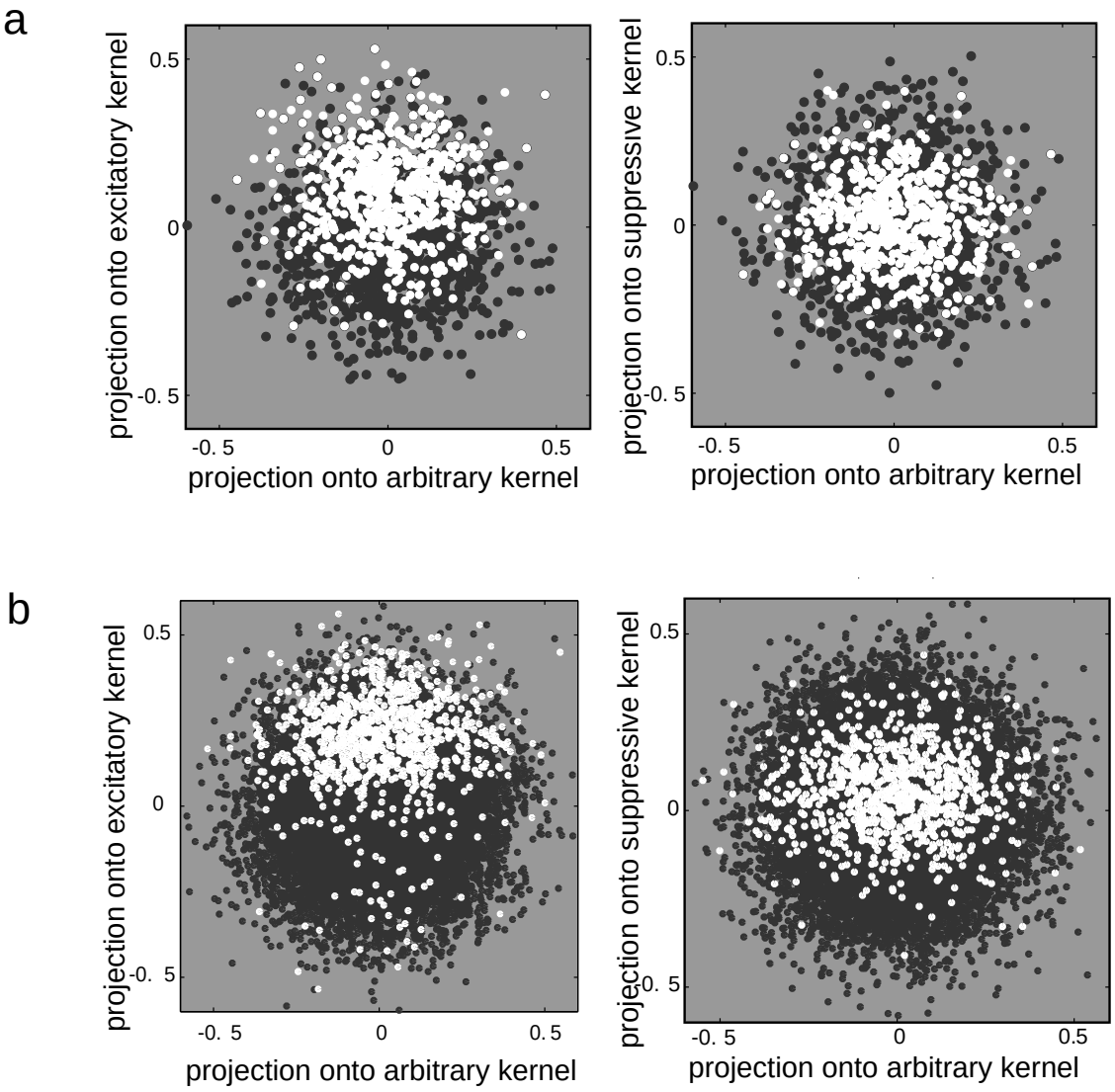

Figure 4.9. Scatter plots of stimuli projected onto estimated kernels for monkey ganglion cells. Top, Monkey retina (cell 2001-09-29-E6A). Bottom, Monkey retina (cell 2001-09-29-0-B5A).

4.6 Fitting Specific Model Within Reduced Subspace

We have recovered an excitatory kernel and suppressive subspace. It is not clear, however, in what way the suppressive kernels effect the excitatory response of the neuron. A reduction in variance (as seen in the eigenvalues and scatter plots) may be caused by a number of mechanisms, including two basic kinds of suppression: subtraction or division. Figures 4.10 and 4.11 illustrate that the suppressive kernels are significantly more divisive than subtractive. This is demonstrated by computing the mean spike count for the excitatory kernel response, conditioned on values of a suppressive kernel response in two bins. The top curve (low values of suppressive kernel response) is better related to the bottom curve (high values of suppressive kernel response) by division than by subtraction. This is consistent with previous notions of gain control in retinal ganglion cells (e.g., [158, 134]).

Based on these observations, we now fit a specific gain control model to the data:

$$\mathcal{P}\left(spike|\vec{s}\right) = \frac{\lfloor(\vec{k}_0 - \alpha)\cdot\vec{s}\rfloor^p}{(\sum_n w_n|\vec{k}_n\cdot\vec{s}|^2)^{p/2} + \sigma^p} \tag{4.3}$$

where the exponent p is incorporated to allow for more realistic contrast-response functions; and α allows for spikes when response values are below zero (see for example monkey cell in Figures 4.9b and 4.10b).

We recover the scalar parameters of this specific model (p, α, w_n and σ) by selecting them to maximize the likelihood of the spike data according to equation (4.3):

$$\{p, \alpha, \vec{w}, \sigma\} = \arg\max_{p,\alpha,\vec{w},\sigma} \prod_k \mathcal{P}\left(spike|\vec{S}_k\right) \prod_j (1 - \mathcal{P}\left(spike|\vec{S}_j\right)) \tag{4.4}$$

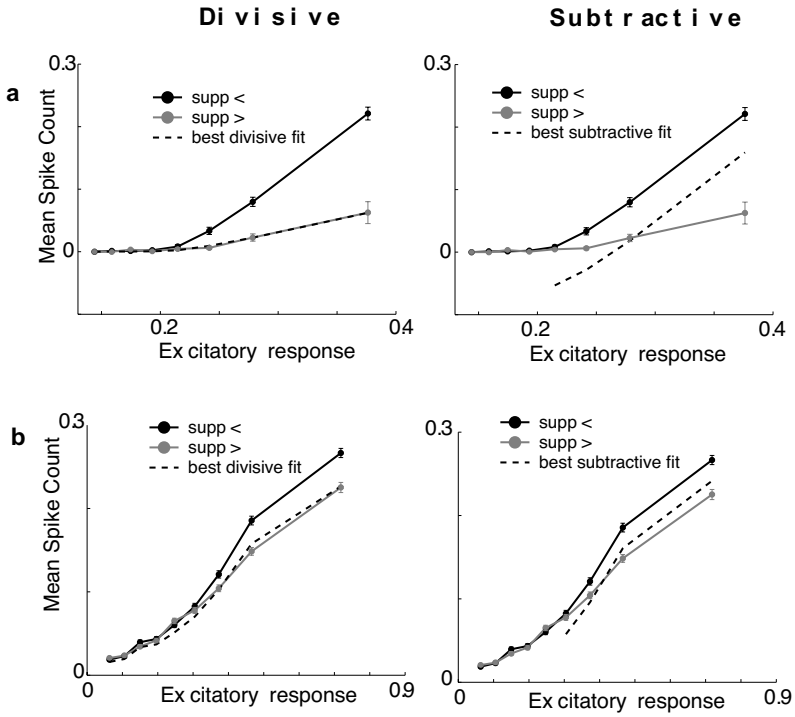

Figure 4.10. Division versus subtraction in salamander retinal ganglion cells. Mean spike count for the excitatory kernel response, conditioned on two different ranges of suppressive kernel response in Salamander cells. Responses of suppressive kernel are divided into two bins of equal number of stimuli. The top curve is fit to the bottom curve by minimizing the mean squared error of a divisive scalar (left) and subtractive scalar (right). In the subtractive model, we only consider points in the two curves whose difference is above a threshold, yielding slightly better fits than a pure subtractive model. In all cells, the divisive fit produces a smaller mean squared error than the subtractive fit. **a**, Salamander (1999-11-12-B6A). Division: scalar 0.2815, mean squared error 2.4422e-05; Subtraction: scalar .0309, mean squared error .0215. **b**, Salamander (1999-01-19-G2A). Division: scalar 0.85, mean squared error 3.9861e-04; Subtraction: scalar 0.0249, mean squared error 0.0041.

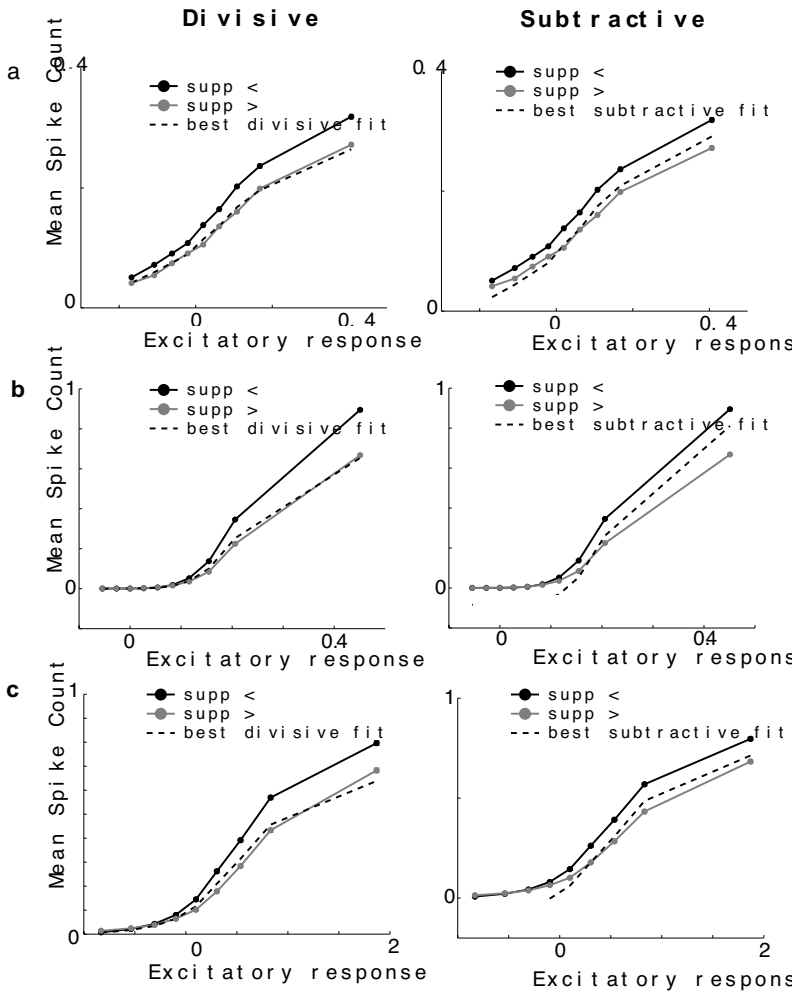

Figure 4.11. Division versus subtraction in monkey retinal ganglion cells. Mean spike count for the excitatory kernel response, conditioned on two different ranges of suppressive kernel response in Monkey cells (analogous to Figure 4.10) **a**, Monkey (2001-09-29-0-B5A) Division: scalar 0.7312, mean squared error 0.0013; Subtraction: scalar .0841, mean squared error .0696. **b**, Monkey (2001-09-29-1-E6A). Division: scalar 0.8289, mean squared error 2.3385e-04; Subtraction: scalar .0275, mean squared error .0014. **c**, Monkey (1999-01-28-B1A) Division: scalar 0.809, mean squared error 0.0047; Subtraction: scalar .0834, mean squared error .0105.

Note that a direct maximum likelihood solution on the raw data would have been impractical due to the high dimensionality of the stimulus space. In the subsequent section, we use this technique to recover a specific model for ganglion cells. But we first draw the attention of the reader to a potential problem with the method described thus far for recovering the kernel subspace.

4.7 Correcting for Bias in Kernel Estimates and Explaining Changes in the STA

The kernels in the previous section were all recovered from stimuli of a single contrast. However, when the STA is computed in a ganglion cell for low and high contrast stimuli, the recovered kernels typically differ in shape ([35], and see figure 4.12). This would appear inconsistent with the method we describe, in which the STA is meant to provide an estimate of a single excitatory kernel. This behavior can be explained by assuming a model of the form given in equation 4.3, and in addition dropping the constraint that the gain control kernels are orthogonal (or identical) to the excitatory kernel.

First we show that when the orthogonality constraint is dropped, the STA estimate of the excitatory kernel is biased by the gain control signal. Consider a situation in which a suppressive kernel contains a component in the direction of the excitatory kernel, $\vec{k}_0$. We write $\vec{k}_n = \alpha\vec{k}_0 + \vec{k}'_n$, where $\vec{k}'_n$ is perpendicular to the excitatory kernel. Then, for example, a stimulus $\vec{s} = \vec{k}_0 + \beta\vec{k}'_n$, with $\beta > 0$, produces a suppressive component along $\vec{k}_n$ equal to $\alpha||\vec{k}_0||^2 + \beta||\vec{k}'_n||^2$, but the

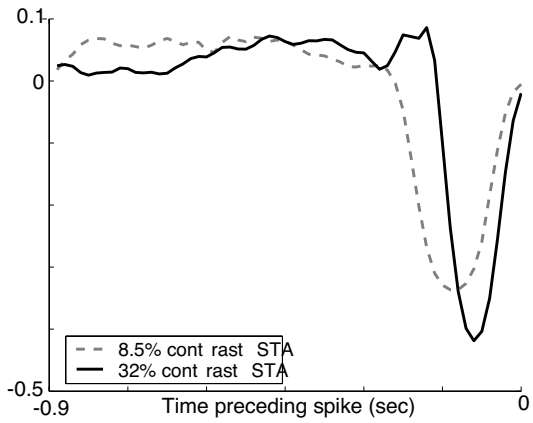

Figure 4.12. Change in STA kernels for different contrasts in salamander retinal ganglion cell. STA kernels derived from salamander ganglion cell spikes for 8.5 percent and 32 percent contrast stimuli (1999-11-12-B6A). Input stimuli are full-field 33Hz flickering binary white noise. Kernels are normalized to unit energy. Note the change in kernel shape as a function of contrast, specifically larger onset latency and slower timecourse at low contrast.

corresponding paired stimulus vector $\vec{s} = \vec{k}_0 - \beta\vec{k}'_n$ produces a suppressive component of $\alpha||\vec{k}_0||^2 - \beta||\vec{k}'_n||^2$. Thus, the two stimuli are equally likely to occur but not equally likely to elicit a spike. As a result, the STA will be biased in the direction $-\vec{k}'_n$. Figure 4.13 illustrates an example in which a non-orthogonal suppressive axis biases the estimate of the STA.

Now consider the model in equation 4.3 in the presence of a non-orthogonal suppressive subspace. Note that the bias is stronger for larger amplitude stimuli because the constant term σ^2 dominates the gain control signal for weak stimuli. Indeed, in chapter 3 we have hypothesized that changes in receptive field tuning can arise from divisive gain control models that include an additive constant.

Since suppressive axes bias the STA kernel estimate away from the true excitatory kernel, one can try and use the full recovered subspace to obtain a better

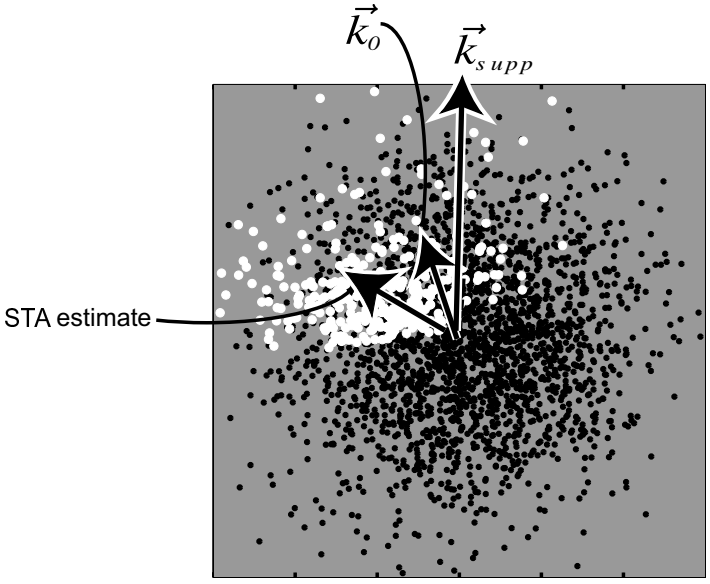

Figure 4.13. Demonstration of estimator bias. When a gain control kernel is not orthogonal to the excitatory kernel, the responses to one side of the excitatory kernel are suppressed more than those on the other side. The resulting STA estimate is thus biased away from the true excitatory kernel, $\vec{k}_0$.

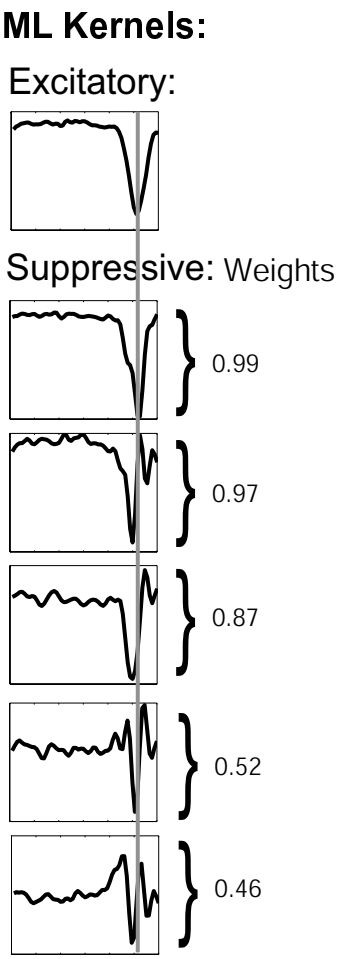

Figure 4.14. ML retrieved model for salamander cell (1999-11-12-B6A). Remaining model parameters are: $p = 7.18$, $\sigma = 0.4126$ and $k = 0$.

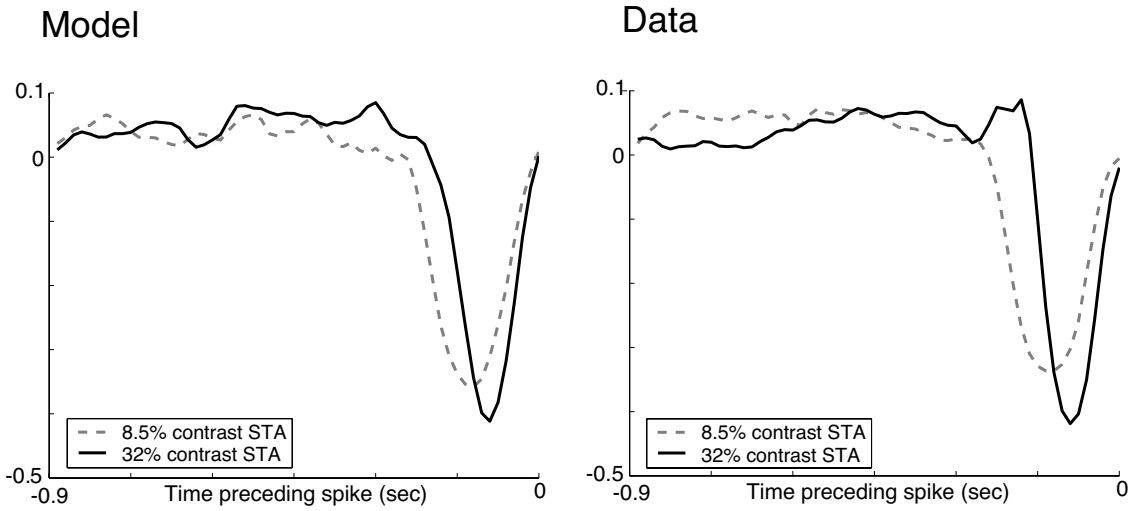

Figure 4.15. ML estimated model producing changes in the STA as a function of contrast for salamander retinal ganglion cell. Changes in the STA for 8.5 percent and 32 percent contrast in salamander cell (1999-11-12-B6A). Input stimuli are full-field 33Hz flickering binary white noise. Left, STA kernels derived from simulated spikes using ML-estimated model. Right, STA kernels derived from ganglion cell spikes (same as Figure 4.12). Kernels are normalized to unit energy.

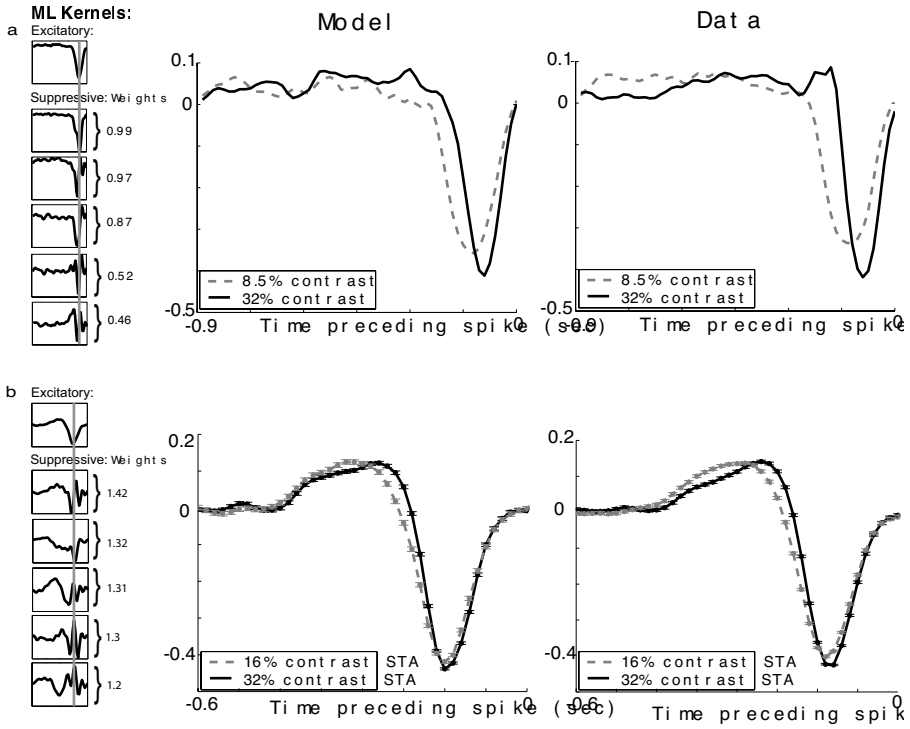

Figure 4.16. Data and ML retrieved model changes in the STA for different contrasts in salamander retinal ganglion cells. Input stimuli are full-field 33Hz flickering binary white noise. **a**, 1999-11-12-B6A. In the data, 8.5 percent contrast resulted in 623 spikes with spike rate of 0.0078; and 32 percent contrast in 1677 spikes with spike rate of 0.0404. Remaining model parameters are: $p = 7.18$, $\sigma = 0.4126$, $k = 0$. **b**, 1999-01-19-G2A. In the data, 16 percent contrast resulted in 8074 spikes with spike rate of 0.0447; and 32 percent contrast in 8877 spikes with spike rate of 0.0490. Remaining model parameters are: $p = 3.9206$, $\sigma = 1.6712$, $k = -0.6126$. Kernels are normalized to unit energy. Error bars are standard deviations computed by boot-strapping 1000 times from the set of stimuli eliciting spikes.

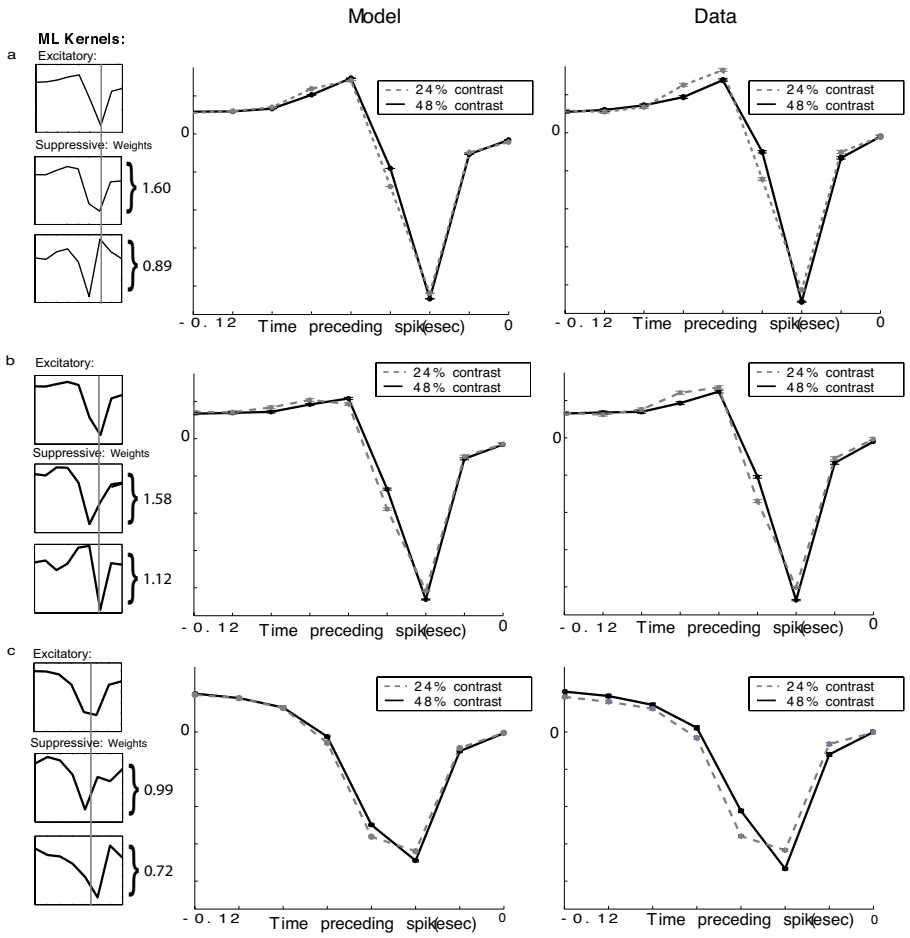

Figure 4.17. Data and ML retrieved model changes in the STA for different contrasts in monkey ganglion cells. Input stimuli are full-field 33Hz flickering Gaussian white noise. **a**, 1999-01-28-A4A. In the data, 24 percent contrast resulted in 25778 spikes with spike rate of 0.2309; and 48 percent contrast in 23346 spikes with spike rate of 0.2776. Remaining model parameters are: $p = 2.7687$, $\sigma = 1.8504$, $\alpha = -0.9914$ **b**, 1999-01-28-B1A. In the data, 24 percent contrast resulted in 18867 spikes with spike rate of 0.1690; and 48 percent contrast in 17477 spikes with spike rate of 0.2078. Remaining model parameters are: $p = 2.7580$, $\sigma = 1.88$, $\alpha = -0.8857$ **c**, 1999-01-28-D2A. In the data, 24 percent contrast resulted in 7623 spikes with spike rate of 0.0683; and 48 percent contrast in 9466 spikes with spike rate of 0.1125. Remaining model parameters are: $p = 2.3809$, $\sigma = 1.0016$, $\alpha = 0$. Error bars are standard deviations computed by boot-strapping 1000 times from the set of stimuli eliciting spikes.

estimate of the excitatory kernel. Specifically, we assume that the true excitatory kernel lies within the subspace spanned by the estimated (biased) excitatory and suppressive kernels. So, assuming a particular gain control model, we maximize the likelihood of the data, but now allowing both the excitatory and suppressive kernels to move within the subspace spanned by the initial estimated kernels. The resulting suppressive kernels need not be orthogonal to the excitatory kernel.

We maximize the likelihood of the full two-contrast data set using the model described by equation (4.3). The subspace is given by the STA and STC retrieved from the high contrast data (since the high contrast suppressive kernels are typically more salient than the low contrast kernels). The excitatory axis is initially set to the STA and the suppressive axes are set to the low-eigenvalue eigenvectors of the STC, along with the STA (e.g., to allow for self-suppression). But during the fitting procedure, the excitatory and suppressive axes are allowed to take on any linear combination of the original STA and STC estimates. The recovered axes, weights and parameters are shown in Figure 4.14. The recovered suppressive axes typically extend farther in time (preceding a spike) than the recovered excitatory axis. Once again, these suppressive axes should not be interpreted as individual neurons. Whereas the axes recovered from the STA/STC analysis are orthogonal, the axes determined during the maximum likelihood stage need not be (and in the data example are not) orthogonal.

Figure 4.15 demonstrates that the fitted model accounts for the change in STA observed at different contrast levels. Specifically, we simulate responses of the model (equation (4.3) with Poisson spike generation) on each of the two contrast stimulus sets, and then compute the STA based on these simulated spike trains.

Figures 4.16 and 4.17 demonstrate similar results for additional salamander and monkey cells. Although the fitted model is based on a single fixed excitatory kernel, the model exhibits a change in STA shape as a function of contrast very much like the ganglion cell data.

4.8 Extensions

We have limited the discussion thus far to recovering a suppressive subspace. But in addition to low-variance axes, the analysis can recover high-variance axes (which we will denote as excitatory axes). Excitatory axes are detected in some ganglion cells, as demonstrated in figure 4.18,a. A scatter plot of the projection onto the STA versus the projection onto the STC excitatory axis reveals an interesting shape; one that resembles a portion of an annulus (figure 4.18,b). We are still in the process of analyzing those neurons with excitatory axes and formulating a more generalized characterization.

In addition, we have limited the discussion to full-field stimuli that vary only in time. But the method can be used more generally to probe spatio-temporal dimensions. Preliminary data gathered by Rust and Movshon in area V1 (unpublished) reveal suppressive axes in one dimension of space and time in complex direction-selective cells (Figure 4.19). The stimuli presented to the neuron are randomly chosen bars. The STA is close to zero. This can occur, for example, if there are equally excitatory axes symmetric about the origin. Then, a stimulus and its reversed polarity (which are equally likely to occur in spherical input) are equally likely to elicit a spike along each of these axes. Thus, the stimuli eliciting spikes

along these axes will average to zero. The STC, in contrast to the STA, can detect axes that are symmetric about the origin. The STC analysis for this cell retrieves a number of non-arbitrary axes, both excitatory and suppressive. The excitatory axes are consistent with Touryan et al. [152], in which each pair of kernels obtained from STC analysis are approximately 90 degrees out of phase. This is expected for complex cells, which respond roughly equally to counter-phase gratings. But in addition to Touryan's result on complex cells, the method also detects a suppressive subspace. The nature of the suppression, and whether it is divisive or subtractive, is yet to be determined.

4.9 Discussion

We have described a spike-triggered covariance method for characterizing a neuron with a suppressive subspace (and specifically, gain control). We have demonstrated the plausibility of the technique through simulation and analysis of neural data. We have shown that under a gain control model, axes of low variance can reveal information about the response properties of neurons. In addition, this type of model can explain the observed biases in computing the STA under different contrast levels.

The suppressive axes recovered from retinal ganglion cell data appear to be significant because: (1) As in the model simulation, a small number of eigenvalues are significantly below the rest; (2) The eigenvectors associated with these axes are concentrated in a temporal region immediately preceding the spike, unlike the remaining axes; (3) Projection of the multi-dimensional stimulus vectors onto these

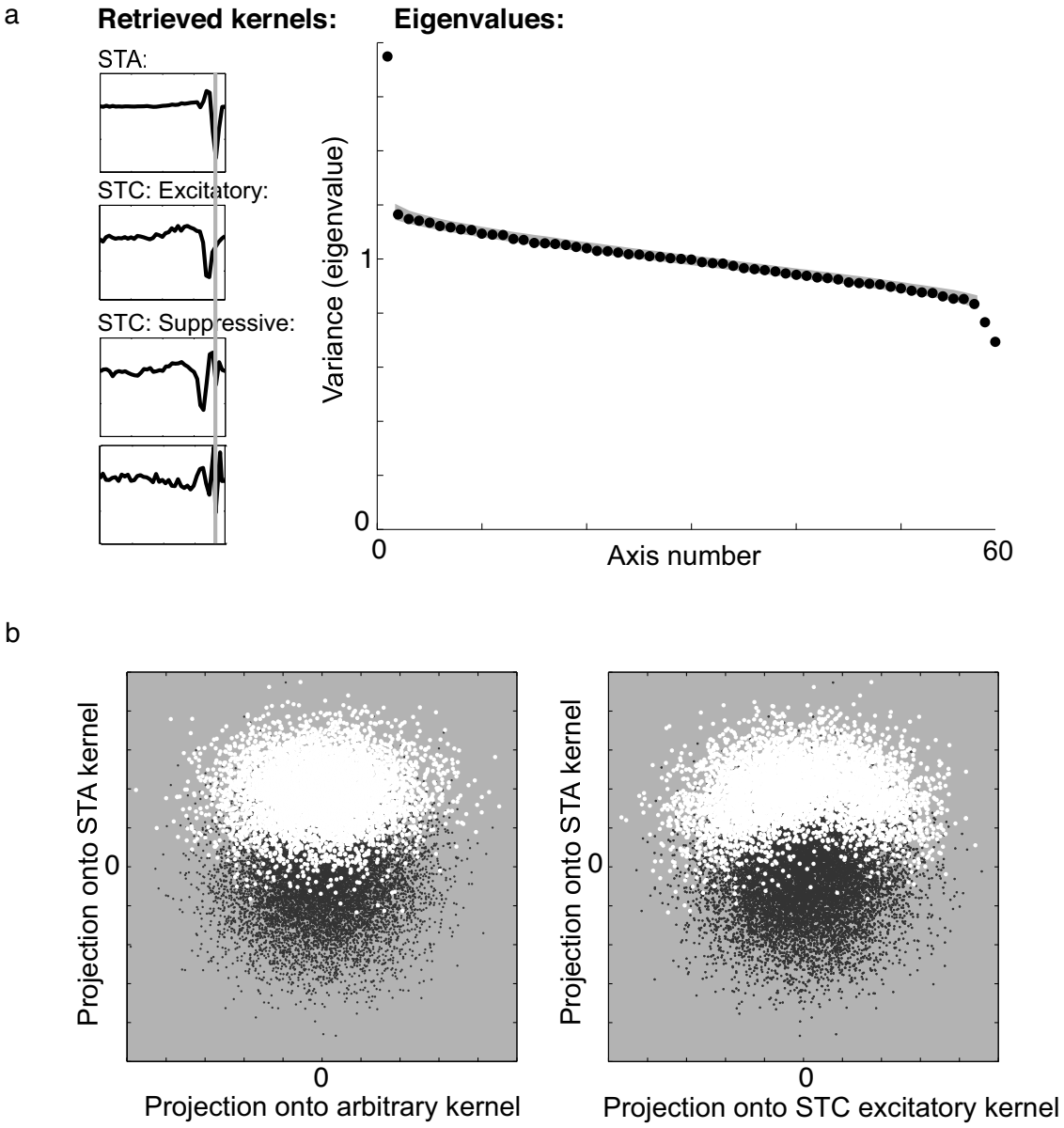

Figure 4.18. Facilitation in monkey retinal ganglion cell (2002-05-03-0-B7A). **a**, Left: Kernels from STA and STC analysis. STC analysis reveals excitatory and suppressive kernels. Right: Sorted eigenvalues of covariance matrix of stimuli eliciting spikes. **b**, Scatter plots of stimuli projected onto recovered axes.

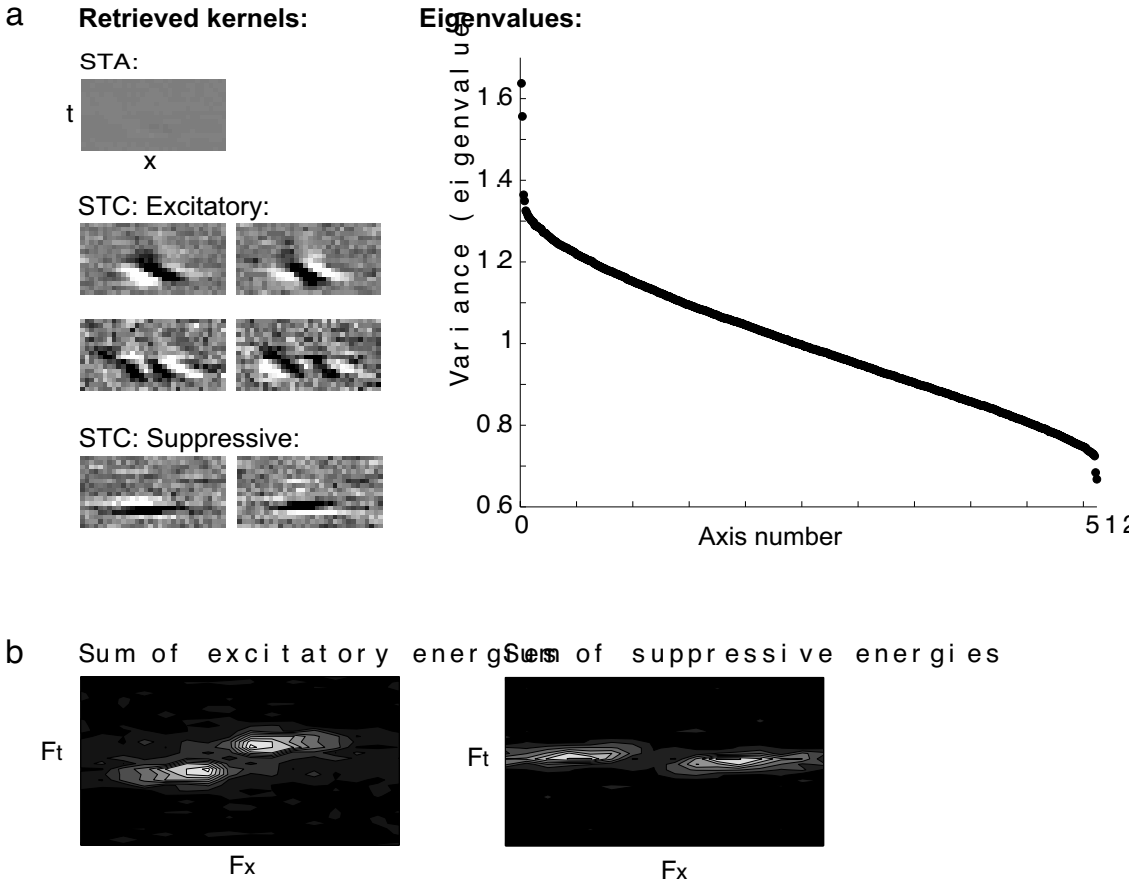

Figure 4.19. Estimated kernels in V1 direction selective complex cell. Input stimuli are 32 bars of binary white noise, and the stimuli preceding a spike are taken over a 16-segment time window. Thus, the full stimulus dimensionality over time and one dimension of space is 512. There are 32441 spikes in the analysis. **a**, Left, Kernels from STA and STC analysis. STC analysis reveals excitatory and suppressive kernels. Each pair is approximately 90 degrees out of phase. Right, Sorted eigenvalues of covariance matrix of stimuli eliciting spikes. **b**, Sum of kernel energies of excitatory and suppressive kernels.

axes reveal reductions of spike probability; (4) The full model, with parameters recovered through maximum likelihood, explains changes in STA as a function of contrast.

Models of retinal processing often incorporate gain control [135, 134, 158, 78, 90]. We have shown for the first time how one can use white noise analysis to recover a gain control subspace directly from a neuron. The kernels defining the suppressive subspace correspond to relatively short timescales of milliseconds. This appears consistent with the rapid gain control model of Victor [158], in which contrast adjusts retinal dynamics with a time constant of 15 milliseconds. In Victor's model, a feedback gain signal is formed from a lowpass filtered version of the rectified response, and is used to modulate the neural response. Victor's model is also capable of explaining temporal changes in the STA. It is difficult, however, to compare Victor's model directly to our characterization: his model consists of a series of dynamical (and sometimes) nonlinear filter descriptions; our characterization, in contrast, defines an input output relationship.

Slow gain control has also been documented in retinal ganglion cells [35, 147, 78]. An increase in contrast typically results in the reduction of the firing rate (and STA amplitude) over the course of several seconds. The fact that we do not see evidence for slow gain control in the analysis might indicate that these signals do not lie within a low-dimensional stimulus subspace. For example, if the slow gain is set by many axes over a significant temporal window, one might expect to see many slightly suppressive axes.

Our analysis is sensitive to the constraint that the input stimuli should be spherically symmetric. In particular, the proofs in the chapter for both the STA and

STC analysis rely on spherical symmetry. All of the salamander stimuli were binary. Binary stimuli can potentially introduce artifacts because they do not spherically explore the input space. We partially corrected for this by whitening the input, which makes the covariance structure sphered; but this cannot in principle make the data entirely spherical. For the salamander binary data, it is possible that artifacts may have mixed in with the true suppressive axes. We believe this did not substantially effect the results. In particular, the ML fit takes into account both the stimuli that elicited spikes and those that did not (as opposed to the STC analysis, which is performed only on the stimuli eliciting spikes). As a result, the ML stage is unlikely to include artifactual axes, because suppressive axes will be those that really brought about less spikes (as opposed to axes that were absent from the raw stimulus set). This was also reflected in the results of the ML fit: even for the binary stimuli, the full ML model could explain the changes in the STA as a function of contrast. Nevertheless, future experiments should clearly be performed with stimuli that are spherically symmetric. Indeed, this will be required if one wishes to make a thorough comparison across cells and species.

The model underlying the analysis is a phenomenological description of the response of a neuron. Thus, the analysis is not capable of distinguishing between physiological mechanisms that could underlie gain control behaviors. The model assumes a Poisson spiking mechanism. We stress that even though the complete model is able to account for changes in the STA shape as a function of contrast, this does not indicate that the real spiking mechanism of the retinal ganglion cells is Poisson. If spikes are generated according to a Poisson process, then each spike should be independent of previous spikes. But spike trains analyzed in different

regions of the brain indicate that spikes are not independent. For example, Troy and Robson found that the interval statistics of spike trains in retinal ganglion cells are gamma distributed (and not exponentially distributed, as would be expected in a Poisson process) [154]. A more realistic spike generation mechanism should take into account refractory period, as well as larger timescale dependencies on the previous spike. Other modeling work in the lab has suggested that an integrate and fire spike generation can also produce temporal changes in the STA [108]. It will be interesting to understand how much of the changes we describe here may be attributed to spiking mechanism. Physiological mechanisms underlying gain control may also be due to internal biochemical adjustments, synaptic depression, and shunting inhibition driven by other neurons.

The methodology will not recover all forms of suppression in a neuron. The measurements assume that the response of a neuron is suppressively modulated by a projection onto a subspace, and that this suppression is revealed by a reduction in the variance of the response. The analysis is second-order, and thus is blind to suppressive axes that do not effect the variance of the neural response. In particular, the analysis will not be able to pick out any anti-symmetric responses. That is, the analysis is blind to a mechanism that preserves the second order aspects of the response, but changes other aspects, such as the skew or kurtosis. A more generalized approach would be to use measures from information theory to examine arbitrary changes between the distributions of the raw stimuli and spike-triggered stimuli. For example, the Kullback-Leibler divergence provides a quantitative measure for how different two distributions are. The limitation in using a more generalized approach is, of course, that experimentally there is not enough data to describe any arbitrary

distribution in a high dimensional space. Therefore, simplifying assumptions are required (see for example, [104]).

Even in the spike-triggered covariance approach, an important theoretical question is how much spike data is actually needed to detect suppressive axes. For example, in some cells (salamander cell 1999-11-12-B6A) only hundreds of spikes were sufficient to detect suppressive axes; but other cells required thousands of spikes. The amount of spike data required largely depends on the strength of suppression in the neuron, and the dimensionality of the stimulus space to be recovered.

Figure 4.20 illustrates the error in kernel estimation, as a function of the variance reduction of a suppressive axis in simulation. We assume a suppressive kernel that reduces the response variance by a factor of either 0.16 or 0.85. We run a Monte carlo simulation, selecting Gaussian samples and projecting them onto a 60-dimensional space. All the vectors in the space have variance of 1 and only one of the vectors has reduced variance (i.e., the suppressive kernel). Therefore, all but one of the eigenvalues will be distributed around 1, and we are interested in how many samples are required to significantly pick out the suppressive eigenvalue. First, note that in both cases for large error, the suppressive eigenvalue is within the range of the gradually descending eigenvalues of the random vectors, and cannot be recovered. Thus, the suppressive eigenvalue is undistinguishable from a typical axis eigenvalue. For smaller errors (and larger sample size) the spread of gradually descending eigenvalues becomes tighter and the suppressive eigenvalue becomes distinct. The STC analysis can only detect suppressive axes that are below the spread of eigenvalue variances expected for random axes. In addition, the number of samples required to detect a suppressive axis depends dramatically on the variance

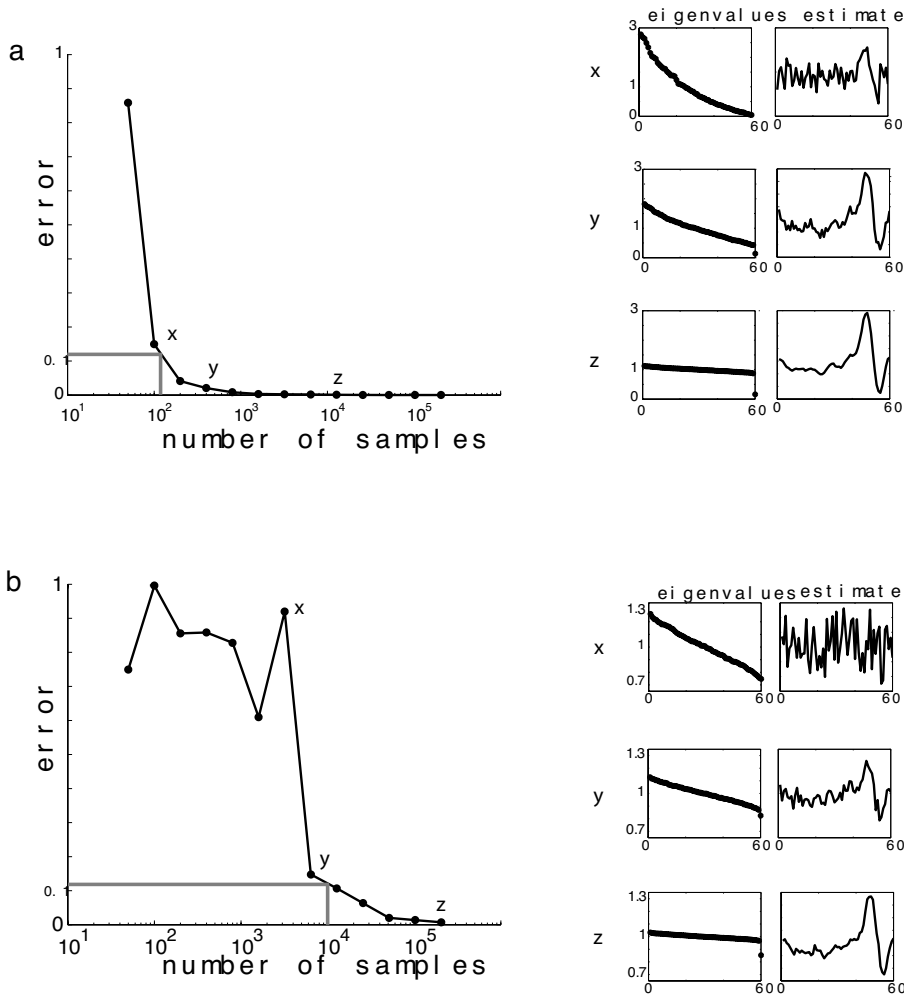

Figure 4.20. Estimation error as a function of sample size. For each sample size, Gaussian stimuli are projected onto a 60-dimensional space. The space is a an ellipsoid, with all axes in the space of variance one and only one axis of reduced variance (i.e., the suppressive axis). An STC analysis pulls out the lowest eigenvalue eigenvector, which is the estimated suppressive kernel. The error is the angle between the true and estimated suppressive kernel, such that the maximal error is 1. Gray line illustrates an example point of equal error on both graphs. Also shown are the eigenvalues and estimated suppressive kernel for particular points. a, Variance reduced by factor of 0.16 along suppressive direction. b, Variance reduced by factor of 0.85 along suppressive direction.

of the suppressive kernel: for a fixed error (gray solid line in figure) between one and two hundred samples are needed for 0.16 variance, and around 10000 samples for 0.85 variance.

Related work in statistics provides a closed-form equation that can indicate asymptotically how much data is needed. The expected distribution of eigenvalues of white noise stimuli is computed as the number of samples N (i.e., spikes) and the dimensionality of the space p tend to infinity with a fixed ratio [88]. In particular, the minimal/maximal eigenvalues of this distribution are $\sigma^2(1 \pm \sqrt{p/N})^2$, where σ^2 is the noise variance. In experimental settings the data is, of course, finite. However, the equation can still provide a rough estimate of the number of spikes required as a function of reduction in variance of a suppressive axis. To pull out a suppressive axis, its variance reduction must be smaller than the smallest eigenvalue of a random selection of white noise. Thus (for a noise variance of 1) , $V_{supp} < (1 - sqrt(p/N)^2$, where V_{supp} is the variance of the suppressive axis. For $p = 60$ and $V_{supp} = 0.16$ we obtain 166.667; and for $p = 60$ and $V_{supp} = 0.856$ we obtain 9.8504e+03. The ratio is roughly a factor of 60, similar to Figure 4.20 (and unrelated to the 60 dimensions of the stimulus space!).

The equation also indicates that for a fixed suppression variance, the number of required samples scale linearly with the dimensionality of the space. This could place a constraint on the ability to gather enough data in high dimensional spaces, unless the suppression is considerably strong. A challenging avenue of work will be to come up with more efficient stimuli that explore the input space, especially in higher neural areas in which Gaussian white noise does not drive the cells very well. An interesting direction will be to develop methodologies using stochastic stimuli

that are more realistic than white noise, but perhaps not as complex as naturalistic stimuli.

As stated in the extensions section, future work will probe retinal ganglion cells with stimulus dimensions that vary not only over time, but also over space. Whereas suppression over time may be explained using non-Poisson spiking mechanisms, spatial suppression is likely to require other explanations. For example, spatial suppression has been documented from regions outside the classical receptive field of a neuron. Will the method pull out suppressive axes outside the classical receptive field, and will these axes be useful for characterizing a neuron? We expect that in V1 the STA spatial dimensions should vary as a function of contrast [128]. Will the method be able to recover a subspace that can account for these variations in the STA? And how will this bear out in the retina? Will such representations also predict a wider range of experimental data?

Though the gain control representation is also overly simplistic, it will potentially offer a more powerful representation than the STA. Though Wiener and Volterra analysis can represent any nonlinearity with an infinite amount of higher order kernels [168, 160], our methodology requires only computation of the second order (covariance) matrix, which is more feasible for a finite amount of data. In particular, gain control behaviors are not well captured by a small number of power series kernels. The concept of gain control is central to many neural models, and thus our work provides a technique with potentially widespread applicability.

4.10 Proofs

We assume that the response of a neuron is given by the projection onto a subspace followed by a nonlinearity, and that the input to the neuron are samples of independent Gaussian white noise with a variance of 1. We now prove that for a stimulus direction outside the subspace, the variance of the response of the neuron will be equal to 1 (at the limit of infinite data).

Specifically, we assume that spikes are generated by a Poisson process with instantaneous rate determined by:

$$\mathcal{P}\left(spike|\vec{s}\right) = f(\vec{K} \cdot \vec{s}) \tag{4.5}$$

where K is a subspace and $f(\cdot)$ is a generalized nonlinearity. For example, for the model discussed in the chapter, the subspace K would be comprised of excitatory and suppressive axes; and the nonlinearity would be a gain control function. We assume without loss of generality that K is a tall and skinny matrix with orthogonal columns.

According to Bayes rule:

$$\mathcal{P}\left(\vec{s}|spike\right) = \mathcal{P}\left(spike|\vec{s}\right)\mathcal{P}\left(\vec{s}\right)/\mathcal{P}\left(spike\right) \tag{4.6}$$

The spike-triggered covariance (STC) is defined as follows:

$$STC = \int_{\vec{s}} \vec{s}\vec{s}^{T}\mathcal{P}\left(\vec{s}|spike\right) \tag{4.7}$$

where we assume that the mean has been subtracted off of $\vec{s}$.

And plugging in Bayes rule (equation 4.6):

$$STC = \int_{\vec{s}} \vec{s}\vec{s}^{T}\mathcal{P}\left(spike|\vec{s}\right)\mathcal{P}\left(\vec{s}\right)/\mathcal{P}\left(spike\right) \tag{4.8}$$

Plugging in equation 4.5:

$$STC = \int_{\vec{s}} \vec{s}\vec{s}^T f(\vec{K}^T \cdot \vec{s}) \mathcal{P}(\vec{s}) / \mathcal{P}(spike) \tag{4.9}$$

We want to show that for any stimulus direction outside the subspace K, the variance of the response of the neuron will be equal to 1 in the limit of infinite data.

Any stimulus $\vec{s}$ can be decomposed into a component in the direction of K and a component in a perpendicular direction, $K^{\perp}$:

$$\vec{s} = \vec{s}_K + \vec{s}_{\perp} \tag{4.10}$$

where $\vec{s}_K = (KK^t)\vec{s}$ and $\vec{s}_{\perp} = (I - KK^t)\vec{s}$.

Then from orthogonality we obtain the following two equations:

$$\vec{s}_{\perp}^T \vec{s}_K = 0 \tag{4.11}$$

$$\vec{K}^T \vec{s_{\perp}} = 0 \tag{4.12}$$

Now since the input stimuli are Gaussian white noise, the samples are statistically independent, so: $\mathcal{P}(\vec{s}) = \mathcal{P}(\vec{s_K})\,\mathcal{P}(\vec{s_{\perp}})$

Therefore equation 4.9 becomes:

$$STC = \int_{\vec{s}} \vec{s}\vec{s}^T f(\vec{K}^T \cdot \vec{s}) \mathcal{P}(\vec{s_K})\,\mathcal{P}(\vec{s_{\perp}}) / \mathcal{P}(spike) \tag{4.13}$$

Separating out $\vec{s_K}$ and $\vec{s_{\perp}}$ in the integral of 4.13 we obtain (also based on equation 4.12):

$$STC = \int_{\vec{s_{\perp}}} \mathcal{P}(\vec{s_{\perp}}) \int_{\vec{s_K}} \vec{s}\vec{s}^T f(\vec{K}^T \cdot \vec{s_K}) \mathcal{P}(\vec{s_K}) / \mathcal{P}(spike) \tag{4.14}$$

Now consider the variance of a unit vector in the direction $\hat{u}$ lying in $K^{\perp}$. The variance is given by $\hat{u}^T(STC)\hat{u}$. That is:

$$\hat{u}^T(STC)\hat{u} = \int_{\vec{s_\perp}} \mathcal{P}\left(\vec{s_\perp}\right) \int_{\vec{s_K}} \hat{u}^T \vec{s}\vec{s}^T \hat{u} f(\vec{K}^T \cdot \vec{s_K}) \mathcal{P}\left(\vec{s_K}\right) / \mathcal{P}\left(spike\right) \tag{4.15}$$

But based on the decomposition of $\vec{s}$ and equation 4.11:

$$\hat{u}^T \vec{s}\vec{s}^T \hat{u} = \hat{u}^T \vec{s_\perp} \vec{s_\perp}^T \hat{u}. \tag{4.16}$$

This term can thus be moved to the outer integral in equation 4.15, since it only depends on $s_\perp$:

$$\hat{u}^T(STC)\hat{u} = \int_{\vec{s_\perp}} \mathcal{P}\left(\vec{s_\perp}\right) \hat{u}^T \vec{s_\perp} \vec{s_\perp}^T \hat{u} \int_{\vec{s_K}} f(\vec{K}^T \cdot \vec{s_K}) \mathcal{P}\left(\vec{s_K}\right) / \mathcal{P}\left(spike\right) \tag{4.17}$$

Now the inner integral is equal to 1 because: $\int_{\vec{s_K}} f(\vec{K}^T \cdot \vec{s_K}) \mathcal{P}\left(\vec{s_K}\right)$ is actually $\int_{\vec{s_K}} \mathcal{P}\left(\vec{s_K}\right) \mathcal{P}\left(spike|\vec{s_K}\right)$, which is equal to $\mathcal{P}\left(spike\right)$. This exactly cancels out with the $\mathcal{P}\left(spike\right)$ in the denominator.

Thus we are now left with:

$$\hat{u}^T(STC)\hat{u} = \int_{\vec{s_\perp}} \mathcal{P}\left(\vec{s_\perp}\right) \hat{u}^T \vec{s_\perp} \vec{s_\perp}^T \hat{u} \tag{4.18}$$

But $\vec{s_\perp}$ is univariate Gaussian, and this is just the variance in direction $\hat{u}$, which must be 1.

CHAPTER 5

GENERAL DISCUSSION

We set out to understand what kind of representations sensory neurons form, and why they form the particular representations that they do. Since sensory neurons exhibit highly nonlinear behaviors, we aimed to go beyond linear receptive field representations. We utilized computational approaches to try and understand the multiple axes (or stimulus dimensions) that might govern the responses of neurons.

We approached the "why" question using the efficient coding hypothesis; this led to a gain control model, known as divisive normalization, whereby the squared response of a primary filter is divided by the weighted squared responses of other filters and an additive constant. We further sought to understand "what" representations neurons form in the context of a gain control model. We demonstrated that a spike-triggered covariance technique can be used to characterize a gain control subspace in spiking neurons.

As noted in the introduction, neural gain control models have been constructed to explain physiological sensory nonlinearities in a number of systems. Gain control has its roots in models of retinal processing [134]. Gain control was also suggested

in the context of movement in the visual system of the fly [113]. The divisive normalization model was coined to explain physiological data inside the classical receptive field in area V1 [64]. Gain control models have also been formulated independently in the context of the peripheral auditory system, spearheaded by the work of Lyon [84]. Our work thus provides theoretical justification to such neural models.

Throughout the thesis we employ two very different computational frameworks for developing a gain control model– theoretical and through neural characterization. We also examine three different neural systems: retinal ganglion cells, V1, and auditory nerve. Nevertheless, some of the conclusions we reach are rather general.

Our work helps to elucidate particular aspects of divisive gain control models. It was originally suggested that constancy of tuning as a function of input strength is an important property for a system that recognizes objects in the world [64]. A gain control model with uniform weighting is able to satisfy this condition, and is in line with the constancy of orientation tuning observed inside the classical receptive field in V1 cells.

However, we have seen three examples in which the tuning properties of sensory neurons are not always constant as a function of input strength. In area V1, the diameter tuning of the mean neural response changed for low and high contrast stimuli; in auditory nerve, the frequency tuning of the mean neural response changed for low and high sound pressure levels; and in retinal ganglion cells the temporal tuning of the mean neural response changed for low and high contrasts.

Changes in tuning shape occur in our model when the weighting of the divisive signal is not uniform over a particular parameter (orientation, frequency, position,

time, and so on) in the operating regime of the primary filter. Stated differently, the fundamental tuning properties of the response is determined by a primary linear filter (or, more generally, a possible combination of several filters, as in V1 complex cells), and by a gain control signal with its own fixed tuning properties. Nonetheless, the combination of these two fixed entities can modulate the overall tuning of the neural response. The gain signal is most modulatory at high signal strengths. But at low signal strengths, the constant dominates the divisive equation and the gain signal is inconsequential; the tuning of the neural response resembles that of the primary filter.

We have also seen that masks with different attributes can exert very different effects on the response properties of neurons. This is apparent both in V1 data from beyond the classical receptive field and in auditory two-tone suppression. The mask typically shifts the response curve of the neuron to the right on a log scale, indicative of a divisive gain change. These shifts are easily explained with a non-uniformly weighted divisive gain control model; stronger weighting along a particular stimulus axis results in a greater rightward shift.

This picture of fixed excitatory and suppressive tuning, should be contrasted with an alternative view of nonlinear sensory behaviors. One can imagine an excitatory and suppressive pool of filters that are not fixed; their tuning changes as a function of contrast. Sceniak et al. have suggested that changes in V1 diameter tuning occur because a neuron recruits more excitatory connections from neighboring neurons [128]. Polat et al. have suggested that low and high contrast stimuli dictate whether neighboring neurons in area V1 exert an inhibitory or excitatory effect [109]. This would imply two different strategies for handling low and high

contrast stimuli. Although such interpretations cannot be ruled out, the weighted divisive gain control model offers an alternative explanation of the data. The tuning properties of a primary filter and a divisive gain signal are fixed, but the relative impact of the divisive gain signal varies as a function of input strength. The advantage of this formulation is that a single model can explain behaviors at both low and high stimulus strengths, and can further explain nonlinear masking data. The fixed model is, of course, overly simplistic. It does not incorporate temporal dynamics or possible adaptation effects (see [162] for adaptation in the context of a gain control model). Nevertheless, the model can explain some basic nonlinearities in sensory neural data.

In chapter 3, this form of model provides a qualitative match to V1 orientation data, with a gain signal that is less selective in the center for orientation, and more strongly selective for orientation in the surround; and with a gain signal that extends spatially beyond the primary excitatory receptive field and falls off with distance. The resulting model can qualitatively explain both data inside the classical receptive field (as in the classical divisive normalization model [64]) and data outside the classical receptive field (as in Cavanaugh et al. [33]). The model also explains basic auditory nerve phenomenon, with a gain signal that is stronger at the preferred temporal frequency of the neuron. This is consistent with a recent gain control model by Robert et al., in which the gain signal is frequency dependent, with strongest weighting at the preferred temporal frequency [119]. The gain signal in their model controls the width of tuning of a primary filter over time. In chapter 4, we recover a low-dimensional gain control signal in retinal ganglion cells. The gain control signal is selective for temporal attributes of the stimulus. This bears

some similarity to Victor's model, in which the contrast of the stimulus over time determines the time constant of the primary filter [157]. In all cases, we have worked with a single form of gain control model, in which the tuning of the primary filter and the gain filter determine particular nonlinear response properties of the neuron.

Gain control has been suggested as a mechanism by which a neuron can deal with a limited dynamic range. It appears that gain control in sensory processing is not only adjusted by the overall amplitude or contrast of the input stimuli. It is, under certain circumstances, also sensitive to other qualities of the stimuli, and is stronger for particular stimulus attributes. Our work suggests that the tuning of the gain signal relative to the primary filter is not an accident; the gain signal might operate along dimensions that are optimal for forming efficient representations of typical stimuli in the environment.

The divisive gain control model provides a functional description, and does not specify the circuitry or biophysics by which these functions are implemented. The specific form of implementation, and the timescales over which gain control operates are likely to be different across sensory systems, levels of processing, and species. For example, in the auditory system outer hair cells have been implicated in providing gain control; but there are species of vertebrates that have no outer hair cells and still exhibit two-tone suppression [59]. Even within the same species, multiple mechanisms appear to be at work. Gain control in the retina is mediated by a number of mechanisms and levels, likely involving both feedforward and feedback connections [134, 153, 157]. Some of the temporal gain control behaviors we describe in chapter 4 might be due to non-Poisson spiking mechanisms (see chapter 4 discussion). In area V1, synaptic depression could potentially explain gain control

in the center [1], but gain control in the surround is likely to involve feedback from other neurons, perhaps via a form of shunting inhibition [29]. Implementation and dynamics, while partially understood in these different systems, remains an active topic of research.

Gain control might also operate in a cascaded fashion, as suggested in some models of cortical processing [143]. One might imagine a system starting with luminance and contrast gain control in the retina; spatio-temporal gain control in area V1; speed gain control in MT; and perhaps color gain control as well. Similar systems might also exist within the auditory system. It is possible that each stage of processing performs gain control within its own capabilities.

Indeed, it appears that sensory systems have come up with multiple, creative, solutions for implementing gain control. Although the details of implementation will vary, our work suggests that the steady state computation underlying gain control might be generic.

The two parts of the thesis offer a potentially interesting path for future exploration. Stochastic stimuli are useful for exploring the response properties of neurons. But one of the disadvantages in stimuli such as white noise is that these are not naturalistic and do not drive cells in higher regions of the brain very well. However, natural images and sounds are often too complex for exploring neural behavior. The statistical properties of natural signals might be utilized for creating experimental stimuli that are more tractable than natural signals, yet more naturalistic than white noise. If we can derive a model of gain control from natural signals, then perhaps we can create stochastic stimuli that are suitable for testing issues of gain control in neural processing. Related work on texture synthesis provides

a potential framework for creating stimuli with specific statistical properties [110]. These directions may offer opportunities for exploring gain control along multiple time scales (as in [49]); and neural adaptation to specific statistical properties, such as mean, variance, and more complex statistical attributes.

A major issue highlighted throughout the thesis is understanding the various components that lead to an overall neural behavior. This can be studied at many different levels of detail. We have specifically focused on the interaction of excitation and suppression in producing a steady state response. If one cannot separate out these components, then the task of interpreting experimental data becomes difficult. But to do so requires an assumption about some form of underlying model, or at least class of models. At this particular juncture, neural modeling can contribute not only to the analysis of data, but also to the design of experimental methodologies that are aimed at testing and characterizing classes of model.

As nonlinearities continue to be found in neural processing, it becomes important to develop computational tools for experimentation, analysis, and modeling of the neural data. It is not uncommon that the same computational tools are utilized in the study of several systems. This has been the case with white noise reverse correlation, wiener analysis, information theory, masking experiments, and so on. As we begin to have a better grasp of the type of nonlinearities that occur in neural processing, the set of tools should be tailored for the study of these systems. Current advances in both computational capabilities and experimental techniques should offer promising opportunities for the field.

BIBLIOGRAPHY

[1] L F Abbott, J A Varela, K Sen, and S B Nelson. Synaptic depression and cortical gain control. *Science*, 275:220–224, 1997.

[2] D G Albrecht, S B Farrar, and D B Hamilton. Spatial contrast adaptation characteristics of neurones recorded in the cat's visual cortex. *Journal of Physiology (London)*, 347:713–739, 1984.

[3] D G Albrecht and W S Geisler. Striate cortex of monkey and cat: contrast response function. *Journal of Neurophysiology*, 48:217–237, 1982.

[4] J Anderson, M Carandini, and D Ferster. Orientation tuning of input conductance, excitation, and inhibition in cat primary visual cortex. *J. Neurophysiol.*, 84:909–926, 2000.

[5] B A Y Arcas, A L Fairhall, and W Bialek. What can a single neuron compute? In *Advances in Neural Information Processing Systems*, volume 13, pages 75–81, 2000.

[6] J J Atick. Could information theory provide an ecological theory of sensory processing? *Network: Computation in Neural Systems*, 3:213–251, 1992.

[7] J J Atick and A N Redlich. Towards a theory of early visual processing. *Neural Computation*, 2:308, 1990.

[8] H Attias. Em algorithms for independent component analysis. In M Niranjan, editor, *Neural Networks for Signal Processing*, volume 8, pages 132–141. IEEE, New York, 1998.

[9] H Attias and C E Schreiner. Temporal low-order statistics of natural sounds. In M Jordan, M Kearns, and S Solla, editors, *Adv in Neural Info Processing Systems*, volume 9, pages 27–33. MIT Press, 1997.

[10] H Attias and C E Schreiner. Coding of naturalistic stimuli by auditory midbrain neurons. *Adv in Neural Info Processing Systems*, 10:103–109, 1998.

[11] F Attneave. Some informational aspects of visual perception. *Psych. Rev.*, 61:183–193, 1954.

[12] R Baddeley. Searching for filters with "interesting" output distributions: an uninteresting direction to explore? *Network*, 7:409–421, 1996.

[13] R Baddeley, L F Abbott, M C Booth, F Sengpiel, T Freeman, E A Wakeman, and E T Rolls. Respones of neurons in primary and inferior temporal visual cortices to natural scenes. *Proc. Roy. Soc. (Lond.)*, B264:1775–1783, 1998.

[14] R M Balboa and N M Grzywacz. The role of early lateral inhibition: More than maximizing luminance information. *Visual Neuroscience*, 17:77–89, 2000.

[15] H Barlow. Redundancy reduction revisited. *Network: Computation in Neural Systems*, 12:241–253, 2001.

[16] H B Barlow. Possible principles underlying the transformation of sensory messages. In W A Rosenblith, editor, *Sensory Communication*, page 217. MIT Press, Cambridge, MA, 1961.

[17] Horace B Barlow and P Foldiak. Adaptation and decorrelation in the cortex. In R Durbin, C Miall, and G Mitchinson, editors, *The Computing Neuron*. Addison-Wellesley, New York, 1989.

[18] L A Bauman and A B Bonds. Inhibitory refinement of spatial frequency selectivity in single cells of the cat striate cortex. *Vision Research*, 31:933–944, 1991.

[19] A D Baylor and A L Hodgkin. Detection and resolution of visual stimuli by turtle photoreceptors. *Journal of Physiology*, 242:729–758, 1973.

[20] A J Bell and T J Sejnowski. An information-maximisation approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, 1995.

[21] A J Bell and T J Sejnowski. Learning the higher-order structure of a natural sound. *Network: Computation in Neural Systems*, 7:261–266, 1996.

[22] A J Bell and T J Sejnowski. The 'independent components' of natural scenes are edge filters. *Vision Research*, 37(23):3327–3338, 1997.

[23] C Blakemore and E A Tobin. Lateral inhibition between orientation detectors in the cat's visual cortex. *Exp Brain Res*, 15:439–440, 1972.

[24] A B Bonds. Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex. *Visual Neuroscience*, 2:41–55, 1989.

[25] A B Bonds. Temporal dynamics of contrast gain in single cells of the cat striate cortex. *Visual Neuroscience*, 6:239–255, 1991.

[26] L J Borg-Graham, C Monier, and Y Frénac. Visual input evokes transient and strong shunting inhibition in visual cortical neurons. *Nature*, 393:369–373, 1998.

[27] R T Born and R B Tootell. Single-unit and 2-deoxyglucose studies of side inhibition in macaque striate cortex. *Proc Natl Acad Sci USA*, 88:7071–7075, 1991.

[28] R W Buccigrossi and E P Simoncelli. Image compression via joint statistical characterization in the wavelet domain. *IEEE Trans Image Proc*, 8(12):1688–1701, December 1999.

[29] M Carandini and D J Heeger. Summation and division by neurons in primate visual cortex. *Science*, 264:1333–1336, 1994.

[30] M Carandini, D J Heeger, and J A Movshon. Linearity and normalization in simple cells of the macaque primary visual cortex. *Journal of Neuroscience*, 17:8621–8644, 1997.

[31] M Carandini, J A Movshon, and D Ferster. Pattern adaptation and cross-orientation interactions in the primary visual cortex. *Neuropharmacology*, 37:501–511, 1998.

[32] J F Cardoso. Source separation using higer order moments. In *ICASSP*, pages 2109–2112, 1989.

[33] J R Cavanaugh. *Properties of the Receptive Field Surround in Macaque Primary Visual Cortex*. PhD thesis, Center for Neural Science, New York University, August 2000.

[34] J R Cavanaugh, W Bair, and J A Movshon. Orientation-selective setting of contrast gain by the surrounds of macaque striate cortex neurons. In *Neurosci Abstracts*, volume 23, page 227.2, 1997.

[35] D Chander and E J Chichilnisky. Adaptation to temporal contrast in primate and salamander retina. *J Neurosci*, 21(24):9904–9916, 2001.

[36] E J Chichilnisky. A simple white noise analysis of neuronal light responses. *Network: Computation in Neural Systems*, 12(2):199–213, 2001.

[37] P Common. Independent component analysis, a new concept? *Signal Process.*, 36:387–314, 1994.

[38] T M Cover and J A Thomas. *Elements of Information Theory*. Wiley Series, 1991.

[39] K J W Craik. *The nature of explanation*. Cambridge: University Press, 1943.

[40] P Dayan and L F Abbott. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. MIT Press, 2001.

[41] P Dayan, G E Hinton, R M Neal, and R S Zemel. The Helmholtz machine. *Neural Computation*, 7:889–904, 1995.

[42] R de Ruyter van Steveninck and W Bialek. Coding and information transfer

in short spike sequences. In *Proc.Soc. Lond. B. Biol. Sci.*, volume 234, pages 379–414, 1988.

[43] G C DeAngelis, R D Freeman, and I Ohzawa. The organization of suppression in receptive fields of neurons in the cat's visual cortex. *Journal of Neurophysiology*, 68:144–163, 1994.

[44] G C DeAngelis, J G Robson, I Ohzawa, and R D Freeman. The organization of suppression in receptive fields of neurons in the cat's visual cortex. *Journal of Neurophysiology*, 68:144–163, 1992.

[45] E deBoer and P Kuyper. Triggered correlation. In *IEEE Transact. Biomed. Eng.*, volume 15, pages 169–179, 1968.

[46] B Delgutte. Physiological models for basic auditory percepts. In H. Hawkins and T. McMullen, editors, *Auditory Computation*. Springer-Verlag, 1996.

[47] A M Derrington and P Lennie. Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. *Journal of Physiology (London)*, 357:219–240, 1984.

[48] D W Dong and J J Atick. Temporal decorrelation: A theory of lagged and nonlagged responses in the lateral geniculate nucleus. *Network: Computation in Neural Systems*, 6:159–178, 1995.

[49] A L Fairhall, G D Lewen, W Bialek, and R de Ruyter van Steveninck. Multiple time scales of adaptation in a neural code. In *Advances in Neural Information Processing Systems*, volume 13, pages 124–130, 2000.

[50] D Ferster and K D Miller. Neural mechanisms of orientation selectivity in primary visual cortex. *Annu Rev Neurosci*, 23:441–471, 2000.

[51] D J Field. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A*, 4(12):2379–2394, 1987.

[52] D J Field. What is the goal of sensory coding? *Neural Computation*, 6:559–601, 1994.

[53] D J Field, A Hayes, and R F Hess. Contour integration by the human visual system: Evidence for a local "association field". *Vision Research*, 33:173–93, 1993.

[54] J M Foley. Human luminance pattern mechanisms: masking experiments require a new model. *J. Opt. Soc. Am. A*, 11(6):1710–1719, 1994.

[55] G Chechik G, N Tishby, and I Nelken. Groups redundancy measures reveal redundancy reduction along the auditory pathway. In T G Dietterich, S Becker, and Z Ghahramani, editors, *Adv. Neural Information Processing Systems*, volume 14, Cambridge, MA, 2002. MIT Press.

[56] T J Gawne, T W Kjaer, A Hertz, and B J Richmond. Adjacent visual cortical complex cells share about 20 percent of their stimulus related information. *Cerebral Cortex*, 6:482–489, 1996.

[57] T J Gawne and B J Richmond. How independent are the messages carried by adjacent inferior temporal cortical neurons? *Journal of Neuroscience*, 13:2758–2771, 1993.

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

[59] D Geisler. *From Sound to Synapse: Physiology of the Mammalian Ear*. Oxford University Press, New York, 1998.

[60] W S Geisler, J S Perry, B J Super, and D P Gallogly. Edge co-occurance in natural images predicts contour grouping performance. *Vision Research*, 41(6):711–724, March 2001.

[61] C D Gilbert and T N Wiesel. The influence of contextual stimuli on the orientation selectivity of cells in primary visual cortex of the cat. *Vision Research*, 30:1689–1701, 1990.

[62] N Graham and A Sutter. Normalization: Contrast-gain control in simple (Fourier) and complex (non-Fourier) pathways of pattern vision. *Vision Research*, 40:2737–61, 2000.

[63] H L Hawkins, T A McMullen, A N Popper, and R R Fay, editors. *Auditory Computation*. Springer, New York, 1996.

[64] D J Heeger. Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9:181–198, 1992.

[65] D J Heeger. Modeling simple cell direction selectivity with normalized, half-squared, linear operators. *Journal of Neurophysiology*, 70(5):1885–1898, 1993.

[66] J A Hirsch, J-M Alonso, R C Reid, and L M Martinez. Synaptic integration in striate cortical simple cells. *J. Neurosci.*, 18(22):2517–9528, 1998.

[67] D Hubel and T Wiesel. Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology (London)*, 160:106–154, 1962.

[68] A Hyvärinen and P Hoyer. Emergence of topography and complex cell properties from natural images using extensions of ICA. In S. A. Solla, T. K. Leen, and K.-R. Müller, editors, *Adv. Neural Information Processing Systems*, volume 12, pages 827–833, Cambridge, MA, May 2000. MIT Press.

[69] A Hyvarinen, J Karhunen, and E Oja. *Independent Component Analysis*. Wiley Series, 2001.

[70] B S Jackson and E M Relkin. A frequency-dependent saturation evident in rate-intensity functions of the chinchilla auditory nerve. *Hearing Research*, 126:75–83, 1998.

[71] E Javel, D Geisler, and A Ravindran. Two-tone suppression in auditory nerve of the cat: Rate-intensity and temporal analyses. *J. Acoust. Soc. Am.*, 63(4):1093–1104, 1978.

[72] E T Jaynes. Information theory and statistical mechanics. *Phys. Rev.*, 106:620–630, 1957.

[73] P I M Johannesma. The pre-response stimulus ensemble of neurons in the cochlear nucleus. In *Symposium on Hearing Theory (IPO)*, pages 58–69, Eindhoven, Holland, 1972.

[74] J P Jones and L A Palmer. The two-dimensional spatial structure of simple

receptive fields in the cat striate cortex. *J Neurophysiology*, 58:1187–11211, 1987.

[75] C Jutten and J Herault. Blind separation of sources, part i: An adaptive algorithm based on neuromimetic architecture. *Signal Process.*, 24(1):1–10, 1991.

[76] M K Kapadia, G Westheimer, and C D Gilbert. Dynamics of spatial summation in primary visual cortex of alert monkeys. *Proc Natl Acad Sci*, 21:12073–12078, 1999.

[77] D Kersten. Predictability and redundancy of natural images. *J. Opt. Soc. Am. A*, 4(12):2395–2400, 1987.

[78] K J Kim and F Rieke. Temporal contrast adaptation in the input and output signals of salamander retinal ganglion cells. *J. Neurosci.*, 21(1):287–299, 2001.

[79] J J Knierim and D C Van Essen. Neuronal responses to static texture patterns in area v1 of the alert macaque monkey. *Journal of Neurophysiology*, 67:961–980, 1992.

[80] S B Laughlin. A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch.*, 36c:910–912, 1981.

[81] J B Levitt and J S Lund. Contrast dependence of contextual effects in primate visual cortex. *Nature*, 387:73–76, 1997.

[82] M S Lewicki. Efficient coding of natural sounds. *Nature*, 5(4):356–362, 2002.

[83] C Y Li and W Li. Extensive integration field beyond the classical receptive field of cat's striate cortical neurons–classification and tuning properties. *Vision Research*, 18:2337–2355, 1994.

[84] R F Lyon. Automatic gain control in cochlear mechanics. In P Dallos et al., editor, *The Mechanics and Biophysics of Hearing*, pages 395–420. Springer-Verlag, 1990.

[85] E Mach. *The Analysis of Sensations, and the Relation of the Physical to the Psychical*. New York: Dover (reprint), 1959.

[86] L Maffei and A Fiorentini. The visual cortex as a spatial frequency analyzer. *Vision Research*, 13:1255–1267, 1973.

[87] L Maffei and A. Fiorentini. The unresponsive regions of visual cortical receptive fields. *Vision Research*, 16:1131–1139, 1976.

[88] V A Marcenko and L A Pastur. Distributions of eigenvalues of some sets of random matrices. *USSR-Sb.*, 1:507–536, 1967.

[89] V Z Marmarelis. Wiener analysis of nonlinear feedback in sensory systems. *Annals of Biomedical Engineering*, 19:345–382, 1991.

[90] M Meister and M J Berry. The neural code of the retina. *Neuron*, 22:435–450, 1999.

[91] S Mikaelian and E Simoncelli. Modeling temporal response characteristics of V1 neurons with a dynamic normalization model. In *Neurocomputing*, volume 38–40, pages 1461–1467. Elsevier, 2001. Presented at Computational NeuroScience, Brugge, Belgium, 16-20 July 2000.

[92] J A Movshon, I D Thompson, and D J Tolhurst. Receptive field organization of complex cells in the cat's striate cortex. *Journal of Physiology (London)*, 283:79–99, 1978.

[93] J A Movshon, I D Thompson, and D J Tolhurst. Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *Journal of Physiology (London)*, 283:53–77, 1978.

[94] J R Müller, A B Metha, J Krauskopf, and P Lennie. Rapid adaptation in visual cortex to the structure of images. *Science*, 285:1405–1408, Aug 1999.

[95] K I Naka and W A H Rushton. S-potentials from luminosity units in the retina of fish (yprinidae). *Journal of Physiology*, 185:587–599, 1966.

[96] J I Nelson and B J Frost. Intracortical facilitation among co-oriented, coaxially aligned simple cells in cat striate cortex. *Exp. Brain Res.*, 61:54–61, 1985.

[97] O. Nestares and D J Heeger. Modelling the apparent frequency-specific suppression in simple cells responses. *Vision Research*, 37:1535–1543, 1997.

[98] Q D Nykamp and D L Ringach. Full identification of a linear-nonlinear system via cross-correlation analysis. *Journal of Vision*, 2:1–11, 2002.

[99] I Ohzawa, G Sclar, and R D Freeman. Contrast gain control in the cat's visual system. *Nature*, 298:266–268, 1982.

[100] I Ohzawa, G Sclar, and R D Freeman. Contrast gain control in the cat's visual system. *Journal of Neurophysiology*, 54:651–667, 1985.

[101] E Oja. A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 15:267–273, 1982.

[102] B A Olshausen and D J Field. Emergence of simple-cell receptive field properties by learning a sparse factorial code. *Nature*, 381:607–609, 1996.

[103] B A Olshausen and D J Field. Natural image statistics and efficient coding. *Network: Computation in Neural Systems*, 7:333–339, 1996.

[104] L Paninski. Convergence properties of three spike-triggered analysis techniques. In *Computational Neuroscience: Trends in Research*, 2002. To Appear.

[105] L Parra, C Spence, and P Sajda. Higher-order statistical properties arising from the non-stationarity of natural signals. In T K Leen, T G Dietterich, and V Tresp, editors, *Adv. Neural Information Processing Systems*, volume 13, Cambridge, MA, 2001. MIT Press.

[106] K Pearson. *The Grammar of Science*. London: Dcott, 1892.

[107] J O Pickles. *An Introduction to the Physiology of Hearing*. Academic Press, New York, 1988.

[108] J W Pillow and E P Simoncelli. Biases in white noise analysis due to non-poisson spike generation. In *Computational Neuroscience: Trends in Research*, 2002. To Appear.

[109] U Polat, K Mizobe, M W Pettet, T Kasamatsu, and A M Norcia. Collinear stimuli regulate visual responses depending on cell's contrast threshold. *Nature*, 5:580–584, 1998.

[110] J Portilla and E P Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *Int'l Journal of Computer Vision*, 40(1):49–71, 2000.

[111] J Portilla, V Strela, M Wainwright, and E Simoncelli. Adaptive Wiener denoising using a Gaussian scale mixture model in the wavelet domain. In *Proc 8th IEEE Int'l Conf on Image Proc*, pages 37–40, Thessaloniki, Greece, Oct 7-10 2001. IEEE Computer Society.

[112] R P N Rao and D H Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87, 1999.

[113] W Reichhardt and T Poggio. Figure-ground discrimination by relative movement in the visual system of the fly. *Biol. Cybern.*, 35:81–100, 1979.

[114] F Rieke, D A Bodnar, and W Bialek. Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. *Proc. R. Soc. Lond. B*, 262:259–265, 1995.

[115] F Rieke, D Warland, R R de Ruyter van Steveninck, and W Bialek. *Spikes: Exploring the Neural Code*. MIT Press, Cambridge, MA, 1997.

[116] D L Ringach, M J Hawken M J, and R Shapley. Receptive field structure of neurons in monkey primary visual cortex revealed by stimulation with natural image sequences. *Journal of Vision*, 2:12–24, 2002.

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

[118] J Risasanen. Modelling by shortest data description. *Automatica*, 11:465–471, 1978.

[119] A Robert and J L Eriksson. A composite model of the auditory periphery for simulating responses to complex sounds. *J. Acoust. Soc. Am.*, 106(1):1852–1864, 1999.

[120] J E Rose, D J Anderson, and J F Brugge. Some effects of stimulus intensity on response of auditory nerve fibers in the squirell monkey. *Journal Neurophys.*, 34:685–699, 1971.

[121] S Roweis and L Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326, 2000.

[122] D L Ruderman and W Bialek. Statistics of natural images: Scaling in the woods. *Phys. Rev. Letters*, 73(6):814–817, 1994.

[123] M A Ruggero. Physiology and coding of sound in the auditory nerve. In *The Mammalian Auditory Pathway: Neurophysiology*. Springer-Verlag, 1992.

[124] M B Sachs and P J Abbas. Rate versus level functions for auditory-nerve fibers in cats: One-burst stimuli. *J. Acoust. Soc. Am.*, 56(6):1836–1847, 1974.

[125] M B Sachs and P J Abbas. Phenomenological model for two-tone suppression. *J. Acoust. Soc. Am.*, 60(5):1157–1163, 1976.

[126] M B Sachs, R L Winslow, and B H.A. Sokolowski. A computational model for rate-level functions from cat auditory-nerve fibers. *Hearing Research*, 41:61–70, 1989.

[127] T D Sanger. Optimal unsupervised learning in a single-layer network. *Neural Networks*, 2:459–473, 1989.

[128] M P Sceniak, D L Ringach, M J Hawken, and R Shapley. Contrast's effect on spatial summation by macaque v1 neurons. *Nat Neurosci.*, 8:733–739, 1999.

[129] O Schwartz and E P Simoncelli. Natural signal statistics and sensory gain control. *Nature Neuroscience*, 4(8):819–825, August 2001.

[130] G Sclar and R D Freeman. Orientation selectivity in the cat's striate cortex is invariant with stimulus contrast. *Exp. Brain Res.*, 46:457–461, 1987.

[131] G Sclar, J H R Maunsell, and P Lennie. Coding of image contrast in central visual pathways of the macaque monkey. *Vision Research*, 30:1–10, 1990.

[132] C Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, July 1948.

[133] C E Shannon and W Weaver. *Nonlinear problems in random theory*. Urbana: University of Illinois Press, 1949.

[134] R Shapley and C Enroth-Cugell. Visual adaptation and retinal gain control. *Progress in Retinal Research*, 3:263–346, 1984.

[135] R M Shapley and J D Victor. The effect of contrast on the transfer properties of cat retinal ganglion cells. *J. Physiol. (Lond)*, 285:275–298, 1978.

[136] R M Shapley and J D Victor. Nonlinear spatial summation and the contrast gain control of cat retinal ganglion cells. *Journal Physiology*, 290:141–161, 1979.

[137] R M Shapley and J D Victor. How the contrast gain control modifies the frequency response of cat retinal ganglion cells. *J. Physiol. (Lond)*, 318:161–171, 1981.

[138] M Sigman, G A Cecchi, C D Gilbert, and M O Magnasco. On a common circle: Natural scenes and gestalt rules. *PNAS*, 98(4):1935–1940, 2001.

[139] A M Sillito, K L Grieve, H E Jones, J Cudeiro, and J Davis. Visual cortical mechanisms detecting focal orientation discontinuities. *Nature*, 378:492–496, 1995.

[140] E Simoncelli and B Olshausen. Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24:1193–1216, May 2001.

[141] E P Simoncelli. Statistical models for images: Compression, restoration and synthesis. In *31st Asilomar Conf Signals, Systems and Computers*, pages 673–678, Pacific Grove, CA, November 1997.

[142] E P Simoncelli, W T Freeman, E H Adelson, and D J Heeger. Shiftable multiscale transforms. *IEEE Trans Information Theory*, 38(2):587–607, March 1992. Special Issue on Wavelets.

[143] E P Simoncelli and D J Heeger. A model of neuronal responses in visual area MT. *Vision Research*, 38(5):743–761, March 1998.

[144] E P Simoncelli and O Schwartz. Image statistics and cortical normalization models. In M. S. Kearns, S. A. Solla, and D. A. Cohn, editors, *Adv. Neural Information Processing Systems*, volume 11, pages 153–159, Cambridge, MA, 1999. MIT Press.

[145] B C Skottun, R D Freeman, G Sclar, I Ohzawa, and R D Freeman. The effects of contrast on visual orientation and spatial frequency discrimination: a comparison of single cells and behavior. *J. Neurophysiol.*, 57:773–786, 1987.

[146] M Slaney. An efficient implementation of the patterson and holdworth auditory filter bank. *Apple Technical Report 35*, 1993.

[147] S M Smirnakis, M J Berry, M J Warland, W Bialek, and M Meister. Adaptation of retinal processing to image contrast and spatial scale. *Nature*, 386:69–73, 1997.

[148] R J Solomonoff. A formal theory of inductive inference. *Information and Control*, 7:1–22, 1964.

[149] J B Tenenbaum, V De Silva, and J C Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323, 2000.

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

[151] N Tishby, F Pereira, and W Bialek. The information bottleneck method. Technical report, 1999.

[152] J Touryan, B Lau, and Y Dan. Nonlinear analysis of complex cells in cat visual cortex [abstract]. In *Journal of Vision (VSS)*, volume 1, page 34, 2000.

[153] D Tranchina, J Gordon, and R N Shapley. Retinal light adaptation: Evidence for a feedback mechanism. *Nature*, 310:314–316, 1984.

[154] J B Troy and J G Robson. Steady discharges of x and y retinal ganglion cells of cat under photopic illuminance. *Visual Neuroscience*, 9:535–553, 1992.

[155] K De Valois and R Tootell. Spatial frequency specific inhibition in cat striate cortex cells. *J. Physiol (Lond)*, 336:359–376, 1983.

[156] J H van Hateren and A van der Schaaf. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. R. Soc. Lond. B*, pages 359–366, 1998.

[157] J Victor. Nonlinear processes in spatial vision: Analysis with stochastic visual textures. In *Inv. Opthalm. Vis. Sci.*, volume 29, page 118, 1988.

[158] J D Victor. The dynamics of the cat retinal X cell centre. *J. Physiol.*, 386:219–246, 1987.

[159] W E Vinje and J L Gallant. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, 287, Feb 2000.

[160] V Volterra. *Theory of Functionals and of Integro and Integro-Differential Equations*. Dover, New York, 1959.

[161] H von Helmholtz. *Physiological Optics*, 3, 1925.

[162] M J Wainwright, O Schwartz, and E P Simoncelli. Natural image statistics and divisive normalization: Modeling nonlinearity and adaptation in cortical neurons. In R Rao, B Olshausen, and M Lewicki, editors, *Probabilistic Models of the Brain: Perception and Neural Function*, chapter 10, pages 203–222. MIT Press, 2002.

[163] M J Wainwright and E P Simoncelli. Explaining adaptation in V1 neurons with a statistically optimized normalization model. In *Investigative Opthalmology and Visual Science Supplement (ARVO)*, volume 40, pages S–573, May 1999.

[164] M J Wainwright and E P Simoncelli. Scale mixtures of Gaussians and the statistics of natural images. In S. A. Solla, T. K. Leen, and K.-R. Müller, editors, *Adv. Neural Information Processing Systems*, volume 12, pages 855–861, Cambridge, MA, May 2000. MIT Press.

[165] M J Wainwright, E P Simoncelli, and A S Willsky. Random cascades on wavelet trees and their use in modeling and analyzing natural imagery. *Applied and Computational Harmonic Analysis*, 11(1):89–123, July 2001. Special issue on wavelet applications.

[166] B A Wandell. *Foundations of Vision*. Sinauer Associates, Sunderland, MA, 1995.

[167] K Wang and S Shamma. Self-normalization and noise-robustness in early auditory representations. In *IEEE Trans. Speech and Audio Proc.*, volume 2, pages 421–435, 1994.

[168] N Wiener. *Nonlinear problems in random theory*. New York: John Wiley, 1958.

[169] G K Yates. Basilar membrane nonlinearity and its influence on auditory nerve rate-intensity functions. *Hearing Research*, 50:145–162, 1990.

[170] R Zemel. A minimum description length framework for unsupervised learning. Technical Report CRG-TR-93-2, University of Toronto, 1994.

[171] C Zetche. Sparse coding: the link between low level vision and associative memory. In *Parallel Processing in Neural Systems and Computing*, 1990.

[172] C Zetche and B Wegmann. Invited address: nonlinear aspects of primary vision: entropy reduction beyond decorrelation. *SID Int. Sympos. Digest Techn. Papers*, 24:933–936, 1993.

[173] H B Zhao and J Santos-Sacchi. Auditory collusion and a coupled couple of outer hair cells. *Nature*, 399(6734):359–362, 1999.