INTRODUCTION

VISUAL CODING

Vision science seeks to understand how the visual system gathers, interprets, and represents information. The complexity of the human visual system (and the brain more generally) is challenging to analyze partly because of the difficulty in deciphering the neural code, the language in which neurons communicate. In studying neural processing, the strategy of neurophysiologists was first to study lesions and ablations, wherein areas of the brain had been damaged or removed, in order to see how perception and function were affected. In the past 40 years, the advent of single-cell recordings has made it possible to trace the physiology, anatomy and processing of each unit and cell type along a sensory stream. For Hubel and Weisel (1962), who recorded electrical signals from individual cat neurons and mapped out those cells' electrical response to spatial patterns, the goal was to understand the system's constituent parts. We have learned a great deal about the visual system from this approach. Studies of visual coding have examined the processing of individual units and they have determined the response properties of many cells in the visual system. But there remain questions about coding. How does the visual system extract useful information from visual stimuli? What evolutionary forces shaped the structure of the visual system?

Simoncelli and Olshausen (2001) describe three broad factors that influence the evolution of visual systems: the uses to which the organism puts its visual system; the constraints imposed by neuronal structure and architecture; and the visual environment in which the organism lives. While the first two factors

are no doubt important, we examine the third factor, the environment, and we are interested specifically in the statistical structure of the visual world. This area of study has grown markedly in the past decade, with the help of physicists, computer scientists, physiologists and a host of other researchers working in an interdisciplinary context.

The general strategy in answering teleological questions in biology is to assume that nature is parsimonious in its designs. Specifically, the visual system is thought to use an economical strategy to encode information. The environment has been long been implicated in the adaptations of sensory modalities (Attneave, 1954, Barlow, 1961) and the visual system in particular is thought to encode visual information using a representation that has less redundancy than the stimuli. But what are the redundant features of the visual world? And how could the visual system achieve an efficient representation for certain types of redundancy?

CHAPTER 1

THE CODING OF NATURAL SCENES

Natural scenes share a number of statistical properties. Humans can readily identify natural scenes and distinguish one scene from another. Given random dot patterns, though, humans recognize immediately that such images are not natural scenes and that it is quite difficult to distinguish between individual random patterns. This suggests that natural images share some unique statistical properties to which the visual system is particularly attuned.

A large and growing collection of studies has examined the link between coding strategies and the statistics of natural scene images, both by examining the statistics of neural responses to natural stimuli (Laughlin, 1981; Rieke et al., 1995; Dan et al., 1996; Baddeley et al. 1998; Vinje and Gallant, 2000); and by using information theory, together with statistical analysis of the natural stimuli themselves, to predict the goals of sensory coding (Field, 1987; Foldiak, 1990; Atick and Redlich, 1990; Atick, 1992; Olshausen and Field, 1996; Bell and Sejnowski, 1997; van Hateren, 1992; van Hateren and van der Schaaf, 1998). Our studies fall into the latter category: we are interested in the statistics that are shared by natural scenes and in how the visual system could reduce such redundancies in its representation.

Consider a state space of possible images, say the space of all possible 512 × 512 pixel images, each with 256 grey levels (8-bit). There exist $(2^8)^{262144} = 2^{2097152}$ such images. Images of our visual world comprise a special distribution in this high-dimensional space, with statistics that are distinct from other regions of this space. We seek the set of statistics that distinguishes natural

scenes from the immense number of possible scenes, and we use this information to put constraints on the types of processing the human visual system performs. Our approach is to look for ways to take advantage of the redundant information in natural scenes and then relate these strategies to their possible neural substrates.

It is unlikely that the objectives of visual coding are dictated exclusively by the need to match the statistics of scenes with an appropriately efficient neural code. The "answer," as is so often the case, is somewhere in the middle, a compromise between making the best possible use of the predictable statistics of scenes and conforming to the rules of biochemistry, ontogeny, cellular structure and neural architecture. But we would like to see how far we can get in describing the goals of visual coding using efficiency arguments. Besides the biological implications of finding the statistical redundancy of scenes, there are direct applications for this knowledge in computing, for image compression and search, and for object recognition.

1.1 Shared Statistics

This thesis deals with the analysis of static natural scene images.[1] Our work has analogs in astronomy, where images of astronomical objects are studied with regard to their spectra among other things. We will often talk about spatial frequency, a measure of the number of cycles per degree in an image.

[1] Studies of natural scene statistics involve collecting hundreds or thousands of images in a variety of environments and lighting conditions. The images must all be calibrated such that their pixel values are a linear measure of the luminance at a particular location in the world. We will return to the process of calibration later but for now it is important to understand that properly calibrated images have pixel values that are independent of the optical properties of the camera or of the film.

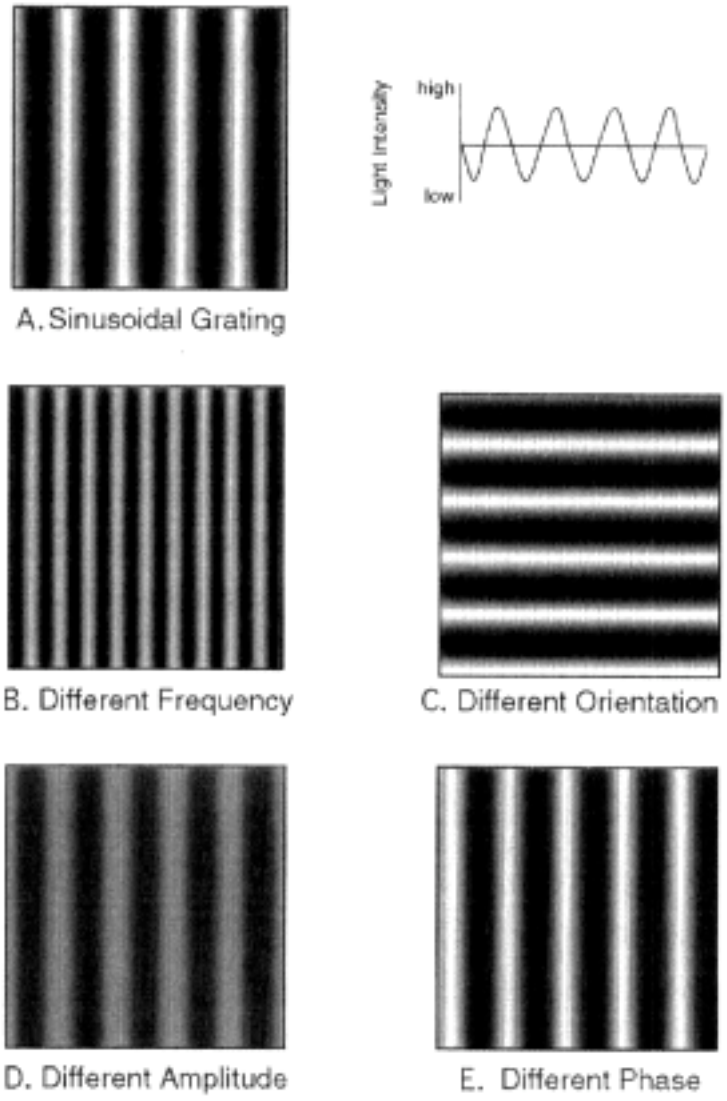


Figure 1.1. Two-dimensional sinusoidal gratings (A) are the primitives of spatial frequency. The grating in (B) has a higher spatial frequency than (A), (C) has a different orientation, (D) has a different amplitude and (E) has a different phase. (Palmer, 1999)

Spatial frequency is based on a Fourier decomposition of an image into two-dimensional gratings given by

(1.1.1) $$F(u,v) = \iint f(x,y)e^{-i2\pi(ux+vy)}\mathrm{d}x\mathrm{d}y$$

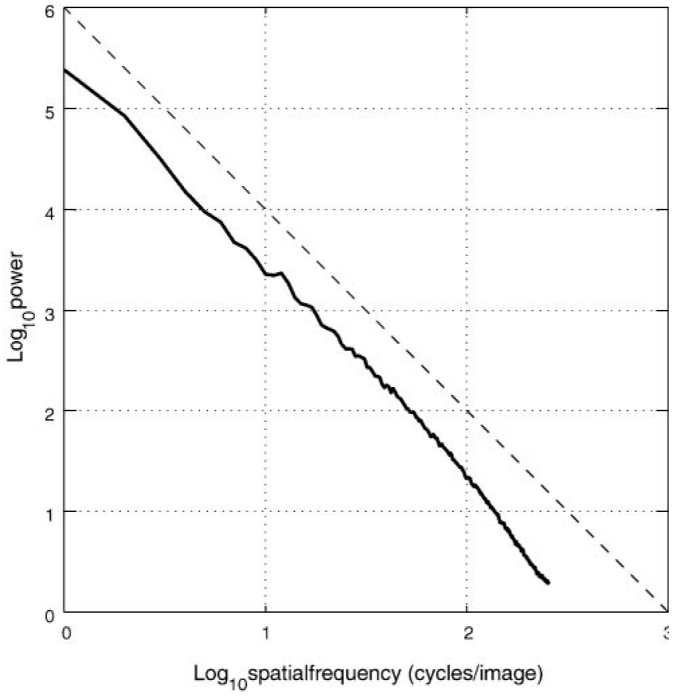
where x and y are the two spatial directions. The intensity along one dimension of the Fourier grating changes according to a sine function and the intensity along the orthogonal direction is constant (Fig. 1.1). In general, the Fourier power spectrum is defined as the square of the Fourier transform. The one-dimensional power spectrum of the two-dimensional Fourier amplitude spectrum (for a given decomposition, the amplitude spectrum gives the amplitude at each frequency and orientation of the basis) is usually found by taking the square of the rotational average of the discretized amplitude spectrum. The power spectrum is also referred to as the structure factor in studies of x-ray crystallography.

Some early researchers tried to quantify the redundancy in television signals (Kretzmer, 1952). Field (1987), Burton and Moorhead (1987) and Tolhurst et al. (1992) later decomposed natural scenes into two-dimensional Fourier gratings and found that the Fourier power spectra fall with spatial frequency as $1/f^n$ where $n \sim 2$. That is, a randomly selected image of our visual world is likely to have a Fourier power spectrum that falls as $1/f^2$ (Fig. 1.2a). Evidently, one thing that distinguishes natural scene images from other images is that natural scenes have similar power spectra. Another way of describing the same statistical property is with the autocorrelation function,

$$R(a) = \lim_{A\to\infty} \frac{1}{2A} \int_{-A}^{A} g(a')g(a+a')da' \qquad (1.1.2)$$

which measures how similar in luminance two points are as the points get farther away from each other (Fig. 1.3). Here, R gives the correlation between two points that are separated by a distance of a. When normalized, $R(0) = 1$.

a.)



b.)

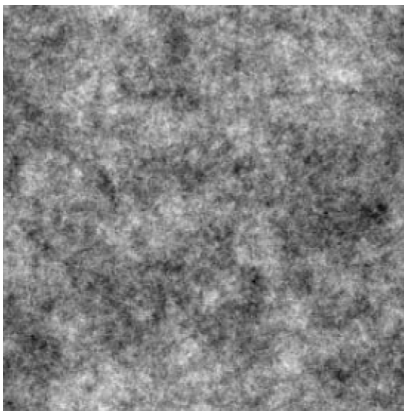


Figure 1.2. The Fourier frequency spectrum (a) for a collection of natural images and (b) an image constructed of random dots whose spatial frequency power spectrum is the same as that for natural images. (Field, 1994)

Mathematically, the power spectrum and the autocorrelation function express the same statistical regularity because the Fourier transform of the

power spectrum is the autocorrelation function. But for scale-invariant images, the functional form of the autocorrelation function depends on the sample size (pixel size) one is considering so we generally discuss the power spectrum and the phase spectrum of images.

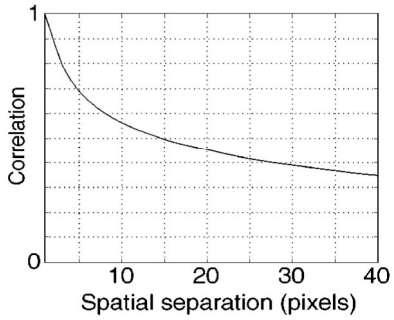


Figure 1.3. The autocorrelation function for a natural image shows that the likelihood that neighboring points are correlated falls steeply with separation. (Olshausen and Simoncelli, 2001)

However, a $1/f^2$ power spectrum is not sufficient to distinguish natural scenes from other possible images. In Fig. 1.2b we show a random dot pattern whose power spectrum falls as $1/f^2$; clearly, this image is not a natural scene. The local structure characteristic of natural scenes can only be captured by higher-order statistics.

Spatiotemporal statistics of natural scenes have also been shown to be predictable. The spatiotemporal correlations are calculated using

$$R(x,t) = \frac{1}{L^2 T} \int_0^{L^2} \int_0^{T} S(x',t') S(x+x',t+t') \mathrm{d}x' \mathrm{d}t', \tag{1.1.3}$$

for which S is the luminance for a particular location x' and time t' and separation between points of x and t, L^2 is the spatial area and T is the

temporal length. Temporal correlations refer to changes at a given spatial location over time. The spatiotemporal power spectrum would be:

$$R(f,\omega) = \frac{1}{L^2 T} \int_0^{L^2} \int_0^{T} R(x,t) e^{2\pi i (fx+\omega t)} \mathrm{d}x \mathrm{d}t . \tag{1.1.4}$$

Dong and Atick (1995a) analyzed commercial movies and home movies (which were found to have largely the same properties) in terms of their spatiotemporal power spectra. They found that the slope of the spatial power spectrum increases (becomes less negative) as temporal frequency is increased and that the slope of the temporal power spectrum increases (becomes less negative) as spatial frequency is increased.[2] The power spectrum therefore cannot be separated into purely spatial and temporal parts. A power-law distribution of velocities is proposed as a way to account for the spatiotemporal power spectrum. The spatiotemporal power spectrum can be described by the function

$$R(f,\omega) = \frac{K}{f^{m+1}} F\left(\frac{\omega}{f}\right), \tag{1.1.5}$$

where K is a constant, f is frequency, $\omega = f \cdot v/r$, m is an exponent (here, it is 2.3, the slope of the spatial power spectrum), F is a function of the velocity distribution P of objects in the world at a distance r:

$$F\left(\frac{\omega}{f}\right) = \int P\left(\frac{vr}{f}\right) r \mathrm{d}r . \tag{1.1.6}$$

[2] And in agreement with Field (1987), Burton and Moorhead (1987) and Ruderman and Bialek (1994), snapshots of these movies exhibited a $1/f^m$ spatial frequency power spectrum, where $m = 2.3$.

That is, the spatiotemporal power spectrum can be described as solely a function of $1/f$.

Another quality shared by natural scenes is that edges of objects tend to be continuous. Quantifying edge continuity is difficult but one group (Geisler et al., 2001) used psychophysical edge detection, in which subjects traced the edges of static scenes, and analysis was done on the tracings. Geisler et al. found that nearby edges (within 1.23 degrees of visual angle) tend to be parallel, though not necessarily collinear. The study also showed that the most likely tilt orientation difference between two edges was likely to be small (i.e., they were likely to be co-circular).

1.2 Efficient Coding

Given these statistical regularities in the input to the visual system, the idea is to use efficiency arguments to find optimal coding solutions and then look at how such solutions could be manifest in the brain. What do we mean by efficient? Efficiency can be defined in a number of ways but our definition relates to the amount of information that a population of neurons is able to carry. The hypothesis that the visual system is constructed in such a way as to maximize information transmission began with Attneave (1954) and Barlow (1961) although Barlow cites Mach (1886) and Pearson (1892) as having discussed these ideas. Attneave stressed the redundancy present in the visual world. Barlow argued that efficient visual coding strategies could take advantage of the statistical likelihood of events and they could use something akin to a Huffman code, wherein the most common event is coded with the shortest code word. In the last decade, work by Field (1987, 1994), Atick and

Redlich (1990), van Hateren (1992) and Rieke et al. (1995) established many of the principles of efficient coding although a great deal of subsequent work has expanded our knowledge of coding efficiency (see Simoncelli and Olshausen, 2001, for a review).

There are two senses in which the visual system is believed to be governed by the need for efficiency: Individual neurons are thought to maximize the use of their information-carrying ability; and the system as a whole should be well-suited to the likely statistics of the input in order to reduce redundancy in its representation (Simoncelli, 2003). This thesis uses efficiency arguments to determine possible ways in which retinal processing is optimized to represent spatial information about natural scenes. We give a brief overview of early visual processing in the next chapter.

CHAPTER 2

THE RETINA AND LATERAL GENICULATE NUCLEUS

2.1 Physiology

Neuronal connections in the retina are organized in layers (Fig. 2.1). Starting from the surface that is radially most distant from the center of the eyeball, the outer nuclear layer contains photoreceptors, whose tips lie in the pigment epithelium, the vascular layer that lines the sclera. The outer plexiform layer contains synaptic connections from photoreceptors to horizontal and bipolar cells. The somas (cell bodies) of these cells and of amacrine cells lie in the inner nuclear layer. Amacrine and bipolar cells project to the inner plexiform layer where they connect to ganglion cell dendrites. Patterns of connection within this layer have been compared to a club sandwich, with different amacrine and bipolar cell types projecting to specific layers (Masland, 2001).

All visual information leaving the retina passes through the ganglion cells, whose axons comprise the optic nerve (Fig. 2.2). The optic nerves from the two eyes cross at the optic chiasm, where the left and right visual fields of the two eyes are segregated. Most of these projections continue on to the midbrain on separate pathways. From the optic chiasm, a few ganglion cell axons synapse to the superior colliculus (which is thought to control involuntary eye reflexes, such as pupil dilation and constriction) but most go to the lateral geniculate nucleus (LGN) of the midbrain. The LGN synapses to the V1 area of visual

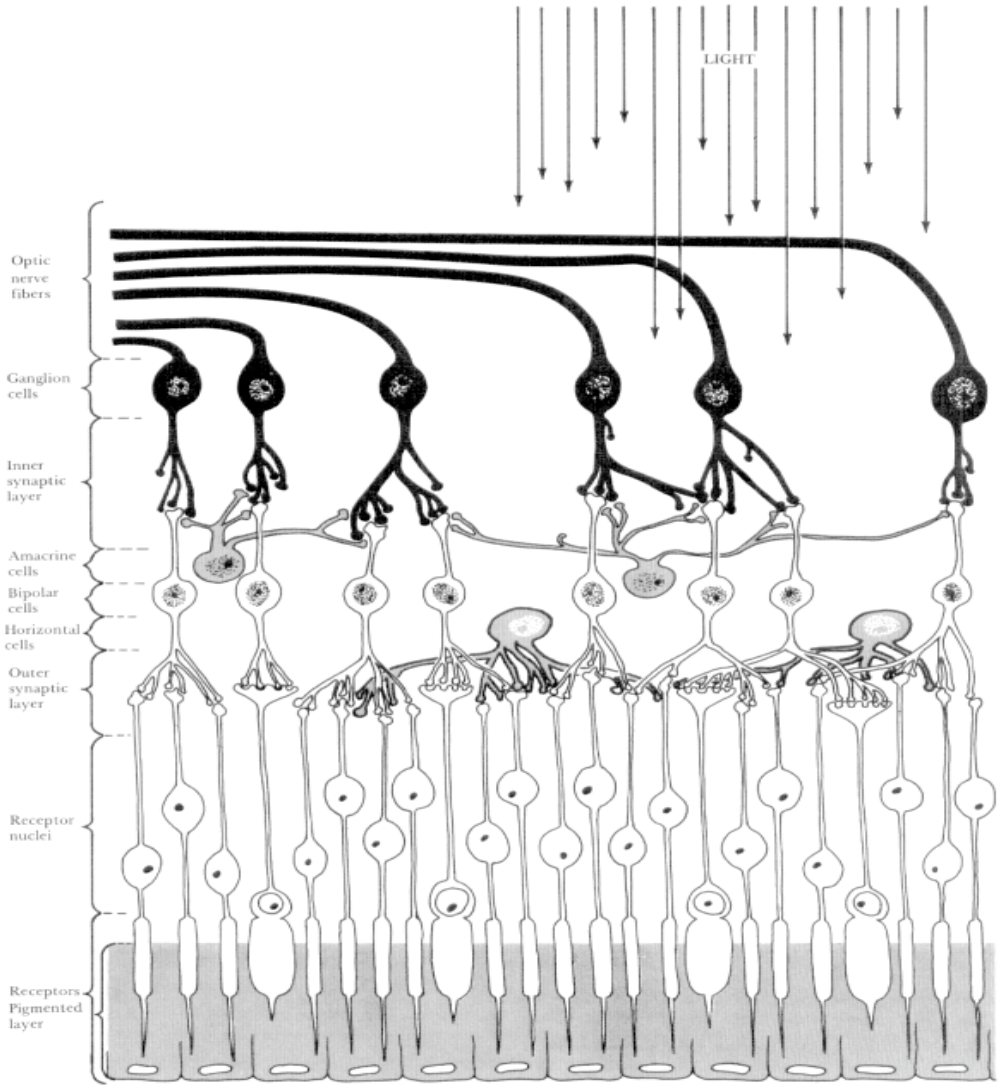


Figure 2.1. Schematic diagram of the retina. Light strikes the photoreceptors after passing through all the other retinal processing cells and blood vessels. The photoreceptors tips penetrate the pigment epithelium where photopigments are replenished. After transduction, photoreceptors communicate with horizontal cells (largely responsible for lateral inhibition) and bipolar cells. Bipolar cells project to ganglion cells and amacrine cells connect ganglion cells to other ganglion cells. (Palmer, 1999)

cortex. More than half of the inputs to LGN originate in cortex, meaning that there is a great deal of feedback in to LGN. The LGN is thought to modulate the weights of the connections from afferent optic nerve fibers with signals

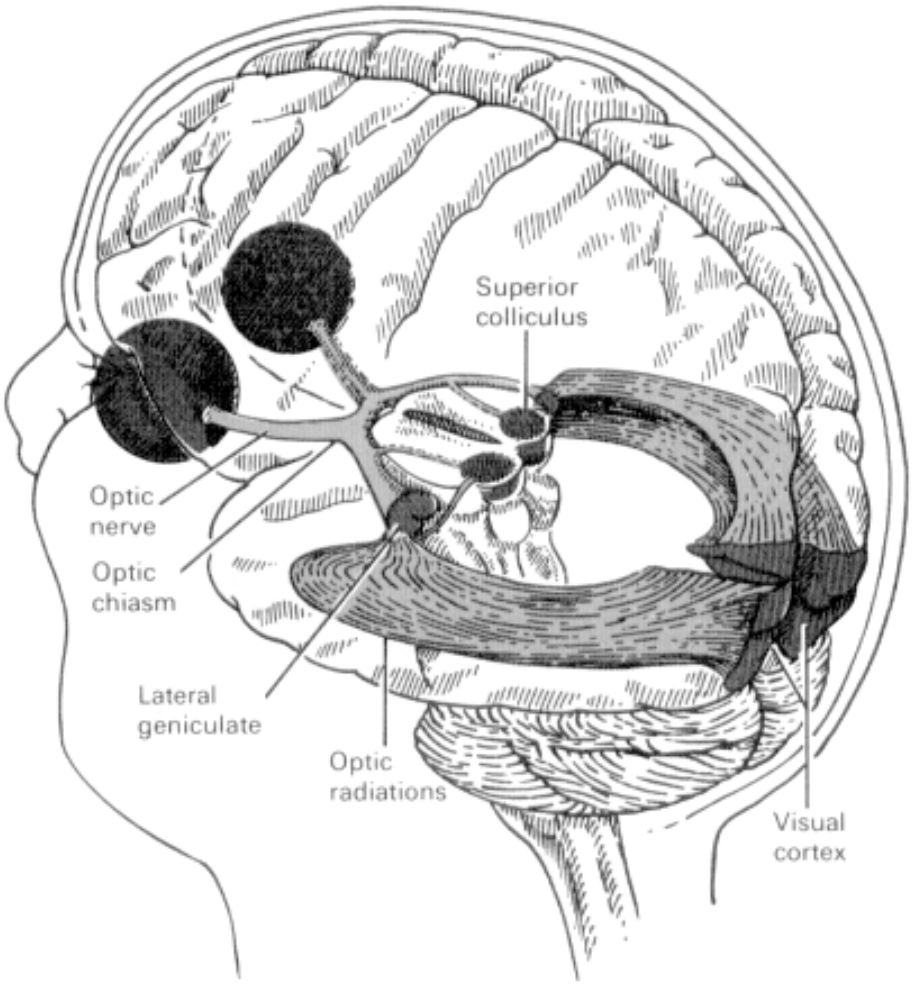


Figure 2.2. Connections among major anatomic units in the visual system. Visual information is passed from the retinae through the optic chiasm (which segregates the left and right hemifields of vision from each eye. The optic nerve projects to the superior colliculus and to the lateral geniculate nucleus, which has projections into visual cortex. (Palmer, 1999)

from cortex. Though fewer than half of the inputs to LGN are from the retina, these connections are often given more weight than those from cortex. Both LGN and V1 preserve topologically equivalent representations of the retinal image though information from the fovea occupies a larger neural area in these units than on the retina (Kandel et al., 2000). This thesis focuses on the

retina in order to characterize the input to V1 and higher areas. Once we understand the goals of retinal processing, we may be able to say more about cortical processing. A general overview of visual processing, including cortical processing is presented in Appendix A.

2.2 Receptive Fields

The two-dimensional arrangement of light stimuli that optimally causes a neuron to fire is called the receptive field of the cell. Mapping a receptive field involves recording current from a particular cell as lights of different shapes are shown on the retina (Hartline, 1940; Kuffler, 1953; Barlow and Levick, 1965). The receptive field is reconstructed using the measured sensitivity at a given location in space. Most retinal ganglion cells are found to have a center-surround organization: they have an inhibitory annulus surrounding an excitatory center or an excitatory annulus and an inhibitory center (Fig. 2.3). Receptive fields of cells in the LGN tend to have a similar center-surround organization. We discuss receptive fields of these units in detail below.
In retinal ganglion cells, the receptive field is the location of the excitatory and inhibitory connections in space. The mapping is frequently a dome-shape that encloses the dendrite tips, although there are other shapes as well. The ganglion cell receptive field is influenced by photoreceptor-horizontal cell interactions; by interactions with bipolar cells; and by the density or geometry of the dendritic tree arbor (Brown et al., 2000). Horizontal cells, the interneurons that connect photoreceptors to each other, play an important role

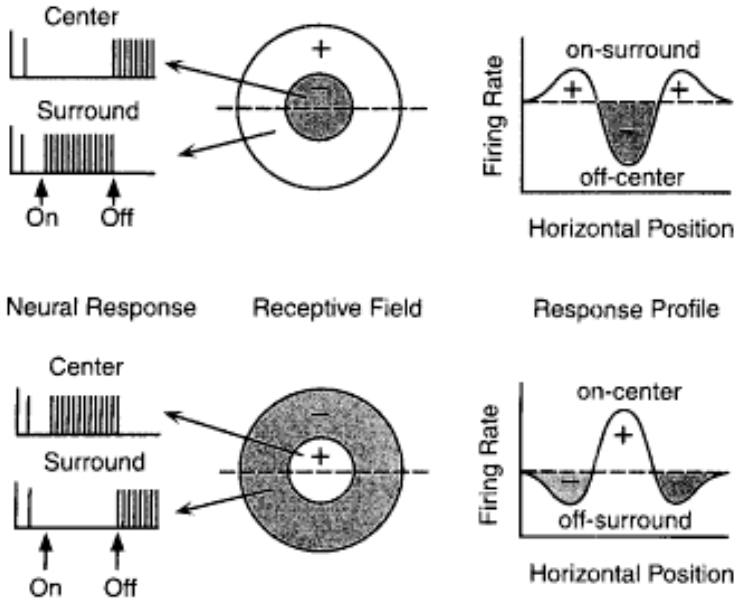


Figure 2.3. Center-surround receptive fields. For the Off-center cell (above), the center is inhibitory and the surround is excitatory. For the On-center cell, the center is excitatory and the surround is inhibitory. (Palmer, 1999)

in creating areas of inhibition. If a light is shown on one photoreceptor, a horizontal cell leading away from it will inhibit other photoreceptors in the neighborhood (Fig. 2.4). This interaction, called lateral inhibition, is seen in

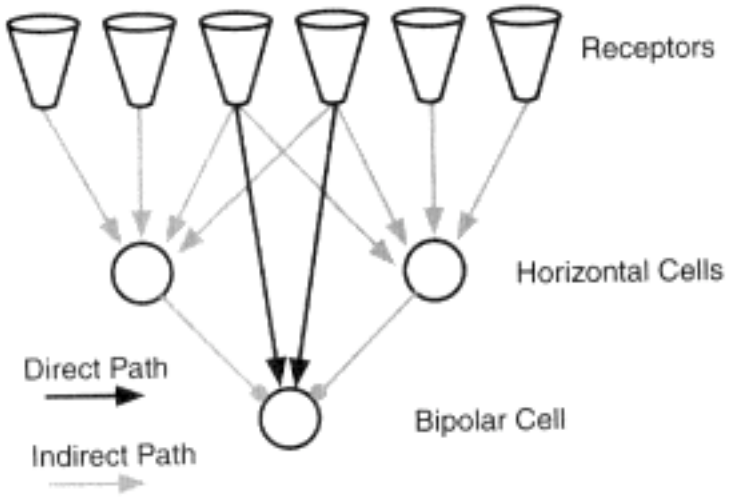


Figure 2.4. A greatly simplified wiring diagram that models center-surround receptive field structure. (Palmer, 1999)

many units in different sensory streams across species. The specific mechanism of lateral inhibition found in mammalian retinae is found to be highly conserved[3] across species: interactions among photoreceptors that are mediated by horizontal cells, and interactions among ganglion cells that are mediated by bipolar and amacrine cells are found across mammalian species (Finlay et al., 2004). Moreover, virtually every functional and structural feature of the mammalian retina as a whole are found to be the same across species. In particular, there is a high degree of conservation in the design of the optics (the cornea and lens that project a visual image onto the retina); in the presence of rod and cone photoreceptors, horizontal, amacrine, bipolar and ganglion cells; in the layering of connections within the retina; and in other features such as oculomotor organization (Finlay et al., 2004).

[3] A trait or structure that is conserved is one that remains unchanged through a period of evolutionary time.

CHAPTER 3

INFORMATION AND EFFICEINCY

3.1 Physics and the Brain

Our goal is to develop a model of the response of retinal ganglion cells to light stimuli, a model that could conceivably generate the observed receptive fields based on simple spatial processing rules. In physics, the general strategy for modeling systems is to search out the handful of physical observables that describes the majority of phenomena. Using dimensionality reduction arguments based largely on conservation laws, classical mechanics, condensed matter physics, electrodynamics, atomic physics and quantum mechanics are hugely successful at describing the physical world in this way. But though energy, charge, momentum, etc. are conserved in the brain, these conserved quantities put only very general restrictions on how coding could take place. Theories of visual coding thus require input from a number of other disciplines, including physiology, evolutionary biology, biochemistry and many others. But statistical mechanics offers some help because it deals with the probabilities of events given their possible outcomes. In particular, the idea of the information content of signals is within the purview of statistical mechanics. Thanks to Shannon (1949), the information content H of a signal is quantifiable:

$$H = -K\sum_{i=1}^{m} p_i \log p_i \tag{3.1.1}$$

where K is a positive constant, m is the number of "letters" or coding units in

the code, and p_i is the probability of the ith letter. This metric is always positive and it ranges from a signal composed of a string of one letter only, which contains zero information, to a signal composed of a random string letters, which contains the maximum amount of information as given by Eqn. 3.1.1.

We are interested in finding the redundancies that are exploited by the visual system to make an efficient code. Kersten (1987) tried to quantify the perceptual information content of natural scenes. Kersten's technique involved a Shannon guessing game, wherein subjects (Kersten and another researcher) replaced missing pixels in natural scene images according to their expectations of what the pixel intensity should be at that point. This method probes human knowledge of scene redundancy. The amount of redundancy, which is what gives rise to this reduction in information, is determined using the following:

(3.1.2) $$Redundancy = 1 - \frac{H}{m}$$

where H is the entropy as defined above and m is the maximum value of entropy. The probability of success at this task was translated into a Shannon entropy per pixel at different resolutions. Kersten found that natural scenes that were displayed with $m = 4$ bits per pixel (16 grey levels), typically have a perceptual information content of 1.4 bits per pixel. But a model that took a simple average of nearest-neighbor pixel values accurately predicted human responses, which means that Kersten's method mainly exploits pairwise correlations and not higher-order redundancies. Human subjects did outperform the model when the images had noise added to them. But Kersten's experimental method does not necessarily capture the true

information content of images.[4]

In addition to scene entropy, neural spike entropy—i.e., the amount of information carried by an action potential or spike-train—has been given bounds. Strong et al. (1998) find that a neuron can transmit 90 bits per second assuming that the code represents information by the mean firing rate during a given time period. Because it is possible that the *pattern* (not just the rate) of spikes contains information, one cannot yet fully specify the information content of neural signals (Victor, 2002).

3.2 Principal Components Analysis

As we mentioned above, physicists generally look for low-dimensional solutions to problems. Given a system—take the simple harmonic oscillator (SHO) for example—we can describe its possible behavior precisely by finding the eigenvectors and eigenvalues of its modes of oscillation. Moreover, the SHO model explains a wealth of phenomena in many settings from objects on springs to molecules in crystals. In neuroscience, we want to describe the relationship between stimulus and neural response and a natural approach is to look for a low-dimensional description of this relationship. Solutions based on finding a small number of eigenvectors that accurately represent a distribution of neural responses are called *compact codes*. With a compact code, a small number of vectors are used to efficiently code an input with higher dimensionality (Field, 1994). Can retinal processing be described with a small set of observables? The general approach in answering this question is to

[4] Work ongoing in the Field lab uses small image patches in a similar guessing game as a way of establishing the full entropy of images. This technique may capture more of the visual system's higher-order redundancy reduction.

invoke the efficient coding hypothesis,[5] which holds that individual neurons and sets of neurons should make maximal use of their firing capabilities. For example, if an individual neuron can only fire with a positive, real value between 0 and M, and the cost of firing at every value is the same, the optimal distribution is uniform. If we had a constraint on our response x of the form

(3.2.1) $$\langle \square(x) \rangle = c$$

where the average value of the constraint $\square(x)$ is a constant c, the maximum entropy distribution of responses $P(x)$ would be an exponential.

(3.2.2) $$P(x) \quad e^{\square\square\square(x)}$$

For a population of neurons, the efficient coding hypothesis suggests that each neuron should be maximizing information transmission and that the neuron's response should be statistically independent from other neurons. That is, there should be an equal likelihood of each possible combination of neural responses across the population. The conditional probability of one neuron's response given the responses of other neurons is therefore a fixed distribution. If this condition is not met (i.e., we can predict the response of one neuron given the response of a different neuron) there is still redundancy in the representation. (Simoncelli and Olshausen, 2001).

[5] This term refers to the notion discussed earlier that the visual system has found an efficient representation of the environment. The umbrella term "efficient coding hypothesis" was deployed recently by Simoncelli (2003) to describe the wealth of work that has been done in this vein.

We can put further restrictions on a possible coding strategy by demanding that it be a linear decomposition (equivalently, a projection onto an orthogonal basis) and that it only consider second-order (covariance) statistics. Solutions of this kind can be obtained with *principal components analysis*, also called the Karhunen-Loève transform. The goal of this method is to find the n-dimensional basis whose axes are aligned with the top n orthogonal directions with the highest variance. This strategy reduces redundancy by not coding regions of state space where there is a low likelihood of finding data (Field, 1994). Suppose we have a distribution whose covariance matrix is diagonal. This distribution will be a Gaussian ellipsoid whose axes are aligned with the coordinate space. If the distribution does not have a diagonal covariance matrix, we can change coordinates into a space

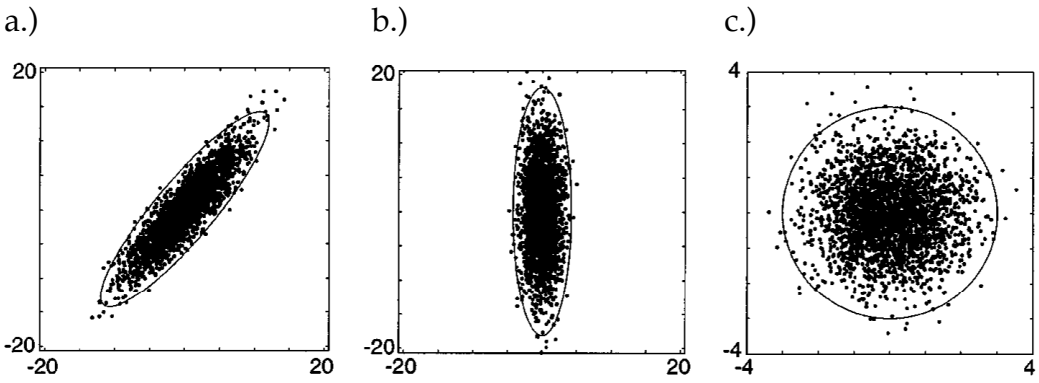


Figure 3.1. Principal components analysis: (a) shows the original data. In (b), we have identified the principal axis and rotated our basis to be parallel with it. (c) shows sphered or whitened data, in which we have removed the correlations making a univariate Gaussian distribution for each component. Sphering (also called whitening) is not strictly part of PCA. (Simoncelli and Olshausen, 2001)

where the axes are aligned with the principal components of the data (Fig. 3.1). PCA finds the directions in state space along which the data have the

greatest variance and rotates the coordinate system to be aligned with these directions. Once the basis vectors of the space are aligned with the directions of highest variance in the data, one typically normalizes the variance in each direction to one (also called whitening or "sphering" the data), although this is not part of PCA.

For example, the data shown in Fig. 3.1 could represent the intensity of two neighboring pixels of a natural scene image, with the intensity of one pixel plotted on the x-axis and the other pixel on the y-axis (Fig. 3.1a). Clearly, the data are correlated; we can remove this redundancy by rotating our coordinate system (Fig. 3.1b). In Fig. 3.1c, we show that after performing PCA, the data have been whitened.

Mathematically, PCA works in the following way: We want to transform the point x into the point c in a new coordinate system. If V is the matrix whose columns are the eigenvectors of the covariance matrix, we take the transpose V^T and multiply it by the matrix $\boldsymbol{z}$ that is defined as

$$\boldsymbol{z} = \boldsymbol{x} - m \tag{3.2.3}$$

where m is the mean. This method projects the point $\boldsymbol{x}$ into $\boldsymbol{c}$ by projecting $\boldsymbol{z}$ onto the principal components axes.

When data are stationary (i.e., ergodic) with respect to space, projecting into a Fourier basis gives a solution that is degenerate with the principal components axes.[6] That is, when we assume translation invariance in images, we can achieve decorrelation by transforming to the Fourier basis. The

[6] Natural scene statistics are not strictly stationary however: scenes are likely to be composed of bright sky and darker ground, separated by a horizon. But to a first approximation, scene statistics can be defined as stationary.

pairwise correlations can then be described by the 1-dimensional power spectrum (or, equivalently, its Fourier transform, the autocorrelation function) taken as a rotational average over Fourier space. The first principal component axis codes the direction with the highest variance, the second principal component codes the direction with the next highest variance and so on.

3.3 Sparse Coding and ICA

The choice of a coding strategy depends critically on the distribution of natural scenes in state space. If the distribution is a Gaussian ellipsoid, PCA is the appropriate strategy for coding and a small subset of vectors (a compact code) can represent the data with a minimal amount of error.[7] PCA will always find the vectors that are aligned with the directions of highest variance. Moreover, anytime we are looking for a solution that removes dimensions along which the data have zero variance, compact codes are the appropriate representation. Indeed, dimensionality is reduced a hundredfold between photoreceptors and ganglion cells, meaning that compact coding is being performed.

But if the distribution of natural scenes in state space does not form a multidimensional Gaussian, PCA would not be *sufficient* as a coding strategy because it may not capture the full redundancy of the input. For example, suppose we have a state space distribution with a shape that is more like the X-shape in Fig. 3.2. This two-dimensional distribution could represent a set of

[7] In the case of the coding of color (wavelength) information, trichromatic cone responses do appear to be a form of compact coding. For natural spectral reflectances (i.e., the wavelength-dependent reflectance distributions found in the world), the first three principal components appear to capture most of the variance within the visible wavelengths of light (Maloney, 1986). It should be noted that it is not clear whether the three cone types correspond to these first three principal components.

data that has either strong positive correlations (i.e., if one detector has a positive response, the other has a high likelihood of also having a positive response) or strong negative correlations (if one detector has a positive response, the other has a high likelihood of having a negative response). The distribution has principal components that are degenerate in this plane but because this PCA code has already found a representation wherein the data are decorrelated (which is the goal of PCA) this strategy does not capture the full extent of the distribution's redundancy (Field, 1994).

Given the statistics of the input, a more efficient code—one that captures the redundancy of this distribution—would align the basis vectors with the data (Fig. 3.2). A solution of this kind is called a sparse code and it has interesting properties. A sparse code does not necessarily achieve independence of the components, although they can be independent. One class of implementations of sparse coding is called independent components analysis (ICA), which is a misnomer because the components are not necessarily independent (Bell and Sejnowski, 1996).

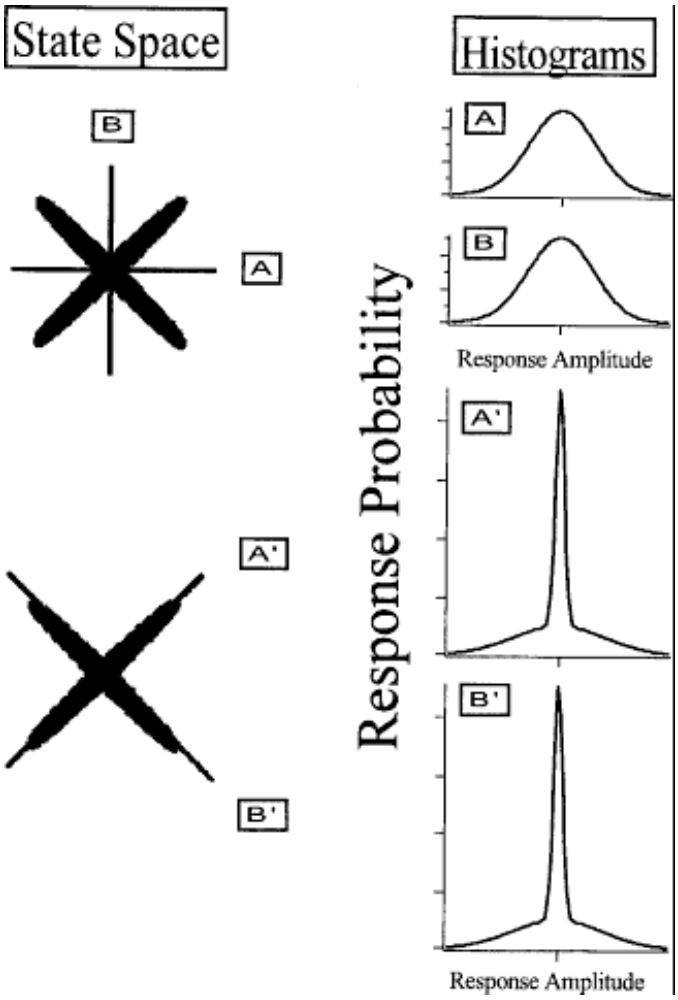


Figure 3.2. In the top left panel shows a distribution that has the *A*-*B* basis as its principal components, but this basis clearly does not capture all the redundancy in the distribution. The *A*-*B* basis in this panel has a Gaussian response profile, as shown in the top right panel. If we choose a sparse code, where our axes are aligned with the data (lower left panel), we see a highly peaked distribution along the *A'*-*B'* basis. The sparse bases have a high probability of having a response of zero and a low probability of values in the tails of the distribution. Note that no two-dimensional basis could possibly capture the variability of images; and in high-dimensional sparse codes, the basis vectors are not necessarily orthogonal. (Field, 1994)

As noted above, the *A* and *B* axes do not capture all of the redundancy. Axes *A'* and *B'* are aligned with the data in a sparse representation and are themselves independent (orthogonal). In contrast, the two plots in Fig. 3.3a

show data (again a mixture from two Gaussian sources) that have the same principal components as the superimposed Gaussian ellipse and in Fig. 3.3b frame, the data have been aligned with the principal components. But the "independent components" (the directions aligned with the data) would not be orthogonal in this case.

a.) b.)

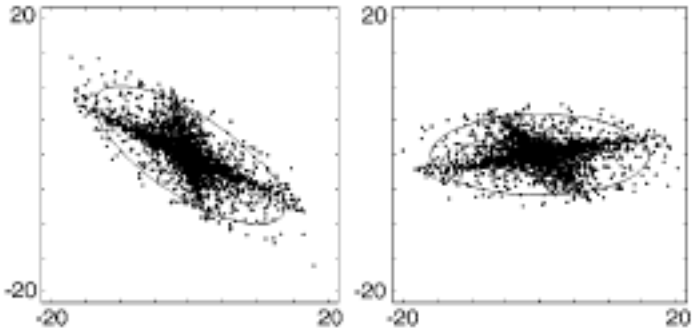


Figure 3.3. Here we show data that have the same principal components as the superimposed Gaussian ellipse (a) and in (b), the data have been aligned with the principal components. But the "independent components" (the directions aligned with the data) would not be orthogonal in this case. (Simoncelli and Olshausen, 2001)

In a sparse code, each detector has a low likelihood of firing for a given stimulus. This has implications in terms of neuronal firing. If visual coding involved reducing the dimensionality of the representation, using a code based on the principal components, all of our basis functions would be in use (i.e., firing) to code most stimuli. This notion goes against physiological evidence that ganglion cell firing generally occurs in bursts of high activity surrounded by periods of relative quiescence (Reinagel, 2001). We note here an important distinction between two forms of sparseness, proposed by Willmore and Tolhurst (2001): *population sparseness* refers to the sparseness of

firing across a population of neurons for a stimulus. Population sparseness describes the typical sparseness of a response to images encountered in the world. *Lifetime sparseness* measures how sparse an individual neuron's response is over its lifetime; Reinagel's study is describing lifetime sparseness. Population sparseness is of prime concern in the studies presented below since we are dealing with static images.

We also note an important distinction between sparse coding and what is called sparse-distributed coding (Field, 1994). A sparse coding strategy says nothing about normalization but a sparse-distributed code, however, demands that the probability that an individual detector will fire is the same for all detectors.

3.4 Measuring Sparseness

In our example of two-pixel correlations, we can examine a histogram of the responses along each basis (Fig. 3.2). For the principal components (axes *A* and *B*), the response histogram is Gaussian (since the ellipsoid was Gaussian by definition). For the "sparse" representation, whose axes (*A'* and *B'*) are aligned with the data, the response profile is more peaked and has rather heavy tails. One thing that distinguishes this peaked distribution from a Gaussian is that it has a high kurtosis, which is the fourth moment of the distribution:

(3.4.1) $$Kurtosis = \left[\frac{1}{M} \sum_{i=1}^{M} \frac{(x_i - \bar{x})^4}{(\sigma_x)^2} \right] - 3$$

A normal distribution has a kurtosis of 0.[8] Kurtosis measures the peakedness of a distribution, although there is disagreement about what specific property of a distribution is being measured (Rinaman, 1993). Kurtosis is also a measure that is sensitive to outlying values in a distribution (Huber, 1985).

One practical difficulty in using kurtosis arises for distributions without a mean of zero. The assumption in a sparse code is that a typical neuron has a low likelihood of firing. That is, the highest probability of an event is to not spike and there is a low probability of firing at any other frequency. A high kurtosis is a good measure of this property of a distribution when the peak is at zero. Real neurons cannot fire with negative spike frequencies, nor do they have mean frequencies of zero. Indeed, most neurons are active even when no stimuli are present: In photoreceptors, there is a continual exchange of ions across the cell membrane in darkness which maintains the cell's potential at about -70 mV. In ganglion cells there is spontaneous firing when there are no visible photons entering the eye (Kandel et al., 2000). In terms of response, however, ganglion cells can be thought of as responding negatively. Inhibition can reduce firing below the rate of spontaneous firing. But there is reason to believe that kurtosis is a useful measure of sparseness. Another sparseness metric (Treves and Rolls, 1991) is of the form

(3.4.2)
$$a = \frac{\left[\sum_{i=1}^{n} \frac{x_i}{n}\right]^2}{\sum_{i=1}^{n} \frac{x_i^2}{n}}$$

[8] Some definitions of kurtosis do not subtract 3 from the sum; in this case, a normal distribution has a kurtosis of 3.

where the sparseness a for a set of n stimuli is related to the firing rates x_i. This metric has been applied to positive distributions. But Willmore and Tolhurst (2001) find that when the Treves-Rolls metric is rescaled, it shows high correlations with kurtosis in the measurement of response sparseness. Both kurtosis and the modified Treves-Rolls metric also agree with what is called an activity sparseness metric, which measures the fraction of cells that are "on," as determined by a threshold.[9] That these three sparseness metrics show strong covariance indicates that they are reliably measuring the same quantity, namely sparseness.

3.5 Evidence for Sparse Codes

Evidence that the visual cortex uses sparse coding comes from neural network[10] simulations (Field, 1987; Olshausen and Field, 1996; Bell and Sejnowski, 1997; van Hateren and van der Schaaf, 1998; Lewicki and Olshausen, 1999; Hyvarinen and Hoyer, 2000) and from physiological recordings (Vinje and Gallant, 2000). Using the method that is today often called independent components analysis (ICA),[11] Olshausen and Field (1996) trained an unsupervised network to develop a basis of receptive fields that optimized sparseness (using kurtosis as a metric) and losslessness when given natural scenes. The resulting non-orthogonal basis bears a striking resemblance to cortical simple cell receptive field mappings produced with spots and gratings (Fig. 3.4). This result shows that a system whose only goals

[9] This latter metric is used in a number of physiological studies such as Berry et al. (1997).
[10] A neural network is a massively parallel system of nodes that have weighted connections. Their mechanism is to adjust the weights of the connections over time in order to maximize a given, small set of parameters. The more input nodes there are, the more one can capture nonlinearities. See Fig. 3.5.
[11] Solutions of this kind are related to the problem of blind source separation (Sejnowski, 2000)

are to develop a sparse code that represents the data with minimal error can produce a set of receptive fields that strongly resemble those found in the visual system.

Vinje and Gallant (2000) showed that physiological recordings from macaque cortex were sparse in response to natural scene movies. They also showed that different sparseness metrics (including a modified Treves-Rolls metric and a kurtosis metric) were in agreement about the sparseness of their data. In order to compensate for the lack of negative response activities in real neurons, Vinje and Gallant reflected their response profiles about zero and then calculated sparseness.

3.6 Metabolic Efficiency

A visual code that is exclusively a compact code is metabolically costly because it requires that the few components that account for the directions with the highest variance be active all the time. In a sparse code, we need not use all our basis functions all of the time, which saves on energy consumption: the vast majority of our detectors are not in use for a given stimulus. This is an important consideration because the body has a limited amount of available energy and using a great deal of it to interpret visual information could hamper other necessary functions. But arguments based strictly on metabolic efficiency are difficult to support, not least because energy consumption is hard to quantify in the brain. Vision is thought to account for 20% of the brain's energy consumption (Kety, 1957; Ames, 1997). At some level, there is a metabolic cost of coding a particular amount of information with a certain amount of uncertainty. But the actual energy cost is difficult to measure. A study of the cost of neuronal computations implies that an efficient system

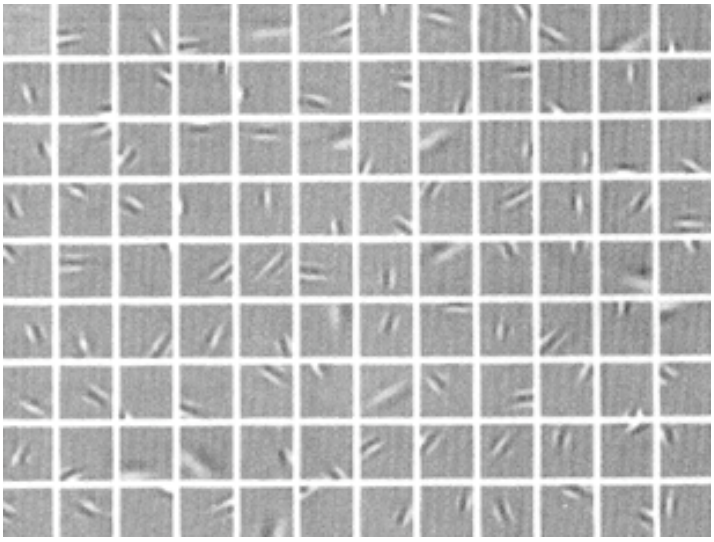


Figure 3.4. Olshausen and Field (1996) trained an ICA neural network to represent a set of natural scenes using an overcomplete, nonorthogonal basis that was optimized for losslessness and sparseness (kurtosis). The resulting basis is remarkably similar to the receptive fields found in cortical simple cells. (Olshausen and Field, 1996)

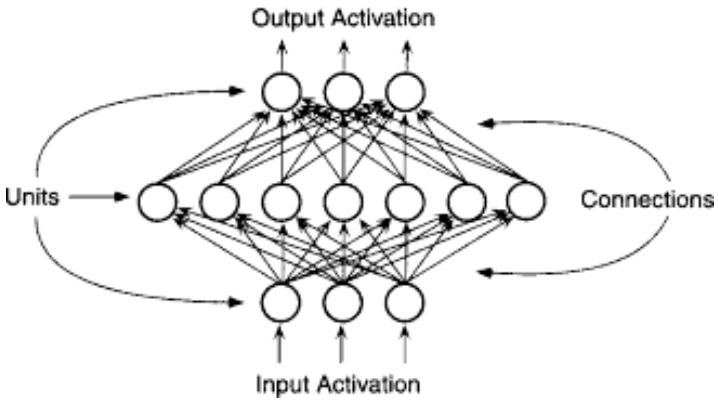


Figure 3.5. Schematic diagram of a neural network. Input nodes link via weighted connections to so-called hidden units. Hidden units often perform a nonlinear thresholding routine (usually a sigmoidal function). Hidden units connect to output nodes. For this type of network, where we are trying to put approximate bounds on nonlinear data with a collection of lines, the number input nodes determines the number of lines, and the number of hidden nodes determines the dimensionality of the representation. Neural networks should be thought of as existence proofs that a system of nodes and activations can accomplish a given task. While they have features that resemble many aspects of neural processing, they are not accurate models of the brain. (Palmer, 1999)

would have each neuron fire at 1 Hz, which indicates that there is significant quiescence (Atwell and Laughlin, 2001, Lennie, 2003).

Levy and Baxter (1996) developed a model of the cost of neural firing and maximized the ratio of the amount of information transmitted C to the cost of a spike E using the following equation:

$$\frac{C}{E} \propto \frac{H(p)}{1 + p(r - 1)} \tag{3.6.1}$$

where $H(p)$ is the Shannon entropy for a binary neuron with a probability p of firing, and r is the cost of firing in a time interval t (note that this model is time-independent). Given estimates of the parameter r, they found that this function is maximized when 2-16% of neurons are firing. Similar results were obtained for their model of an analog neuron suggesting that codes that result in low firing probabilities (like a sparse code) could be advantageous.

CHAPTER 4

CODING IN RETINAL GANGLION CELLS

4.1 Ganglion Cell Physiology

What do retinal ganglion cells accomplish in terms of processing? All the visual information gathered by the eye that reaches the brain passes through retinal ganglion cells. Most ganglion cells have receptive fields comprised of either an inhibitory annulus surrounding an excitatory center (on-center) or an excitatory annulus surrounding an inhibitory center (off-center). This so-called center-surround antagonism is implicated in the processing of other modalities including the auditory and somatosensory streams. It is found also in the early visual processing of other species, including those with compound eyes (Land, 1985). Moreover, retinal processing has been remarkably conserved across mammalian species: all mammals are found to have similar processing units, layout, development[12] and structure, suggesting that the form of processing achieved by the retina is an exceedingly good one.

If center-surround organization is found to be common and conserved across species and in important units of multiple modalities, why is this pattern so common? Does the precise nature of its shape bear on its processing? That is, does the center-surround organization maximize information transmission in some way given the constraints of neural architecture? Or is the shape of the receptive field just an artifact of the common evolution of all neurons? We find that center-surround filters achieve a degree of sparseness in their responses and they may attempt to equalize

[12] Across mammalian species, ganglion cells are start and finish neurogenesis before all other retinal neurons (La Vail et al., 1991; Rapaport, 1992, 1996).

their responses across spatial frequency.

First we must define what we mean by the term efficient. There are on the order of 10^8 photoreceptors and only 10^6 ganglion cells, which suggests that we must account for this decrease in bandwidth with some sort of data compression. Studies using flashing spots or sine-wave gratings as stimuli (Kuffler, 1953, Rodieck and Stone, 1965, Enroth-Cugell and Robson, 1966, Croner and Kaplan, 1995) and more recently using spike-triggered averages for white noise (e.g., Brown and Masland, 2001) have mapped out the size and sensitivity of the receptive fields in various vertebrate species. Although there are at least a dozen varieties of ganglion cells (Wassle and Boycott, 1991) most can be classified into two broad classes based on their responses and projections:

- **Magnocellular ganglion cells** (M-cells) are generally large cells that respond to contrast differences but not to color differences. They have large receptive fields, they respond transiently to changes in stimulation, and they project to magnocellular layers in LGN.
- **Parvocelluar ganglion cells** (P-cells) are smaller and they are sensitive to color differences. These cells have a sustained response to stimulation and they project to parvoccllular layers in LGN (Palmer, 1999).

The M- and P-types correspond to the anatomical classes called parasol and midget cells, respectively. Approximately 70% of the monkey's 1.5 x 10^6 ganglion cells are midget cells (Masland, 2001). These cells are the dominant form found in the primate fovea: they have high spatial acuity, and receive input (via a pathway that includes associated midget bipolar cells) often from

single photoreceptors. Both the M- and P-cell types, which together comprise the majority of ganglion cells, show center-surround antagonism. We model ganglion cell receptive fields as spatial filters whose response to a visual stimulus is the convolution of the image and the filter. In particular, the receptive fields are modeled with radially symmetric difference-of-Gaussian (DoG) functions following Rodieck (1965).

4.2 Nonlinearities in the Retina

Ganglion cell responses are clearly dependent on a number of factors besides spatial contrast. Rodieck (1965) made an important first step in defining the temporal response of ganglion cells as an impulse followed by an "undershoot" that decays exponentially in time:

(4.2.1) $$A(t) = \delta(t) + he^{-t/\tau}$$

where $\delta(t)$ is a delta function impulse, h is the size of the undershoot and τ is the decay constant. In Rodieck's model, then, the response to any stimulus would be

(4.2.2) $$R(t) = R_0 + \iiint I(x,t')B(x)A(t - t')\,dx\,dt'$$

where R_0 is the basal firing, $I(x, t')$ is the light intensity pattern, and $B(x)$ is the DoG spatial sensitivity function. This model is appealing and it has influenced a great deal of subsequent research. But it assumes both space-time separability as well as response linearity. In regard to space-time separability,

a number of studies have shown that the cell center responds faster than does the surround (Enroth-Cugell and Freeman, 1987; Sakai and Naka, 1995; Benardete and Kaplan, 1997a). Responses are linear only in limited conditions (Enroth-Cugell and Robson, 1966; Victor, 1987; Benardete and Kaplan, 1997b). A number of schemes have found success in predicting these nonlinear responses. One popular model (Victor, 1987) uses full-wave rectification and low-pass filtering in the temporal domain. This technique essentially models the fact that ganglion cells are faster but less sensitive to large changes in light, in agreement with physiological studies (Shapley and Victor, 1979, 1981).

There are a number of other nonlinearities in retinal processing that affect ganglion cell responses. For example, 52% of retinal ganglion cells are thought to be sensitive to spectral differences of light (Zrenner, 1983). P-cells in particular have color opponancy in their receptive fields, wherein cells are have an excitatory response to red or blue wavelengths in the center and an inhibitory response to green or yellow wavelengths in the surround (or vice versa). This is an important feature of visual processing that a full model of retinal processing would address. The present work deals only with spatial, monochromatic processing.

P-cells and M-cells have different sized receptive fields, different temporal responses and different spectral responses. In terms of their contrast response, many studies and reviews (e.g. Kaplan, 1991; Meister, 1999) treat both cells as linear up to saturation. P-cells appear to be linear up to 80% contrast, while M-cells are linear only to 10% contrast (Kaplan, 1991). The saturation characteristic of real neurons is highlighted by Borghuis et al. (2002), who showed that a linear model of P-cell response strongly overestimates the response at high firing rates, where saturation is thought to occur.

4.3 Noise Reduction

Srinivasan et al. (1982) stress the noise-reduction requirements of the retina as the guiding principle in receptive field design. Although noise is present at every level of visual processing, there is physiological evidence that photon absorption noise is the dominant form of noise in the visual system (Pelli, 1981, 1991; Banks et al., 1987; Cohn, 1976; Krauskopf and Reeves, 1980). In their predictive coding hypothesis, Srinivasan et al. propose that lateral inhibition in general and center-surround filter design in particular maximizes the use of the dynamic range of detectors in order to transmit information at the highest possible resolution for a given level of noise. The dynamic range of a neuron is necessarily limited by physiological factors and there is a limit to the number of discernable levels at which a neuron can fire (Barlow and Levick, 1976). Even without input noise, visual interneurons have a limited number of responses because of discretization error. For example, if the system can code information with spike frequencies between 0 and 200 Hz and there is an uncertainty of 10 Hz at every frequency, the system can only code with 20 distinguishable levels (Srinivasan et al., 1982).

Srinivasan et al. propose that the fact that natural scenes are spatially correlated indicates that the visual system could use lateral inhibition to code spatial information with a minimal dynamic range. They show that a group of detectors sampling over space will make the best possible prediction (in terms of statistics) by taking a weighted linear sum over the group of detectors. Specifically, for a 1-dimensional set of m detectors that gives samples of x_i, the best prediction at location 0 is given by:

(4.3.1) $$\hat{x}_0 = \sum_{i=1}^{m} h_i x_i$$

where h_i are the solutions to the set of simultaneous linear equations that describe the spatial autocorrelations R_{ij}:

(4.3.2) $$\begin{aligned} R_{11}h_1 + R_{12}h_2 + \ldots + R_{1m}h_m &= R_{01} \\ R_{21}h_1 + R_{22}h_2 + \ldots + R_{2m}h_m &= R_{02} \\ &\vdots \\ R_{m1}h_1 + R_{m2}h_2 + \ldots + R_{mm}h_m &= R_{0m} \end{aligned}$$

The coefficients R_{ij} express the average value of the product of samples at locations i and j.

(4.3.3) $$R_{ij} = \langle x_i x_j \rangle .$$

This so-called *predictive coding* minimizes the error between the predicted value and the actual value at a given location. The standard deviation of the mean square error for this example would be:

(4.3.4) $$E = \sqrt{R_{00} - \sum_{i=1}^{m} h_i R_{0i}} .$$

Redundancy in the signal is reduced as a byproduct of this scheme. Retinal processing in their view is "concerned primarily with coding the visual image

to protect it from subsequent intrinsic noise rather than with reconstructing the scene or extracting specific features from it." Srinivasan et al. argue that ganglion cell receptive field organization is mainly guided by the signal-to-noise ratio of the signal: For low-intensity signals (for example in scotopic conditions, where signal and noise intensity are similar) receptive fields should sample over a larger area in order to have a higher signal-to-noise ratio; High intensity signals require less sampling and a smaller dynamic range than would be necessary without predictive coding. They show that these optimal receptive field designs are matched by large monopolar cells (LMC) in the visual system of the fly. LMCs are units that sum signals from photoreceptors in the fly's compound eye and they show the same center-surround receptive fields as ganglion cells in vertebrates.

4.4 Redundancy Reduction

The prevailing view of retinal coding was shaped by the argument of Srinivasan et al. and by Atick and Redlich (1992), who focus on spatial pairwise redundancy reduction as the goal of retinal processing. It should be noted that both of these theories start with the same assumption about the statistics of natural scenes, namely their pairwise correlations. Atick and Redlich start with the finding by Burton and Moorhead (1987) and Field (1987) that the power spectra of natural scene images goes as $1/f^2$, which is the same statistical regularity as that expressed by the autocorrelation function. The output of the *j*th ganglion cell is obtained from linear filtering:

(4.4.1) $$O(\mathbf{x}_j) = \int K(\mathbf{x}_j - \mathbf{x})L(\mathbf{x})\mathrm{d}\mathbf{x} \equiv K \cdot L$$

where $L(x)$ is the light intensity, $K(x_j - x)$ is the receptive field filter kernel and x_j is the center of the receptive field. Since we know the spectrum of the input in general, we need only multiply the power spectra together. Atick and Redlich did just this and multiplied a $1/f^2$ power spectrum by a plot of the output intensity as a function of stimulus frequency measured in the retina (recorded from single cells). This multiplication is allowed because of the convolution theorem

$$T\big(f(x) \quad g(x)\big) = T\big(f(x)\big) \cdot T\big(g(x)\big), \tag{4.4.2}$$

which states that the Fourier transform of the convolution of two functions f and g is equal to the product of the transforms of each function. Atick and Redlich found that the average output spectrum

$$\left\langle O(\mathbf{f}) O^*(\mathbf{f}) \right\rangle = \left\langle (K(\mathbf{f}) L(\mathbf{f}))(K(\mathbf{f}) L(\mathbf{f}))^* \right\rangle \tag{4.4.3}$$

is equal to a constant up to the peak of activity:

$$\left\langle O(\mathbf{f}) O^*(\mathbf{f}) \right\rangle = const. \tag{4.4.4}$$

The same qualitative result was obtained when psychophysical contrast sensitivity data were used instead of neural responses. Atick and Redlich argue that the flattening or whitening of the output spectrum would lead to statistical independence of ganglion cell responses in space. If a flat power

spectrum is Fourier transformed, we get an autocorrelation function that is a delta function. That is, a distribution with the same power at each Fourier frequency will have no pairwise correlations. In the case of the retina, whitening only proceeds to the peak of intensity (or sensitivity for the psychophysical data) but in general the average value of the output in space would be a delta function:

(4.4.5) $$\left\langle O(\mathbf{x}_i)O(\mathbf{x}_j)\right\rangle \propto \delta_{ij}$$

Atick and Redlich make an important point about noise. As we mentioned, whitening of the spectrum in these experiments proceeds only through the peak response (the peak is around 4 cycles/degree), not into the high frequencies. Because the noise spectrum is usually flat and not declining in the same way as the signal ($1/f^2$), whitening in the high-frequency regime would serve to amplify unwanted noise. Also, if the input signal is decorrelated in the output, there is no way to distinguish signal from noise. This argument is consistent with findings that ganglion cells lose their inhibitory surrounds at low luminance and become lowpass filters. As Atick and Redlich point out, lowpass filtering increases the signal to noise ratio because signal power becomes small at high frequency whereas noise power is constant across frequencies.

There have been recent studies investigating ganglion cell temporal coding that use movie stimuli as well. In a paper that complimented their other work with natural movies, Dong and Atick (1995b) found that a system that is optimized for removing temporal correlations (whitening) showed a similar

power spectrum (at low frequencies) to that obtained from physiological recordings from the cat LGN. As we mentioned, receptive fields in the LGN are similar in their spatial properties to those of ganglion cells; LGN cells have different temporal filtering properties, though, which Dong and Atick say could be related to temporal whitening.

Dan et al. (1996) also found that the temporal frequency response of LGN cells was relatively flat over a range of temporal frequencies for natural movie stimuli. Moreover, they found that movies of Gaussian noise resulted in an output power spectrum from the LGN cells that increased linearly with frequency, showing that only images that display the second-order correlations typical of natural images will result in a flat (whitened) output. While time-varying statistics are certainly an important area of study, this thesis deals only with static images.

Meister (1999) has cited unpublished data that suggest that retinal ganglion cell firing is "sparse and precise" when natural movies are shown to salamander and rabbit retinae. Reinagel and Reid (1998) found the same precise firing in cat LGN. These studies are particularly important because they used natural stimuli.

4.5 Measuring Correlations

As Atick and Redlich point out, Barlow (1989) argued that having statistically independent outputs means that the joint probability for any combination of outputs O_i can be calculated directly from individual probabilities $P_{ijk\ldots}$. Using Shannon's theory, the output strengths can be taken as being proportional to their "improbabilities," given by $-\log(P_i)$, which also represents the information content of the output. In this statistically

independent basis, each output represents a "feature," which would correspond to a "word" in language. The discovery of such independent features reduces the *predictable* redundancy of the representation. This could well be one of the goals of retinal processing.

An important question to ask at this point is How correlated are real ganglion cells in their firing? Physiological studies agree that ganglion cell responses are highly correlated but there is debate about how much the correlated activity contributes to coding. Multineuron recording studies testing the independence of ganglion cell responses find that nearby ganglion cells of similar functional classes have highly correlated firing patterns across species (Arnett, 1978, Arnett and Sparker, 1981, Johnson and Levine, 1983, Mastronarde, 1989, Meister, 1996, DeVries, 1999). There is also evidence that in development, correlated firing is necessary for retinal neurons to innervate correctly. Synchronized firing has been proposed as a mechanism that helps coordinate the proper development of neural wiring (see Wong, 2000).

But using a controversial information theoretic measure, Nirenberg et al. (2001) argue that information transmission is not strongly *improved* by correlated firing. In particular, they show that when retinal ganglion cells are shown natural scene stimuli, 90% of the information about stimuli can be extracted without examining correlations among responses, suggesting that ganglion cells act as "independent encoders." There is a great deal of debate about this study but the fact that ganglion cell responses are correlated is not disputed.

Furthermore, decorrelation arguments are insufficient to describe the center-surround organization of retinal ganglion cell receptive fields. The first experiment described below uses phase-randomized receptive fields to

achieve the same degree of whitening as center-surround filters. We show that phase-randomized center-surround filters—which are utterly incompatible with physiology—achieve the same level of whitening and high-frequency attenuation as phase-aligned center-surround filters. We propose that what distinguishes receptive fields with center-surround organization from phase-randomized filters is the *sparseness* of their responses to natural scenes.

While we do not dispute that whitening of the power spectrum is achieved by retinal processing, our explanation for how it comes about does not define whitening in terms of decorrelation. If the goal of retinal processing were to equalize the response of units tuned to different frequencies, we would also observe a flat power spectrum in the output. That is, if we remain agnostic as to whether decorrelation is occurring, we can achieve the same whitened spectrum by demanding that the retina seek a uniform response across frequency. This notion is called response equalization (Field, 1987; Field and Brady, 1997; Brady and Field, 2000) and it is appealing because cells that respond to different frequencies will have similar dynamic ranges.

The goal of a sparse-coding network is to translate the redundancy in the stimulus into a redundancy in the response magnitude of the detectors, and any stimulus can be coded by a different, small set of units. For a sparse-distributed code, all detectors are designed to have the same low probability of firing. Such representations retain high dimensionality like the input, but the individual detectors have a low likelihood of firing.

CHAPTER 5

EXPERIMENTS

5.1 Study 1: Phase Alignment and Sparseness

For the first study, images from van Hateren's database (van Hateren and van der Schaaf, 1998) were randomly selected. Images were discarded if they did not conform to two criteria: They were required to be devoid of human-created forms and of significant blur. The restriction on blur is the more crucial one: these images were captured with a head-mounted camera that was timed to open the shutter at various intervals. If the photographer's head moved when the shutter was open, the images were blurry, which introduces uncertainties into the data. After our selection process, we arrived at a set of 137 stimuli that shows a range of scenes at different scales. The stimuli do not show a wide variety of environments: they were all obtained in Holland, a country not known for its wealth of diverse environments. Most of van Hateren's images show grass or forest scenes, some have bodies of water and none has any large vistas. These are significant drawbacks to using van Hateren's database. But this database has basically the only large set (~4000) of calibrated images and it has been cited in 96 subsequent studies. The difficulties in creating such a database are many: The goal is to make photometric maps of scenes wherein pixel values correspond linearly to luminance. Films and CCD chips have nonlinear transfer functions that must be corrected by sampling a scene with a photometer and a set of standard reflectors. The idea is to sample a card with different reflectances on it in the same lighting as the scene and map pixel values to these photometric readings. But changing lighting conditions, clouds and wind all make this

very difficult to achieve in practice. While far from perfect, van Hateren's database is reasonably representative of the patterns of luminance found in the world.

We used a difference of Gaussians (DoG) model of retinal ganglion cell receptive fields as the basis of our filter kernels (Fig. 5.1). The radially symmetric DoG function *R* is described by

$$R(x,y) = C_1 \frac{1}{\pi\sigma_1^2} e^{-(x^2+y^2)/\sigma_1^2} - C_2 \frac{1}{\pi\sigma_2^2} e^{-(x^2+y^2)/\sigma_2^2} \tag{5.1.1}$$

where C_1 and C_2 are constants that determine the height of the center and surround Gaussians, respectively, and σ_1 and σ_2 are the variances of the center and surround respectively (Rodieck, 1965). In our study, $\sigma_1 / \sigma_2 = 6.0$ and $C = 20$ and filter kernels were created in a frame of 64×64 pixels then centered and zero-padded to make them 1024×1024 pixels (the size of the stimuli). Convolutions were performed with phase-aligned DoG filters and with DoG filters whose Fourier frequency components had been phase-randomized (that is, they were given a norm-preserving, random rotation in phase space). As allowed by the convolution theorem, we calculated the convolutions by multiplying the power spectra of the images and filters. Clearly, the power spectra of the phase-aligned and the phase-randomized filters are identical.[13] Six phase-randomized filters were each convolved with the image set and we took the mean of these trials as our phase-randomized power spectrum. The

[13] The convolved power spectra are identical when the image and filter are the same size, as we found in a separate trial. But in our experiment, because the images are larger than the filters, zero padding is necessary, which leads to differences in the convolved power spectra at low frequencies. However, the mean convolved power spectrum for the randomized filters falls within the error bars of the mean DoG filter power spectrum.

64 pixels at the edges of the image were cropped in order to remove edge effects (this was necessary because we did not use periodic boundary

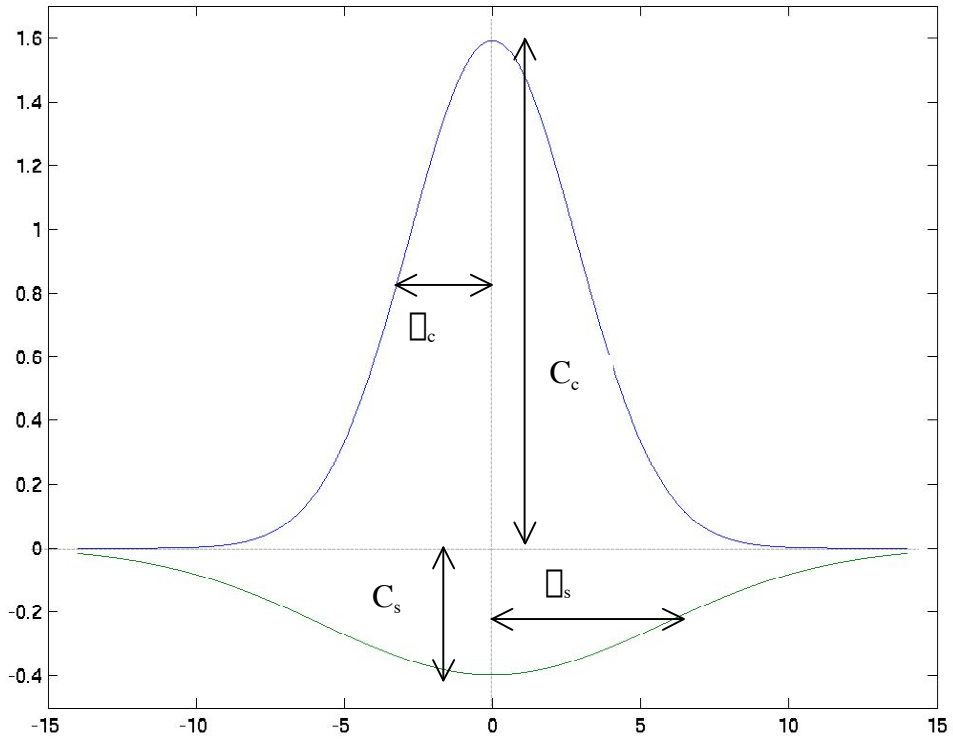
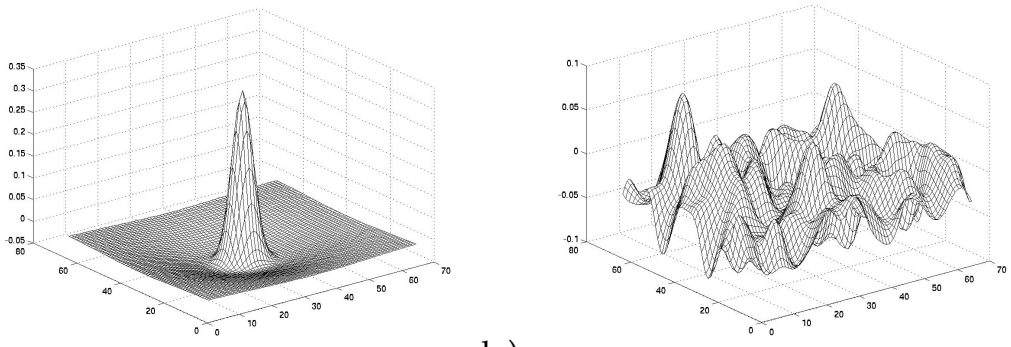


Figure 5.1. Difference of Gaussians function in one dimension showing the two Gaussian functions before they have been added together and the parameters that can be varied in the function (equation 3): C_c is the center sensitivity, σ_c is the center width, C_s is the surround sensitivity and σ_s is the surround width.

conditions). Phase-randomized filters achieve the same whitening and high-frequency noise attenuation as center-surround filters in the 0.3 to 3 cycles/degree range described by Atick and Redlich. But the phase-randomized filters do not resemble ganglion cell receptive fields (see Fig. 5.2).



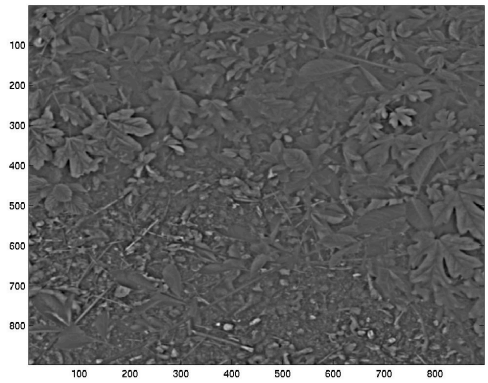
a.) b.)

Figure 5.2. Difference of Gaussian receptive fields based on equations and parameters from Rodieck (1965) are shown phase-aligned (a) and phase-randomized (b). The z-axis represents sensitivity, the x- and y-axes represent space. These filters were convolved with images from the database of van Hateren and van der Schaaf (1998).

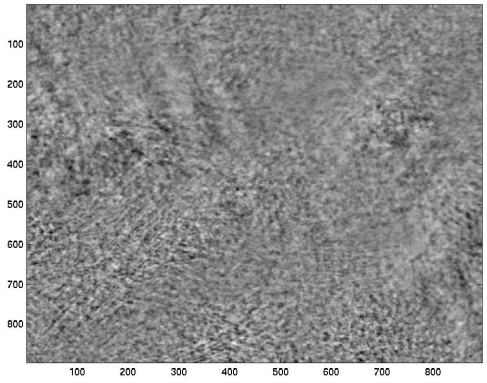
In the space of all possible 64 × 64 pixel filters with this power spectrum, the phase-aligned center-surround pattern is unlikely to be chosen over one of the phase-randomized patterns. Moreover, in the unlikely event of the visual environment undergoing the same phase randomization as the filter, such a filter would be quite appropriate for coding scenes in that environment. For filters with a given power spectrum, each alignment of phases preserves the same amount of information in any convolution with an image—though, to our eyes, the images that have been convolved with phase-randomized filters do not look like a useful representation (Fig. 5.3). Our task, then, is to explain why center-surround organization in space is advantageous.

Ruderman (1994) suggests that phase alignment is desirable in order to preserve rotational symmetry. Phase alignment has often been simply assumed, as with physiological studies that test sensitivity using sine-wave

a.)



b.)



c.)

Figure 5.3. Original linear image (a), DoG filter convolution (b), and randomized filter convolution (c) for image 8.

gratings (e.g., Croner and Kaplan, 1995) although more recent studies measure phase alignment. In such studies, receptive fields are reconstructed from response power spectra, resulting in center-surround patterns, though any arrangement of phases gives the same power spectrum. We are interested in a more practical reason for phase alignment, one that should relate to the information transmission properties of the cells. We suggest that center-surround filters increase the sparseness of the output and that this is one of the main goals of retinal processing.

We measured the sparseness of our convolved images (Fig. 5.3) using kurtosis as our metric and found that the sparseness of the center-surround filtered images is on average 3.5 ± 2.4 (mean ± SD) times greater than that of images convolved with phase-randomized filters (Figs. 5.2a and 5.2b). That is, the mean ratio of the sparseness of a center-surround filtered image and a phase-randomized-filtered image was the number reported (Fig 5.5). This value refers to the difference in population sparseness (see section 3.2).

To gauge lifetime sparseness (that is, sparseness of a single neuron through its lifetime), we compiled a total histogram for all images after the two types of filtering. In this case, the center-surround filtered images had a sparseness that was 1.4 times greater than that of images convolved with phase-randomized filters. However, this measure is not necessarily an accurate description of a neuron's firing through time since the images are static snapshots of the world, not frames of a movie.

As a control, the same convolutions were performed on Gaussian white noise and on white noise whose power spectrum was given by $1/f^2$, both of which gave a kurtosis of 0 for all cases.

5.2 Compressive Nonlinearities

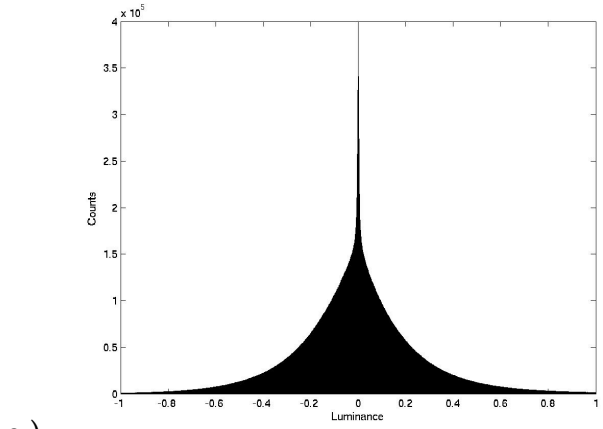
In keeping with the proposal of Srinivasan et al., we performed another experiment in which we convolved the same sets of filters with log-transformed images. The rationale for taking a log of the image is based on physiological studies of frogs such as Norman and Werblin (1974), who showed that photoreceptor sensitivity, when adaptation over time is taken into account, goes as the log of the intensity. This nonlinearity is used in models of image processing (e.g. Field, 1994; Brady and Field, 2000) although it is by no means the only one that has been proposed. Other so-called compressive nonlinearities, such as the so-called Naka-Rushton equation (Naka and Rushton, 1966), which was based on recordings from minnows, and that of Baylor et al. (1987), which was based on recordings from the macaque, have different nonlinear relationships between intensity and cone response. While it is likely that photoreceptors perform some form of nonlinear transform, it is difficult to compare nonlinearities in terms of data compression properties since, by definition, there is no linear relationship between them. Moreover, different nonlinearities may preserve different amounts of information in the representation.

With respect to information theory, as Field (1987) pointed out, a log transform would recast intensity differences as ratios, a property that could be advantageous for the cell:

(5.2.1) $$\log(a/b) = \log(a) - \log(b).$$

Difference of Gaussians filters, which use the cone response as their input, can thus be thought of as Ratio of Gaussian filters. Compressive nonlinearities, as

the name implies, reduce the dynamic range over which the cell must operate.



a.)

b.)

Figure 5.4. Total filtered, log-transformed image histograms (for all 137 images) for DoG filters (a.) and for phase-randomized filters. The sparseness (kurtosis) of (a.) is 4.3, the sparseness of (b.) is 2.9.

We used a log nonlinearity for the images from the previous study then convolved each with the same sets of filters. Sparseness for the DoG filters was 1.9 ± 1.0 times higher than for the phase-randomized filters (Fig. 5.5). This result cannot be compared to the untransformed (linear) result because the input data have undergone a nonlinear transformation. We report the log case

in order to show that sparseness is higher for the DoG filters than for the phase-randomized filters when a model of the cone nonlinearity is included. The other response nonlinearities that have been proposed (Baylor et al., 1987; Naka and Rushton, 1966) achieve a contrast-like response, whereas a linear response is based on the absolute amplitude of the signal. Lifetime sparseness, as defined above, was found to be 1.5 times greater for the center-surround filtered images than for those filtered with phase-randomized filters (Fig. 5.4).

We note that the DoG filter and the phase-randomized filter are indistinguishable solely based on the mean (the first statistical moment) because they were both designed to have a mean of zero. Nor could they be distinguished based on variance (second moment): Because the filters have the same power spectrum, they have the same variance, since the integral of the power spectrum over frequency is the variance. Differences in the skew (third moment) of the filtered images showed no clear pattern.

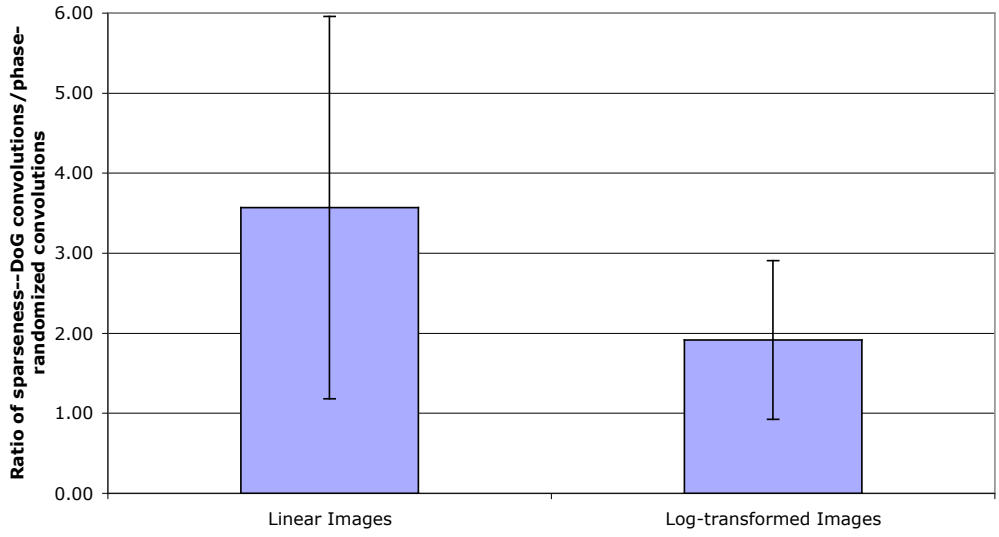


Figure 5.5. Mean ratio (across images) of sparseness for DoG convolutions : phase-randomized filter convolutions for linear images and for log-transformed images.

5.3 Study 2: Contrast Sensitivity and Vector Length

As a group, then, retinal ganglion cells achieve a whitened, sparse output by virtue of their center-surround receptive field organization. The whitening aspect of the output could be defined in terms of two different goals of retinal processing: In the Atick and Redlich sense, whitening is a consequence of decorrelation. But a sparse-distributed coding strategy could also give a flat or "whitened" response across frequency because it seeks to equalize the response magnitudes, which would mean that cells sensitive to different spatial frequencies have the same dynamic range. This latter definition does not say anything about the decorrelation goals of the retina.

What distribution of sensitivity to contrast will produce such a response given natural scenes? That is, how should sensitivity be distributed across frequencies in order to achieve these goals? Contrast sensitivity in humans is thought to be bandpass, with a peak around 4 cycles/degree and then fall with frequency (Fig. 5.6). But perhaps this peak represents the point where the ratio of signal to noise is maximized. We have used a sensitivity measure based on the vector length of the filters that increases monotonically with frequency for some cells, as proposed by Field and Brady (1997) and Brady and Field (2000). Instead of peaking at 4 cycles/degree, sensitivity increases to the high-frequency cut-off. This method of measuring the vector length (L^2-norm) of the model receptive fields presumably takes account of the specific shape and size of the filter better than the standard measure, which is defined as the peak sensitivity at the cell center. When the distribution of vector lengths is multiplied by the power spectrum of an image with power distributed as $1/f^2$, the response will be flat. But we stress that this scheme

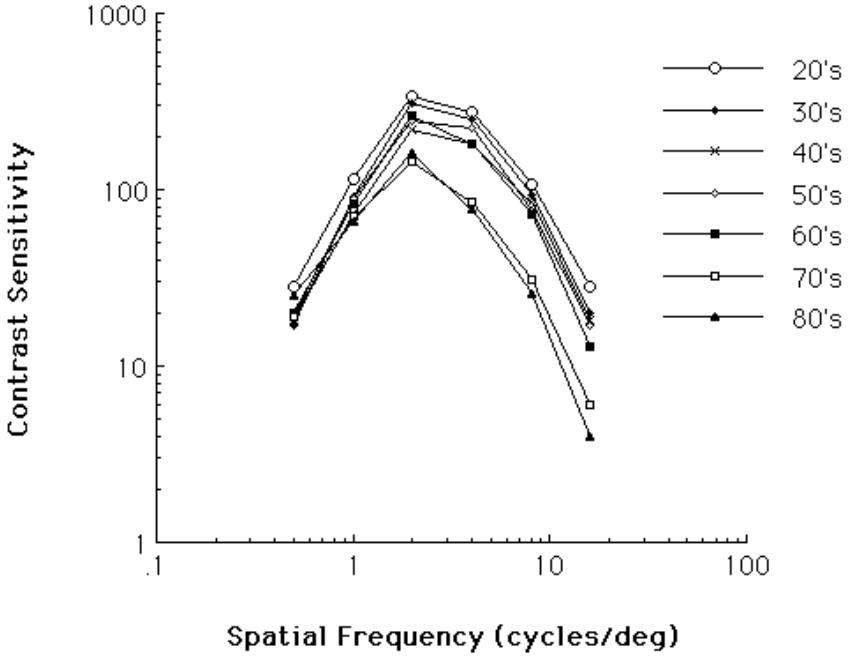


Figure 5.6. Contrast sensitivity as a function of spatial frequency for humans in different age groups (Schieber, 1982).

says nothing about spatial correlation: The flat response in this case is a result of the sensitivity of the cells being distributed proportional to frequency in such a way as to maximize the use of the cell's dynamic range, not as a way to give statistically independent responses. In other words, cell sensitivity is designed to produce the same response magnitude across frequency, which means that cells tuned to different frequencies can have the same dynamic range. If this is indeed the goal of retinal processing, contrast sensitivity in the ganglion cells would appear to be designed to take advantage of the statistics of the environment; namely, its $1/f^2$ structure.

5.4 Vector Length

Our second study involved retinal ganglion cell sensitivity measurements in the macaque by Croner and Kaplan (1995). It is one of the only recent studies of its kind that tested sensitivity across the retina. This study measured synaptic potentials in cells across the retina in awake, behaving macaques when they were shown gratings of different frequencies. Sensitivity data is therefore collected as a function of frequency in the form of a power spectrum. The receptive fields' spatial structure is then reconstructed (using the inverse Fourier Transform) to determine sensitivity. But without phase information the power spectrum is insufficient for establishing a unique solution. Croner and Kaplan assume phase-alignment in the way they reconstruct the spatial receptive fields from responses to grating stimuli. Phase-alignment is generally assumed in receptive field mapping studies that use a basis of orthonormal gratings. In our study of the Croner and Kaplan data, we used the measured values of their experimentally determined parameters for the DoG function describing the cells' receptive fields. We calculated the vector length (the L^2-norm) for the DoG functions, given by

(5.4.1) $$\int |\square(x)|^2 dx,$$

for P-cells in the study.[14] P-cells are the dominant class found in the primate fovea and they have high spatial acuity compared to M-cells. We plotted our calculated sensitivity values as a function of the mean spatial frequency (i.e., the mean frequency value of the spatial frequency power spectrum) of each cell, shown in Fig. 5.7. Parameters for the 84 total P-cells of different sizes

[14] This measure was first used by DiLorenzo (1989) in models of neurons in the taste modality.

represent the median value within bins corresponding roughly to cells of the same temporal-equivalent eccentricity on the retina. There are five such median values for the parameters that describe the receptive field function. The results of our measurement when applied to Croner and Kaplan's data show a function increasing monotonically with mean frequency (Fig. 5.7) in a log-log plot. Using this sensitivity measure, sensitivity increases well past 4 cycles/degree.

What does this vector-length sensitivity curve tell us about ganglion cell responses? When this contrast sensitivity function is multiplied by a $1/f^2$-function representing the power spectrum of a natural image, the response function is flat, indicating that the response at each frequency will be approximately uniform (Fig. 5.8). This property is called *response equalization*

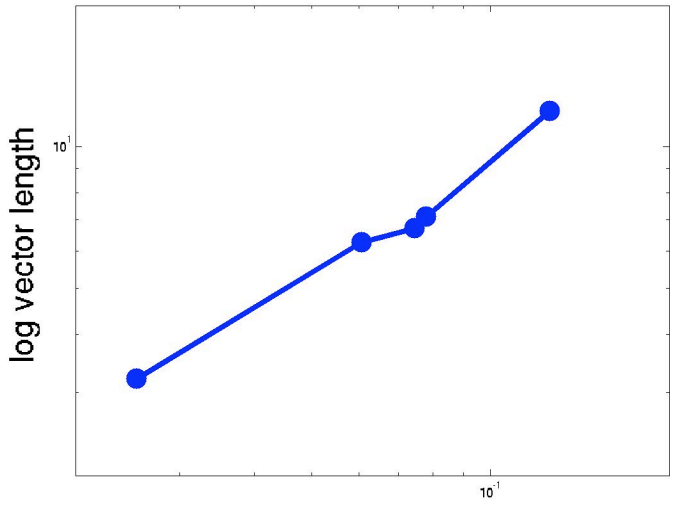


Figure 5.7. Plot of the vector length sensitivity of cells from data by Croner and Kaplan (1995). Vector length sensitivity is a monotonically increasing function, which gives an equalized response (see Fig. 5.8). The *x*-axis represents the log mean spatial frequency of each of the cells. A similar plot resulted when the median frequency value was used as the *x*-coordinate (not shown).

(Field and Brady, 1997). Flattening of the power spectrum in other contexts (namely decorrelation; see section 4.5) is called whitening as we mentioned. But we stress that we have remained agnostic on the question of decorrelation. The idea of using the term response equalization is to shift focus from the whitening aspects of retinal processing to the uniformity of their responses. Uniform responses across frequency mean that the cells are maximizing the use of the range of firing rates over which the cell responds (i.e., the dynamic range) given the regular statistics of the environment.

We note that the data given show no high-frequency cut-off but we assume that such a cut-off exists. High-frequency noise will be attenuated above this cut-off. Without knowledge of how sensitivity was affected by mean luminance in the Croner and Kaplan study, we cannot say any more about the

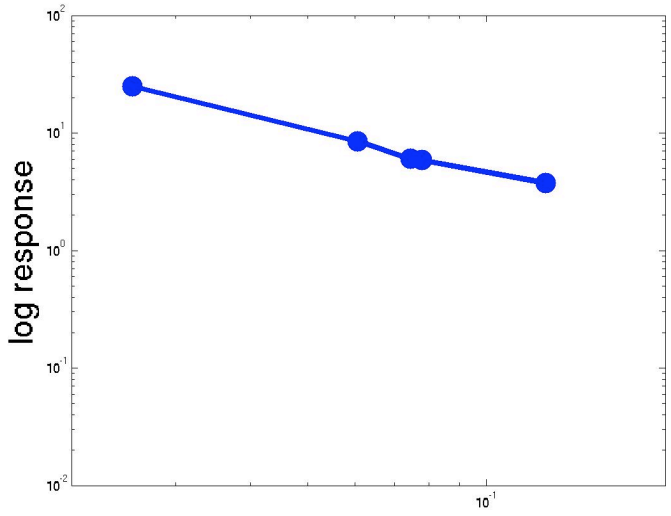


Figure 5.8. Plot of the response magnitude of ganglion cells—found from the product of the measured vector length (L^2-norm; see Fig. 5.7) sensitivity of the cells and a distribution of 1/frequency^2, an "input" that represents a typical natural scene. This plot, a function of the log mean spatial frequency of each cell, shows a generally flat shape and a decreased response at high frequencies.

noise-reduction properties of primate ganglion cells in this study.

How do we explain the observed human contrast sensitivity function, which is found in many studies to peak at 4 cycles/degree? The contrast sensitivity function measures the psychophysical threshold at which humans are able to detect contrast at a given frequency (using gratings) and as such it takes account of the entire visual system. Vector length sensitivity, on the other hand, measures the sensitivity of ganglion cells. We propose that the 4 cycles/degree peak of the psychophysical contrast sensitivity function is the point at which the signal to noise ratio is maximized. The power spectrum of noise in the visual system is flat (see section 4.4) whereas signal power decreases (Pelli, 1981, 1991). Therefore, the signal to noise ratio goes down (i.e., it has a negative slope) at high frequencies. We propose that the positive slope observed at low spatial frequencies corresponds to a regime that is coded by the lowest-frequency channel used by the visual system. That is, large (low-frequency) ganglion cells have a limit to their radial size and stimuli above this size are all coded by this lowest-frequency channel.

5.5 Discussion

One of our assumptions is that the retina has some knowledge of the redundant structure of images in the design of its processing. The prejudice in the past has been to assume that because the retina is the earliest major processing unit, it "is not expected to have knowledge beyond the simplest aspects of natural scenes" (Atick and Redlich, 1992). This has led to an effort to understand retinal processing purely in terms of its effect on the power spectrum. We believe the sparse output of our DoG filters implies that the retina is well-matched to other scene statistics.

We limit our findings to include only those cells with receptive fields that are well-described by a DoG model (center-surround). Some classes of ganglion cells do not have center-surround receptive fields (these represent a minority of the total number of cells). In addition, studies of ganglion cell sensitivity that use spike-triggered averages produced for white noise movies show that there could be additional microstructure in center-surround receptive fields that is not described by the DoG model (Brown et al., 2000).

Given that there is more than one way to remove spatial pairwise correlation in a signal (using, for example, a center-surround filter with any arbitrary amount of phase alignment), our task then is to see what advantages a center-surround organization offers over other organizations. We find that center-surround organization of DoG filters gives a sparse response compared to other whitening filters given natural scenes. We also find that the vector length of the filter increases with frequency in such a way as to maximize the use of the dynamic range available in ganglion cells. That is, if cells with different frequency sensitivity give the same response magnitude on average, they have clearly made the best use of the dynamic range of this cell type. So the early visual system would appear to be using two complimentary strategies in the way it codes natural scenes: In addition to equalizing responses across frequency, ganglion cells produce a sparse response given natural scenes.

But a response that is sparse across the population and uniform across spatial frequency is not the primary goal of retinal processing. If we sought a maximally sparse response, no cells would ever fire since this response distribution would have a very high kurtosis. These studies do not resolve the question of why center-surround organization is highly conserved for

receptive fields across species. The answer to this question likely requires a broader theory that incorporates noise reduction, nonlinearities, adaptation and other temporal properties.

5.6 Future Work

Future work will examine nonlinearities in detail, as well as temporal properties, for which vector length sensitivity may again be useful. An important consideration in evaluating the sparsifying properties of the retina is to examine how the retina responds to white (uncorrelated) noise stimuli. We believe that the retina produces a sparse output because natural scenes are sparse. If the input to the retina does not have the same sparse statistics as natural scenes, we would expect a less sparse response. For example, we would expect that white noise would produce a less sparse response than natural scenes in retinal ganglion cells. Physiological recordings in the retina (Berry et al. 1997) and LGN (Reinagel and Reid, 2000) suggest that random flickering and white noise stimuli both produce sparseness in these cells' output. Berry et al. use activity (i.e., the fraction of the time that the tested neuron spends firing) as their sparseness measure. We would expect sparseness to be much lower for random stimuli. But to our knowledge, no one has directly compared the sparseness of responses to natural scenes with responses to white noise. The results of Berry et al. (1997) and Reinagel and Reid (2000) are important and should be investigated further.

It is possible that a center-surround arrangement requires a minimum of dendritic wiring given its task. Such arguments have been used in the context of cortex (Mitchison, 1991). Given the task of achieving a whitened, sparse output, retinal ganglion cells could be organizing themselves in such a way as

to require the same amount of wiring for all regions of the cell. Anatomical studies (Freed and Sterling, 1988) suggest that dendritic trees of ganglion cell centers do a linear summation of nonlinear inputs but there is debate as to the nature of this process. Regardless of the relationship between sensitivity and dendritic tree structure, a strategy involving minimum wiring would mean that a simple set of rules could guide construction of the receptive fields. One possible test of the minimum wiring hypothesis in ganglion cells would involve a neural network architecture that is designed to search for receptive fields that whiten and/or sparsify with a minimum of connectivity.

The study of retinal ganglion cell processing is crucial to our understanding of higher visual processing. That retinal structure and organization is remarkably well-conserved across mammalian species implies that the retina has found an exceedingly effective solution to visual coding given the requirements and objectives of mammalian vision and the constraints of neurophysiology.

APPENDIX A

NEURONAL PHYSIOLOGY AND CORTICAL ORGANIZATION

A.1 Neurons

Neurons come in three broad classes: motor neurons, interneurons and sensory neurons, of which the latter two are important for us. Sensory neurons (Fig. A.1) transduce electrical signals from physical stimuli. Interneurons are

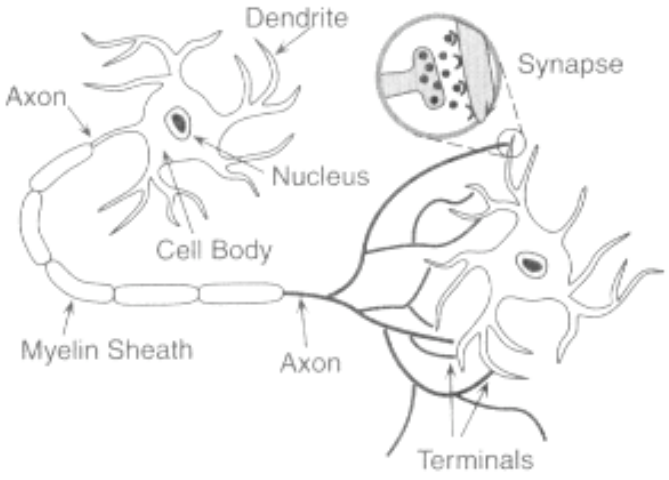


Figure A.1. Schematic diagram of a sensory neuron. Information is transmitted across the synapse by neurotransmitters. The signal propagates from the dendrites to the cell body and down the axon which synapses with other neurons. (Palmer, 1999)

by far the most common type of neuron and the neurons that comprise this class include all neurons that are not specifically involved in movement or transduction (Kandel et al., 2000). One major distinction between sensory neurons and interneurons is that the former generally respond with graded signals, meaning that they respond to stimuli with changes in membrane

potential. Most interneurons respond with action potentials, which are discrete spikes in voltage, each of which has approximately the same change in voltage. Action potential responses are measured in spikes per second. When given a stimulus of a certain duration, spiking neurons in general respond either with a transient signal at the onset and/or offset of the stimulus or with a sustained response throughout stimulation. Neurons adapt to unchanging stimuli by reducing their sensitivity. For example, retinal neurons (besides the photoreceptors) have a fast adaptation (<0.1 s) and a slow adaptation (~10 s) to changes in contrast (Baccus and Meister, 2002).

Excitatory signals at the dendrites increase the firing rate of the post-synaptic cell above the basal rate. These signals are passed by neurotransmitters such as glutamates. Inhibitory connections reduce the firing rate of below the basal rate and are often passed by the neurotransmitter GABA (Kandel et al., 2000). The cell sums these signals and if the input activation exceeds the threshold, the cell fires.

We assume that the cell's firing rate contains all of the information communicated by the cell. Some studies assume that a spike train contains information in the specific timing of spikes, which is called temporal coding (see Victor, 1999; Sen et al 1996; Softky, 1994; Gawne, 2000). This is a crucial point: the majority of studies of neuron sensitivity look for correlations between the firing rate and the specific pattern of stimulation. There is evidence that temporal coding is at work in the nervous system: some chemoreceptors code different chemical stimuli with different patterns of spike intervals (Kandel et al., 2000); some auditory neurons in cortex code location with spike patterns (Middlebrooks et al., 1994); and the patterns of firing across neural populations (Gray and Swets, 1989, Meister et al., 1995)

has been proposed as a factor in coding.[15] This a large and growing area of research that could have important implications for our ideas about efficient coding of sensory information.

A.2 Photoreceptors

Light entering the eye is refracted by the cornea and the lens onto the retina. Rod and cone photoreceptors cover the retina but rods are excluded from the fovea (the fovea is the central 2° of the retina that has the largest concentration of cones and the highest spatial acuity). Photons in the range of visible light (400-700 nm) enter the photoreceptors where they change the physical structure of photopigment molecules, which leads to transduction into neural signals.

In humans, there are three types of cones and one type of rod. Rods are sensitive to low-light (scotopic) conditions and saturate at high (photopic) intensities. Cones are sensitive in photopic conditions and are maximally sensitive to either short-, medium- or long-wavelengths of light. Color is a perceived quality of an object that is reconstructed from the trichromatic luminance information gathered by cones. It is based on incident light and on the reflectance of objects, an unchanging physical attribute of materials. The fact that perceived color is relatively invariant to ambient lighting (i.e., incident lighting with different frequency spectra) is called color constancy. Trichromatic vision has been shown in principle to be capable of achieving

[15] In this last type of code—so-called population codes—the number of cells that responds to a stimulus is found to increase with stimulus intensity. Recent work suggests that population codes express the uncertainty about the stimulus in terms of the probability density function across neurons (Pouget, 2003) although there is much debate about the role of population codes (Pola, 2003).

color constancy over a limited range of frequencies (Maloney, 1986).

Photoreceptors are one of the few neuronal classes that pass graded responses. They synapse to two types of interneurons: to horizontal cells, which convey information among photoreceptors; and to bipolar cells, which connect to amacrine cells (another interneuron) and to ganglion cells. Ganglion cells are one of the first units to respond with true action potentials.[16] Each neural spike has approximately the same amplitude change in electrical potential, and the rate of spiking varies with stimulus intensity according to rules we will discuss (Kandel et al., 2000). Ganglion cell physiology and processing is discussed in Chapter 5.1. The optic nerve, which is composed entirely of ganglion cell axons, projects to the lateral geniculate nucleus, as described in Chapter 2.1.

A.3 Cortex

Projections leaving the retina from the left and right visual fields connect to LGN in alternating layers that are parallel to the unit's surface. Neurons from LGN project to the V1 area of cortex (Fig. A.2), where neurons that project from the left and right eyes are stacked into ocular dominance columns or "slabs" (Palmer, 1999). These slabs are each comprised of orientation columns perpendicular to the surface, whose cells respond only to oriented patterns of luminance from a given eye (Fig. A.3). That is, orientation columns contain cells whose receptive fields are oriented in a particular spatial direction and respond to input from a specific eye.

Three classes of neurons dominate V1: simple, complex and hypercomplex

[16] Some amacrine cells also spike.

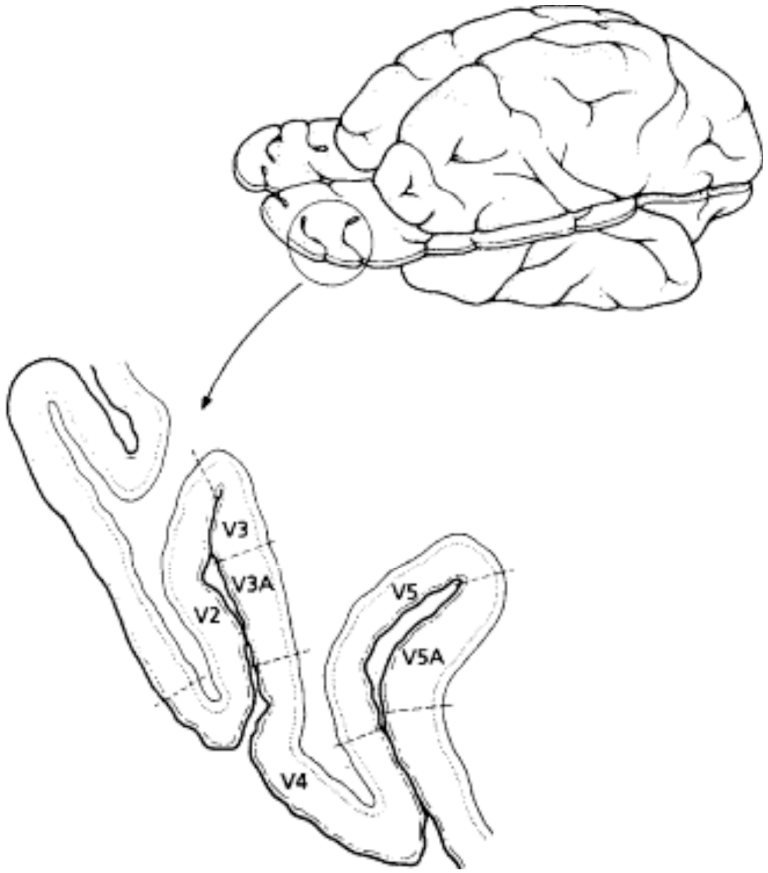


Figure A.2. Locations of visual areas of striate cortex. (Palmer, 1999)

cells.[17] Simple cells respond to bright and dark lines or edges that have a particular orientation, as Hubel and Weisel (1962) showed. But simple cell responses are affected by stimuli that are outside of the so-called classical receptive field (CRF) of Hubel and Weisel (De Valois and De Valois, 1988). In particular, lines parallel to but outside of the CRF can increase or decrease firing. The second group, complex cells, fires only weakly when presented with small, stationary spots but are sensitive to moving lines and edges. Many complex cells fire strongly for stimuli moving in a certain direction but are

[17] Because of sampling biases (and other biases), our current understanding of cortical processing may be drastically oversimplified. See Olshausen and Field (2004).

largely insensitive to where in the receptive field the motion occurs (Palmer,

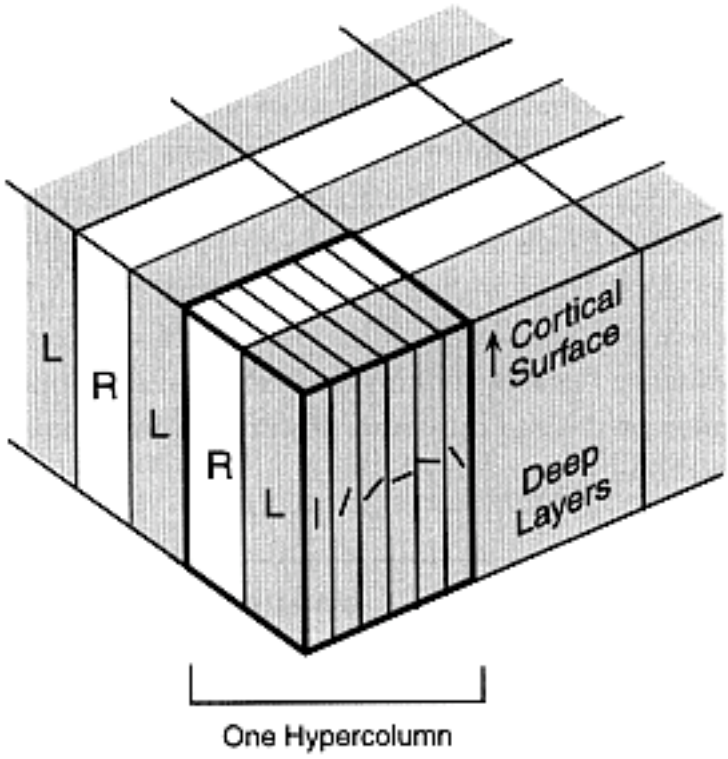


Figure A.3. A layer of cortical area V1 showing ocular dominance columns or "slabs" composed of projections that come from the two eyes (L and R). Perpendicular to the cortical surface are orientation columns that generally contain cells with oriented receptive fields. (Palmer, 1999)

1999). Hypercomplex cells, the third class, respond to lines and edges of a particular length, and their responses are modulated in a continuous way by stimuli that are longer or shorter than their receptive fields (De Angelis, Freeman and Ohzawa, 1995).

After V1, there are connections among dozens of distinct regions of visual cortex. Broadly, visual information leaving V1 travels along either the ventral of dorsal pathways, whose names refer to areas of the brain where high-level processing takes place. The ventral stream is thought to analyze form information, to figure out "what" we are looking at. The dorsal stream

analyzes "where" features are in the world and it may help coordinate motor interactions with the world (Ungerleider and Mishkin, 1982). There are believed to be subpathways that analyze and integrate color, form, motion and binocular information. The important point, however, is that our understanding of processing in V1 and beyond is incomplete.

REFERENCES

Ames, A. 1997. Energy requirements of brain function: when is energy limiting? In *Mitochondria and free radicals in neurodegenerative disease*. Eds. M. F. Beal, N. Howell and I. Bodis-Wollner. New York: Wiley.

Atick, J. J. 1992. Could information theory provide an ecological theory of sensory processing? *Netw. Comput. Neural Syst.* **3**, 213-215.

Atick, J. J. and Redlich A. N. 1990. Towards a theory of early visual processing. *Neural Comput.* **2**, 308-320.

Atick, J. J. and Redlich A. N. 1992. What does the retina know about natural scenes? *Neural Comput.* **4**, 196-210.

Attneave, F. 1954. Some informational aspects of visual perception. *Psychol. Rev.* **61**, 183-193.

Atwell, D. and Laughlin, S. B. 2001. An energy budget for signaling in the grey matter of the brain. *J. Cereb. Blood Flow Metab.* **21**, 1133-1145.

Arnett, D. W. 1978. Statistical dependence between neighboring retinal ganglion cells in goldfish. *Exp. Brain Res.* **32**, 49-53.

Arnett, D. and Sparker, T. E. 1981. Cross-correlation analysis of maintained discharge of rabbit retinal ganglion cells. *J. Physiol.* **317**, 29-47.

Baccus, S. A. and Meister, M. 2002. Fast and slow contrast adaptation in retinal circuitry. *Neuron* **36**, 909-919.

Baddeley, R., Abbott, L. F., Booth, M. C., Sengpiel, F., Freeman, T., et al. 1998. Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proc. Roy. Soc. B (London)* **264**, 1775-1783.

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

Banks, M. S., Geisler, W. S. and Bennett, P. J. 1987. The physical limits of grating visibility. *Vision Res.* **27**, 1915-1924.

Barlow, H. B. 1961. Possible principles underlying the transformation of

sensory messages. In: *Sensory Communication*, ed. W. A. Rosenblith. Cambridge, MA: MIT Press.

Barlow, H. B. 1989. Unsupervised learning. *Neural Comput.* **1**, 295-311.

Barlow, H. B. and Levick, W. R. 1965. The mechanism of directionally selective units in rabbit's retina. *J. Physiol.* **178**, 477-504.

Barlow, H. B. and Levick, W. R. 1976. Threshold setting by the surround of cat retinal ganglion cells. *J. Physiol.* **259**, 737-757

Baylor, D. A., Nunn, B. J. and Schnapf, J. L. 1987. Spectral sensitivity of the cones of the monkey *Macaca fascicularis J. Physiol.* **390**, 145

Bell, A. J. and Sejnowski, T. J. 1997. The "independent components" of natural scenes are edge filters. *Vision Res.* **37**, 3327–3338.

Benardete, E. A. and Kaplan, E. 1997a. The receptive field of the primate P retinal ganglion cell, I: linear dynamics. *Vis. Neurosci.* **14**, 169-186.

Benardete, E. A. and Kaplan, E. 1997a. The receptive field of the primate P retinal ganglion cell, II: nonlinear dynamics. *Vis. Neurosci.* **14**, 187-205.

Berry, M. J. Warland, D. K. and Meister, M. 1997. The structure and precision of retinal spike trains. *Proc. Natl. Acad. Sci. USA* **94**, 5411-5416.

Borghuis, B. G., van de Grind, W. A. and Lankheet, M. J. M. 2002. *The role of nonlinearities in front-end visual responses to natural stimuli.* part of Ph.D. dissertation, University of Utrecht, Utrecht, Holland.

Brady, N. and Field, D. J. 2000. Local contrast in natural images: normalisation and coding efficiency. *Perception* **29**, 1041-1055.

Brown, G. D., Yamada, S. and Sejnowski, T. J. 2001. Independent components analysis at the neural cocktail party. *Trends Neurosci.* **24**, 54-63.

Brown, S. P., He, S. and Masland, R. H. 2000. Receptive field microstructure and dendritic geometry of retinal ganglion cells. *Neuron* **27**, 371-383.

Burton, G.J., Moorhead, I. R. 1987. Color and spatial structure in natural scenes. *Appl. Optics* **26**, 157-170.

Buxton R. B. 2001. *Introduction to Functional Magnetic Resonance Imaging.* Cambridge: Cambridge University Press.

Chichilnisky, E. J. 2001. A simple white noise analysis of neuronal light responses. *Netw. Comput. Neural Syst.* **12**, 199-213.

Cohn, T. E. 1976. Quantum fluctuations limit foveal vision. *Vision Res.* **16**, 573-579.

Croner, L. J. and Kaplan, E. 1995. Receptive fields of P and M ganglion cells across the primate retina. *Vision Res.* **15**, 7-24.

Dan Y., Atick J. J., Reid R. C. 1996. Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. *J. Neurosci.* **16**, 3351–3362.

Darwin, C. *On the Origin of Species*. London: John Murray, 1859 (186).

Dayan, P. and Abbott, L. F. 2001. *Theoretical Neuroscience.* Cambridge: MIT Press.

De Angelis, G. C., Freeman, R. D. and Ohzawa, I. 1994. Length and width tuning of neurons in the cat's primary visual cortex. *J. of Neurophysiol.* **71**, 347-374.

De Valois, R. L., Morgan, H., and Snodderly, D. M. 1974. Psychophysical studies of monkey vision-III. spatial luminance contrast sensitivity tests of macaque and human observers. *Vision Res.* **14**, 75-81.

De Valois, R. L. and De Valois, K. K. 1988. *Spatial Vision*. New York: Oxford Press.

DeVries, S. H. 1999. Correlated firing in rabbit retinal ganglion cells. *J. Neurophysiol.* **81**, 901-920.

DiLorenzo, P. M. 1989. Across unit patterns in the neural response to taste: vector space analysis. *J. Neurophysiol.* **62**, 823-833.

Dong, D. W. and Atick, J. J. 1995a. Statistics of natural time-varying images. *Netw. Comput. Neural Syst.* **6**, 345-358.

Dong, D. W. and Atick, J. J. 1995b. Temporal decorrelation: a theory of lagged and nonlagged responses in the lateral geniculate nucleus. *Netw. Comput.*

Neural Syst. **6**, 159-178.

Dong, D. W. 2000. Spatiotemporal inseparability of natural images and visual sensitivities. In: *Motion Vision: Computation, Neural and Ecological Constraints.* Eds. J. M. Zanker and J. Zeil. Berlin: Springer Verlag.

Enroth-Cugell, C. and Robson, J. G. 1966. Contrast sensitivity of retinal ganglion cells of the cat. *J. Physiol.* **187**: 517-552.

Enroth-Cugell, C. and Freeman, A. W. 1987. The receptive field structure of cat retinal Y cells. *J. Physiol.* **384**, 49-79.

Fernald, R. D. 1997. The evolution of eyes. *Brain Behav. Evol.* **50**, 253-259.

Fernald, R. D. 2000. Evolution of eyes. *Curr. Opin. Neurobiol.* **10**, 444-50.

Field, D. J. 1987. Relations between the statistics of natural images and the response profiles of cortical cells. *J. Opt. Soc. A.* **4**, 2379-2394.

Field, D. J. 1993. Scale-invariance and self-similar 'wavelet' transforms: an analysis of natural scenes and mammalian visual systems. In: *Wavelets, Fractals and Fourier Transforms: New Developments and New Applications*, eds. M. Farge, J. C. R. Hunt and J. C. Vassilicos. Oxford: Oxford University Press.

Field, D. 1994. What is the goal of sensory coding? *Neural Comput.* **6**, 559-601.

Field, D. J. and Brady, N. 1997. Wavelets, blur and the sources of variability in the amplitude spectra of natural scenes. *Vision Res.* **37**, 3367-3383.

Finlay, B. L., de Lima Silveira, L. C. and Reichenbach, A. 2004. Comparative aspects of visual system development. To appear in: *The Structure, Function and Evolution of the Primate Visual System*, ed. J. Kremers. New York: John Wiley.

Foldiak, P. 1990. Forming sparse representations by local anti-hebbian learning. *Biol. Cybernet.* **64**, 165-170.

Freed, M. A. and Sterling, P. 1988. The ON-alpha ganglion cell of the cat retina and its presynaptic cell types. *J. Neurosci.* **8**, 2303-2320.

Gawne, T. J. 2000. The simultaneous coding of orientation and contrast in the responses of V1 complex cells. *Exp. Brain Res.* **133**, 293-302.

Geisler, W. S., Perry, J. S., Super, B. J. and Gallogly, D. P. 2001. Edge co-occurrence in natural images predicts contour grouping performance. *Vision Res.* **41**, 711-724.

Gray, C. M. and Swets J. A. 1989. Chattering cells: superficial pyramidal neurons contributing to the generation of synchronous oscillations in the visual cortex. *Science* **274**, 109-113.

Hartline, H. K. 1940. The receptive fields of optic nerve fibers. *Am. J. Physiol.* **130**, 690-699.

Hubel, D. H., Weisel, T. N., 1962. Receptive fields, binocular interaction and functional architecture in the cat's striate cortex. *J. Physiol.* **160**, 106-154.

Huber, P. J. 1985. Projection pursuit. *Annals of Statistics* **13**:435-475.

Hyvarinen, A. and Hoyer, P. O. 2000. Emergence of phase- and shift-invariant features by decomposition of natural images into independent feature subspaces. *Neural Comput.* **12**, 1705-20.

Johnsen J. A. and Levine, M. W. 1983. Correaltion of activity in neighbouring goldfish ganglion cells: relationship between latency and lag. *J. Physiol.* **345**, 439-449.

Kandel, E. R., Schwartz, J. H. and Jessell, T. M. 2000. *Principles of Neural Science*. New York: McGraw-Hill.

Kaplan, E. 1991. The receptive field structure of retinal ganglion cells in cat and monkey. In: *The Neural Basis of Visual Function*. Ed. A. G. Leventhal. Boca Raton, FL: CRC Press.

Kersten, D. 1987. Predictability and redundancy of natural images. *J. Opt. Soc. Am. A* **4**, 2395-2400.

Kretzmer ER (1952) Statistics of television signals. *The Bell System Technical Journal* 751-763

Kelly, D. H. 1972. Adaptation effects on spatio-temporal sine-wave thresholds. *Vision Res.* **12**, 89-101.

Land, M. 1985. The Eye: Optics. In *Comprehensive Insect Physiology, Biochemistry and Pharmacology*. Eds. G. A. Kerkut and L. I. Gilbert. London: Pergamon.

Kety, S. S. 1957. The general metabolism of the brain in vivo. In *Metabolism of the Nervous System*. Ed. D. Richter. London: Pergamon.

Krauskopf, J. and Reeves, A. 1980. Measurement of the effect of photon noise in detection. *Vision Res.* **20**, 193-196.

Kuffler, S. W. 1953. Discharge patterns and functional organization of mammalian retina. *J. Neurophysiol.* **16**: 37-68.

Laughlin, S. B. 1981. A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch.* **36C**, 910-912.

Laughlin, S. B., de Ruyter van Steveninck, R. R. and Anderson, J. C. 1998 The metabolic cost of neural information. *Nature Neurosci.* **1**, 36-41.

La Vail, M. M., Rapaport, D. H. and Rakic, P. 1991. Cytogenesis in the monkey retina. *J. Comp. Neurol.* **309**, 86-114.

Lennie, P. 2003. The cost of cortical computation. *Curr. Biol.* **13**, 493-497.

Lewicki, M. S. and Olshausen, B. A. 1999. Probabilistic framework for the adaptation and comparison of image codes. *J. Opt. Soc. Am.* **16**, 1587-1601.

Levy, W. and Baxter, R. A. 1996. Energy efficient neural codes. *Neural Comput.* **8**, 531-543.

Mach, E. 1886. *The Analysis of Sensations.* Trans. C. M. Williams. Chicago: Open Court, 1914.

Maloney, L. T. 1986. Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *J. Opt. Soc. Am. A* **3**, 1673-1683.

Masland, R. H. 2001. The fundamental plan of the retina. *Nature Neurosci.* **4**, 877-887.

Mastronarde, D. N. 1989. Correlated firing of retinal ganglion cells. *Trends Neurosci.* **12**, 75-80.

Meister, M., Lagnado, L. and Baylor, D. A. 1995. Concerted signaling by retinal ganglion cells. *Science* **270**, 1207-1210.

Meister, M. Multineuronal codes in retinal signaling. *Proc. Nat. Acad. Sci. USA*

93, 609-614.

Meister, M. and Berry, M. J. 1999. The neural code of the retina. *Neuron* **22**, 435-450.

Middlebrooks J. C., Clock, A. E., Xu, L., Eddins, A. C., and Green, D.M. 1994. A panoramic code for sound localization by cortical neurons. *Science* **264**, 842-844.

Mitchison, G. 1991. Neuronal branching patterns and the economy of cortical wiring. *Proc. Roy. Soc. B Biol.* **245**, 151-158.

Naka, K. I. and Rushton, W. A. 1966. S-potentials from colour units in the retina of fish (*Cyprinidae*). *J. Physiol.* **185**, 536-555.

Nirenberg, S. Carcieri, S. M., Jacobs, A. L. and Latham, P. E. 2001. Retinal ganglion cells act largely as independent encoders. *Nature* **411**, 698-701.

Norman, R. A. and Werblin, F. S. 1974. Control of retinal sensitivity. I. Light and dark adaptation of vertebrate rods and cones. *J. Gen. Physiol.* **63**, 37-61.

Oliva, A., Torralba, A., Guerin-Dugue, A. and Herault, J. 1999. Global semantic classification of scenes using power spectrum templates. In: *Proceedings of The Challenge of Image Retrieval* (CIR99), Springer Verlag BCS Electronic Workshops in Computing series, Newcastle, UK.

Olshausen, B. A. and Field, D. J. 1996. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381**, 607-609.

Olshausen, B. A. and Field, D. J. 2004. What is the other 85% of V1 doing? In *Problems in Systems Neuroscience.* Eds. T.J. Sejnowski, L. van Hemmen. Oxford: Oxford University Press.

Palmer, S. E. 1999. *Vision Science*. Cambridge, MA: MIT Press.

Pearson, K. 1892. *The Grammar of Science*. London: Walter Scott.

Pelli, D. G. 1981. *Effects of visual noise.* Ph. D. thesis. Cambridge University, Cambridge, England.

Pelli, D. G. 1991. Noise in the visual system may be early. In *Computational Models of Visual Processing.* Eds. M. Landy and J. A. Movshon. Cambridge:

MIT Press.

Pola, G., Thiele, A., Hoffman, K-P. and Panzeri S. 2003. An exact method to quantify the information transmitted by different mechanisms of correlational coding. *Netw. Comput Neural Syst.* **14**, 35-60.

Pouget, A., Zemel, R. and Dayan, P. 2003. Inference and computation with population codes. *Ann. Rev. Neurosci.* **26**, 381-410.

Ramon y Cajal, S. [1906] 1967. The structure and connexions of neurons. In: *Nobel Lectures: Physiology or Medicine, 1901-1921*, 220-253. Amsterdam: Elsevier.

Rapaport, D. H., Fletcher, J. T., La Vail, M. M. and Rakic, P. 1992. Genesis of neurons in the retinal ganglion cell layer of the monkey. *J. Comp. Neurol.* **322**, 577-588.

Rapaport, D. H., Rakic, P. and La Vail, M. M. 1996. Spatiotemporal gradients of cell genesis in the primate retina. *Perspect. Develop. Neurobiol.* **3**, 147-159.

Ratliff, F. 1961. Inhibitory interaction and the detection and enhancement of contours. In *Sensory Communication*, ed. W. A. Rosenblith. Cambridge, MA: MIT Press.

Reinagel, P. 2001. How do visual neurons respond in the real world? *Curr. Op. Biol.* **11**, 437-442.

Reinagel, P. and Reid, R. C. 2000. Temporal coding of visual information in the thalamus. *J. Neurosci.* **20**, 5392-5400.

Reinagel, P. and Zador, A. M. 1999 Natural statistics at the center of gaze. *Netw. Comput Neural Syst.* **10**, 341-350.

Rieke, F., Bodnar, D. A. and Bialek, W. 1995. Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. *Proc. Roy. Soc. London B* **262**, 259-265.

Rieke, F. 2001. Temporal contrast adaptation in salamander bipolar cells. *J. Neurosci.* **21**, 9445-9454.

Rinaman, W. C. 1993. *Foundations of Probability and Statistics*. Fort Worth: Saunders.

Rodieck, R. W. 1965. Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Res.* **5**, 583-601.

Rodieck, R. W. and Stone, J. 1965. Response of cat retinal ganglion cells to moving visual patterns. *J. Neurophysiol.* **28**: 819-832.

Ruderman, D. L. 1994. Designing receptive fields for highest fidelity. *Netw. Comput. Neural Syst.* **5**, 147-155.

Ruderman D. L. and Bialek, W. 1994. Statistics of natural images: scaling in the woods. *Phys. Rev. Lett.* **73**, 814-817.

Sakai, H. M. and Naka, K. 1995. Response dynamics and receptive field organization of catfish retinal ganglion cells. *J. Gen. Physiol.* **105**, 795-814.

Schieber, F. 1992. Aging and the senses. In *Handbook of Mental Health and Aging*. Eds. J.E. Birren, R. Sloan and G. Cohen. New York: Academic Press.

Schnapf, J. L., Kraft., T. W., Nunn, B. J., Baylor, D. A. 1988. Spectral sensitivity of primate photoreceptors. *Vis. Neurosci.* **1**, 255-261.

Sen, K., Jorge-Rivera, J. C., Marder, E. and Abbott, L. F. 1996. Decoding synapses. *J. Neurosci.* **16**, 6307-18.

Shapley R. and Victor, J. D. 1979. Nonlinear spatial summation and the contrast gain properties of cat retinal ganglion cells. *J. Physiol.* **290**, 141-161.

Shapley R. and Victor, J. D. 1981. How the contrast gain control modifies the frequency responses of cat retinal ganglion cells. *J. Physiol.* **318**, 161-179.

Shannon C. E. 1948. A mathematical theory of communication. *Bell System Technical Journal* **27**, 379-423 and 623-656, July and Oct.

Simoncelli, E. P. and Olshausen, B. A. 2001 Natural image statistics and neural representation. *Ann. Rev. Neurosci.* **24**, 1193-1216.

Simoncelli, E. P. 2003. Vision and the statistics of the visual environment. *Curr. Op. Biol.* **13**, 144-149.

Softky, W. Sub-millisecond coincidence detection in active dendritic trees. *J. Neurosci.* **58**:13-41, 1994.

Srinivasan, M. V., Laughlin, S. B. and Dubs, A. 1982. Predictive coding: a fresh view of inhibition in the retina. *Proc. Roy. Soc. B Biol. Sci.* **216**, 427-459.

Strong, S. P., Koberle, R., de Ruyter van Steveninck, R. R., and Bialek, W. 1998. Entropy and information in neural spike trains. *Phys. Rev. Lett.* **80**, 197-200.

Tolhurst, D. J., Tadmor, Y. and Tang C. 1992. The amplitude spectra of natural images. *Ophthal. Physiol. Opt.* **12**, 229-232.

Torralba, A. and Oliva, A. 2003. Statistics of natural image categories. *Netw. Comput. Neural Syst.* **14**, 391-412.

Treves, A. and Rolls, E. T. 1991. What determines the capacity of auto-associative memories in the brain? *Netw. Comput. Neural Syst.* **2**, 371-397.

Ungerleider, L. G. and Mishkin, M. 1982. Two cortical visual systems. In *Analysis of Visual Behavior,* Eds. D. G. Ingle, M. A. Goodale, and R. J. Q. Mansfield. Cambridge: MIT Press.

Valeton, J. M. and van Norren, D. 1983. Light adaptation of primate cones: An analysis based on extracellular data. *Vision Res.* **23**: 1539-1547.

van Hateren, J. H. 1992. Theoretical predictions of spatiotemporal receptive fields of fly LMCs, and experimental validation. *J. Comp. Physiol. A* **171**, 157-170.

van Hateren, J. H. 1997. Processing of natural time series of intensities by visual system of the blowfly. *Vision Res.* **37**, 3407-3416.

van Hateren, J.H. and van der Schaaf, A. 1998. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. Roy. Soc. B Biol. Sci.* **265**, 359-366.

van Hateren J.H., Ruderman D.L. 1998. Independent Component Analysis of Natural Image Yields Spatiotemporal Filters Similar to Simple Cells in Visual Cortex. *Proc. Roy. Soc. B Biol. Sci.* **265**, 2315-2320.

Victor, J. D. 1987. The dynamics of cat retinal X cell centre. *J. Physiol.* **386**, 219-246.

Victor, J. D. 1999. Temporal aspects of neural coding in the retina and lateral geniculate. *Netw. Comput. Neural Syst.* **10**, R1-R66.

Victor, J. D. 2002. Binless strategies for estimation of information from neural data. *Phys. Rev. E* **66**, 051903.

Vinje, W. E. and Gallant, J. L. 2000. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* **287**, 1273-1276.

Wassle H. and Boycott B. B. 1991. Functional architecture of the mammalian retina. *Physiol. Rev.* **71**, 447-480.

Willmore, B. and Tolhurst, D. J. 2001. Characterizing the sparseness of neural codes. *Netw. Comput. Neural Syst.* **12**, 255-270.

Wiskott, L. and Sejnowski, T. J. Slow feature analysis: unsupervised learning of invariances. *Neural Comput.* **14**, 715-770.

Wong, R. O. L. 2000. Retinal waves and visual system development. *Ann. Rev. Neurosci.* **22**, 29-47.

Zrenner, E. and Gouras, P. 1983. Cone opponancy in tonic ganglion cells and its variation with eccentricity in rhesus monkey retina. In *Color Vision: Physiology and Psychophysics.* Eds. J. Mollon and T. Sharpe. New York: Academic Press.