Spatial Vision
Wilson S. Geisler and Duane G. Albrecht

I. Introduction 79
II. Single Neurons and Behavior 80
   A. Levels of Analysis 80
   B. Linking Hypotheses 83
III. Window of Visibility 84
   A. Space and Time: Retinal Coordinates 84
   B. Space and Time: Environmental Coordinates 85
   C. Naturalistic Viewing Conditions 85
   D. Retinal Eccentricity, Luminance, and Color 86
IV. Optics and Photon Noise 88
V. Retina and Lateral Geniculate Nucleus 89
   A. Selectivity 89
   B. Performance 96
VI. Primary Visual Cortex 99
   A. Selectivity 99
   B. Performance 111
VII. Implications for Object Recognition and Scene Interpretation 121

References 123
I. INTRODUCTION

The topic of spatial vision concerns the fundamental mechanisms within the eye and the brain that analyze and represent the distribution of light across the visual field, with the ultimate goal of understanding how these mechanisms contribute to object recognition and scene interpretation in general.

A wealth of psychophysical and physiological research supports the view that stimulus selectivity plays a fundamental role in spatial vision. Psychophysical studies have provided evidence that the human visual system is selective along a number of stimulus dimensions including orientation, size, position, wavelength (color), speed of motion, direction of motion, and binocular disparity. These studies have shown that there are mechanisms ("channels") selective to different regions along each of these stimulus dimensions. Similarly, neurophysiological and anatomical studies have demonstrated that neurons in the visual pathway are selective along a number of stimulus dimensions, and that this selectivity increases from the retina to the primary visual cortex. For example, photoreceptors are selective along a few stimulus dimensions (spatial position, wavelength, temporal frequency), whereas cortical neurons are selective along many stimulus dimensions (spatial position, wavelength, temporal frequency, orientation, spatial frequency, direction of motion, disparity, etc.). Concomitant with this increase in stimulus selectivity, there is an increase in the heterogeneity; that is, there is an increase in the complexity and diver-
sity of the cells along all of the stimulus dimensions. Thus, for example, the intensity response functions of cones are all very similar from cell to cell, whereas the contrast response functions of cortical neurons are quite different from cell to cell.

A number of different explanations have been proposed for this emergence of stimulus selectivity along the visual pathway. One explanation is that this progressive selectivity is part of a hierarchical process, ultimately leading to single neurons that respond uniquely to specific real-world objects (Barlow, 1972, 1995): "The Neuron Doctrine." A second explanation is that this selectivity reflects a low redundancy code that is well matched to the statistics of natural images (Barlow, 1961, 1989; Field, 1987; Olshausen & Field, 1997): "Sparse Coding."

A third explanation is that this selectivity is a critical step in segregating objects from their context: "Object Segregation." Objects of interest within the natural environment are generally located within a very complex context of other objects. In order to recognize an object of interest, the parts of the object must be separated from the parts of other objects. For example, to recognize a longhorn bull behind a barbed wire fence, it is necessary to separate the image features that define the wire fence from those that define the bull. Fortunately, context is much less of a problem on a local scale; it is relatively easy to identify the orientation, position, and color of local image contours. The selectivity of visual cortical neurons permits recognition of these local image properties, thus allowing subsequent grouping mechanisms to bind together the contours that define the fence separately from those that define the bull.

These three different explanations for stimulus selectivity are not necessarily incompatible. Object segregation or sparse coding could be a first step in producing single neurons tuned to real-world objects. On the other hand, the processing following object segregation or sparse coding could be highly distributed. Further, having neurons matched to the statistics of the natural environment must surely be advantageous for both sparse coding and object segregation, given the constraint of limited resources. However, the goals of sparse coding and object segregation are quite different; hence, the specific selectivities, and how they are implemented, could well be different. It is important to keep all three explanations in mind, given that one's theoretical viewpoint can substantively influence the direction of future research.

In this chapter we will rely upon a wealth of psychophysical and physiological research to develop the topic of spatial vision with two themes in mind. The first theme concerns how stimulus selectivity develops along the visual pathway. The second theme concerns how the anatomical and physiological mechanisms of stimulus selectivity contribute to visual performance, and ultimately, object recognition and scene interpretation.

II. SINGLE NEURONS AND BEHAVIOR

A. Levels of Analysis

Research in perception and cognition has been performed at many different levels of analysis: the organism as a whole, the subregions of the brain, the individual neu-
rons, and the components of individual neurons. The main focus in this chapter will be on the responses of the organism as a whole (i.e., visual psychophysics) and on the responses of the individual neurons (i.e., single neuron electrophysiology).

Focusing on the behavioral responses of the organism as a whole is justified because the ultimate goal of perception and cognition research is to understand behavior. Indeed, behavioral abilities define the phenomena of interest in perception and cognition, and hence provide the groundwork for directing and interpreting the measurements at all the other levels.

Specifically, behavioral findings often serve as a guide to neurophysiological and anatomical studies. To begin with, note that there is an enormous amount of information available in the visual stimulus. The visual system is not capable of transmitting and utilizing all of this information because it has finite neural resources. Some of this information will be used by the nervous system and some of this information will be lost. Behavioral performance can tell us what information is used and what is lost. If behavioral performance indicates that a certain kind of information is lost, then there should be some stage in the neural processing where the loss occurs, and neurophysiological measurements should be able to demonstrate this loss. If behavioral performance indicates that a certain kind of information is not lost, then the information must be preserved at every sequential level of the nervous system between the stimulus and the behavior, and neurophysiological measurements should be able to demonstrate the presence of the information.

Consider, for example, the task of color discrimination. There is an enormous number of different wavelength distributions that can be presented to a human observer. Behavioral studies have shown that most of these wavelength distributions cannot be distinguished. This fact indicates that substantial chromatic information has been lost in the nervous system. Neurophysiological measurements have demonstrated that most of this loss occurs at the level of the photoreceptors. On the other hand, behavioral studies have also shown that sufficient chromatic information is transmitted to support trichromacy. Neurophysiological studies have used this fact as a basis for designing chromatic stimuli and locating the neural systems responsible for trichromatic color discrimination.

The justification for focusing on the responses of individual neurons is somewhat more involved. All of the information transmitted by a neuron is contained in the temporal sequence of action potentials generated by the neuron; this information represents the net sum of all the synaptic inputs to the neuron, plus all of the prior neural information processing that those inputs represent. Because of the all-or-none property of action potentials and their short duration, it is relatively easy to accurately measure the temporal sequence (with a resolution finer than milliseconds) and hence all of the transmitted information (the noise in the recording instrument has little or no effect on the accuracy of the measurements). This fact makes it possible to quantitatively relate physiological responses to perception and cognition in a way that is not possible by current methods for measuring the responses of subregions of the brain or subcomponents of individual neurons.
When measuring the responses of a subregion in the brain using one of the neuroimaging techniques, the responses from the individual neurons are measured indirectly through metabolic activity, with relatively poor spatial and temporal resolution. Although certain questions can be answered with these techniques, it is very difficult to know how to relate the measurements to behavioral performance in a quantitative fashion. Similarly, although the responses of some particular subcomponents of individual neurons can be measured (such as the activity of a single synapse), a complete characterization of all the components is not possible at this point in time, and thus it is difficult to relate the responses of the subcomponents to the responses of the neuron as a whole or to behavioral responses.

There are, of course, limitations to the inferences that can be drawn about visual information processing by measuring action potentials from single neurons. To begin with, not all neurons at all levels of the visual system produce action potentials. However, these neurons are generally local circuit neurons. Neural communications that extend over distances greater than a few millimeters are carried via action potentials. Thus, measurement of action potential responses should be sufficient to characterize the output of the information processing that is transmitted from one region of the brain to another. Further, using intracellular recording (and similar techniques) it is generally possible to measure the information transmitted by the voltage potential responses of nonspiking neurons, with reasonable accuracy. A second limitation concerns the variation in sampling probabilities that may occur with different cell sizes and different subpopulation densities. However, anatomical measurements of subpopulation densities can help to lessen this uncertainty; and, increasing the sample size of measured neurons can increase the probability of obtaining measurements from small subpopulations. A third limitation is that it is not possible to record the individual responses of many neurons simultaneously. However, by measuring the responses of many neurons, one at a time, to the same set of stimuli, it is possible to make inferences about the information processing of

---

1 Neuroimaging techniques are useful because they allow simultaneous measurement of large populations of neurons and they can be performed on human beings (without invasive procedures). However, there are a number of factors which make neuroimaging data difficult to interpret. First, there is considerable uncertainty about the degree to which the measurements correspond to neural activity, because most techniques measure neural responses indirectly via the level of metabolic activity. Second, because the signals are summed over a relatively long temporal interval (on the order of seconds) it is difficult to obtain useful measurements of the rapid neural information processing that ordinarily occurs during perception and cognition (on the order of milliseconds). Third, because the signals are summed over space and time, many different patterns of neural activity within a given region could produce exactly the same summed level of metabolic activity, and thus it is impossible to quantitatively characterize the information transmitted by the population of neurons. Fourth, neural circuits for a given behavioral information-processing task are identified when the difference in metabolic activity between a baseline condition and a test condition exceeds a certain criterion level. Thus, to the extent that the neural circuits for a given information-processing task are distributed throughout different regions of the brain, or to the extent that the neural circuits for different information-processing tasks are intermingled, the changes in metabolic activity will fail to exceed the criterion. (For further discussion of this issue see Wandell, 1999.)
the population as a whole. These inferences depend upon knowing the degree of statistical independence of the responses across neurons. New techniques that allow recording from two or three neurons simultaneously suggest that the responses are relatively uncorrelated.

B. Linking Hypotheses

The ultimate goal of measuring the anatomical and physiological properties of the neurons in the visual pathway is to understand behavioral performance. For example, it is possible to measure the responses of visual cortex neurons as a function of spatial frequency and contrast, but ultimately we would like to understand how the responses of these neurons contribute to spatial frequency discrimination, contrast discrimination, and other behavioral performances. To understand how neural responses contribute to behavioral performance, it is necessary to have a linking hypothesis so that predictions for behavioral performance can be generated from the responses of the neurons at a particular level of the visual system (Brindley, 1960; Teller, 1984). A linking hypothesis can be regarded as a specific model of the neural processing which occurs subsequent to the selected level.

Two general categories of linking hypotheses have been widely considered. One consists of the Bayesian ideal-observer models; these models assume that all of the information available in the responses of the neurons is extracted and optimally combined with knowledge about stimulus probabilities. The other consists of simple pooling models; these models assume that the responses of the neurons are combined using a simple operation such as simple summation, Minkowski summation, or “winner take all.”

An advantage of the Bayesian ideal-observer models is that all of the available neural information is used, and hence if the behavioral performance exceeds the predicted performance, then one can be certain there is something missing in the neurophysiological data (e.g., an undetected subpopulation of neurons). Conversely, if the behavioral performance falls below that of the ideal observer, then either there must be something inaccurate about the neurophysiological data, or there must be some loss of information in subsequent neural stages. For some situations, the ideal combination of the neural responses is relatively simple but in general, it is complex and varies from situation to situation. Although the human visual system is sophisticated, it has many complex tasks to perform (with limited resources) and hence it probably does not perform any individual task in the optimal fashion.

An advantage of the simple pooling rules is that they are easy to understand, and they could be implemented with simple neural circuits. However, for many situations, the simple pooling rules probably do not adequately represent the complexity of later neural processing.

Given a linking hypothesis, it is then possible to compare neural performance and behavioral performance along stimulus dimensions of interest. For some linking hypotheses it is possible to compare absolute levels of performance. One example
is the "winner-take-all" rule, where the behavioral performance for each stimulus condition is determined by the neuron with the best performance. However, it is generally not possible to compare absolute levels of performance because of uncertainty about the number of neurons that are combined. In these cases, only the shapes of the behavioral and neural performance functions can be compared. Nonetheless, if the two performance functions have the same shape (i.e., if they only differ by a scale factor) then it suggests that subsequent brain mechanisms are equally efficient in processing the information along the stimulus dimension, and hence that the variations in the behavioral performance are the result of the variations in the neural responses at the selected level (or at a lower level). Of course, matching shapes for the neural and behavioral performance functions could be the result of a fortuitous cancellation of effects in subsequent neural levels, but such a cancellation would be improbable.

As is clear from this discussion, there is uncertainty about what the appropriate linking hypothesis might be in any particular comparison of neural and behavioral performance. Because of this uncertainty it is necessary to consider a range of possible linking hypotheses, and to evaluate, for each linking hypothesis, how accurately the predicted behavioral performance matches the measured behavioral performance. In some situations the choice of linking hypothesis has relatively little effect on the predicted behavioral performance (e.g., assumptions about subsequent neural processing have little effect on the predictions of behavioral color matching performance based upon the spectral sensitivity of the photoreceptors). In other situations, the choice of the linking hypothesis may have a substantial effect on the predictions. In these cases, the large differences in the predictions might be sufficient to exclude some linking hypotheses.

III. WINDOW OF VISIBILITY

The stimulus for vision is completely specified by describing light intensity as a function of position, time, and wavelength. The limits of what can be seen along these dimensions define the selectivity of the visual system as a whole: the so-called *window of visibility*.

A. Space and Time: Retinal Coordinates

The window of visibility has been measured using spatiotemporal sine waves (patterns that vary sinusoidally in intensity or wavelength over space and time). Each contour in Figure 1 represents the spatiotemporal frequencies that define the window of visibility for a given luminance contrast. These curves are based upon contrast thresholds (i.e., the smallest contrast that can be reliably detected) measured at photopic (daylight) mean luminances. All spatiotemporal frequencies that fall within a contour would be visible, and all those outside would be invisible. For example, the central pair of contours defines the set of stimuli that would be visible at 0.5% contrast, and the outer pair of contours defines the set of stimuli that
would be visible at 100% contrast, the maximum contrast. Note that all spatiotemporal frequencies outside of the 100% contour would be invisible, regardless of contrast.

B. Space and Time: Environmental Coordinates

The contours in Figure 1 describe the overall selectivity of the visual system in retinal coordinates. Although this is a very concise and general description, it is nevertheless useful to translate these limits into environmental coordinates (i.e., into objects moving in the natural environment). Each contour in Figure 2 represents the smallest size object that can be detected at a given speed, as a function of viewing distance. For stationary objects (the lowest contour), the smallest size that can be detected increases in direct proportion to distance; at one quarter of a meter, the nearest distance for good focus, the smallest size is approximately 0.06 mm; at 100 m the smallest size is approximately 20 mm. As physical speed increases, the smallest detectable size increases (the contours shift vertically). As viewing distance increases, the effect of speed decreases (the contours converge). Finally, note that at near viewing distances the effect of speed on spatial acuity decreases at higher speeds (the contours compress).

C. Naturalistic Viewing Conditions

In the laboratory, the detectability of spatiotemporal sine waves is often measured during steady fixation, where the duration of the stimulus can be quite long. Under these conditions, there is a substantial increase in contrast threshold for low spatial
frequencies at low temporal frequencies. However, under normal viewing conditions the eyes are almost always moving, and thus low temporal frequencies are relatively uncommon. When inspecting a static scene, the eyes jump from location to location quite frequently (saccadic eye movements), with fixation durations of approximately 200–500 ms (Carpenter, 1991). When tracking objects, the eyes move continuously (smooth pursuit eye movements), interspersed with saccadic corrections. Under these conditions, the most relevant region of the spatiotemporal visibility window is the region above 2–5 Hz (periods less than 200–500 ms). In this region there is little or no increase in threshold at low spatial frequencies. Similarly, when contrast thresholds are measured using short discrete presentations, with durations comparable to those of single fixations, there is little or no increase in threshold at low spatial frequencies (Arend, 1976; Banks, Sekuler, & Anderson, 1991; Robson & Graham, 1981). Figure 3 compares contrast threshold as a function of spatial frequency for short and long presentation durations.

D. Retinal Eccentricity, Luminance, and Color

The description of the window of visibility given above applies to stimuli centered on the point of fixation (i.e., at the fovea, on the visual axis). Although the temporal limits of visibility remain relatively constant across the entire visual field, the spatial limits decrease markedly away from the point of fixation (Wertheim, 1894). Figure 4 illustrates this decrease in spatial resolution for briefly presented stimuli; each contour represents the spatial frequencies that define the window of visibility across
the visual field for a given contrast. These contours show that $40^\circ$ away from the point of fixation the spatial limit of visibility has decreased by a factor of 16 (from 37 cpd to 2.3 cpd). Because of this decrease in spatial resolution away from the line of sight, eye movements are required to bring the high-resolution region of the system onto objects of interest.

The window of visibility changes continuously with light level (mean luminance). As the light level decreases, the overall sensitivity decreases with relatively

**FIGURE 3** Contrast threshold as a function of spatial frequency for short- and long-duration stimuli.

**FIGURE 4** Window of visibility for spatial frequency and eccentricity. Each contour represents the locus of stimuli that have the same contrast threshold, in percent. (Based upon Robson & Graham, 1981.)
greater loss in sensitivity at high spatial and high temporal frequencies. This decrease in sensitivity translates into a significant loss in real-world performance in terms of the visibility of moving objects at different distances.

The window of visibility for purely chromatic modulations in space and time differs from the window of visibility for purely luminance modulations. In comparison to the photopic luminance contours in Figure 1, there is a loss of sensitivity at high spatial and temporal frequencies, and an increased sensitivity at low spatial and temporal frequencies; that is, no low-frequency falloff (for a review see Mullen, 1985). However, under normal viewing conditions, where the eyes are moving and very low temporal frequencies are uncommon, the spatiotemporal chromatic sensitivity will generally be less than or equal to the luminance sensitivity.

IV. OPTICS AND PHOTON NOISE

The optics of the eye (the cornea, the pupil, and the lens) form an image on the retina that is a two-dimensional projection of the three-dimensional visual environment. The size, position, and shape of the retinal image can be determined by tracing rays from points in the environment through the center of the optics (the nodal point) until they intersect the retina.

The quality of the retinal image can be summarized with a point-spread function (a spatial impulse response function), which describes the two-dimensional distribution of light produced on the retina by a single point in space (e.g., a distant star). The visual environment can be regarded as a dense collection of discrete point sources. Thus, given knowledge of the point-spread function, and the center of the optics, it is possible to determine the retinal image for any arbitrary visual scene. The light from each source produces a distribution on the retina; the shape of the distribution is given by the point-spread function, the position by ray tracing, and the amplitude by the intensity of the source. The retinal image as a whole is then determined by summing these distributions.

Equivalently, the quality of the retinal image can be summarized with an optical transfer function, which describes the change in the amplitude and the phase of spatial frequency sine waves as they pass through the optics. Within this framework, the visual scene is decomposed into a collection of spatial frequency components. The amplitude and phase of each component are adjusted according to the optical transfer function. The retinal image as a whole is then determined by summing the adjusted components.

Point spread functions and optical transfer functions have been measured in the living human eye (Campbell & Green, 1965; Charman, 1993; Westheimer & Campbell, 1962). The results of these studies have shown that the quality of the retinal image is dependent upon many different factors; for example, pupil size, state of accommodation, eccentricity, wavelength, and individual differences. The best optical quality occurs on the optic axis, in a well-accommodated eye, with pupil sizes in the range 2–3 mm (the size of the pupil under normal daytime light levels).
Under these conditions, the optical quality approaches what would be expected from an ideal optical system. Nonetheless, even under these conditions, the optics are the major factor limiting the visibility of fine spatial detail (Banks, Geisler, & Bennett, 1987; Sekiguchi, Williams, & Brainard, 1993). For example, the highest spatial frequency transmitted by the optics is in the range of 50–60 cpd, which corresponds approximately to the highest visible spatial frequency (see Figure 1). Under other conditions (e.g., off the visual axis or at low light levels), other factors become more important in limiting the visibility of fine spatial detail.

Even with a perfect optical system, the quality of the image is ultimately limited by photon noise. By way of analogy, the pattern of photons falling on the retina through time is similar to the random pattern of raindrops falling on a dry surface. Specifically, the number of photons incident at a given point in the image, in a given amount of time, is random; as the average number of photons increases the variance increases in direct proportion.

It is important to keep in mind that optics and photon noise are the first factors limiting visual processing, thus their effects are seen all along the visual pathway; for example, it is impossible to recover the fine detail that is lost due to optical blur. The information-processing limits set by any given stage of the visual system will propagate to all subsequent stages.

V. RETINA AND LATERAL GENICULATE NUCLEUS

The image formed by the optical system is transduced by the photoreceptors (cones and rods), processed by the neurons of the inner nuclear layer (horizontal, bipolar, and amacrine cells) and then transmitted to subsequent brain regions via the axons of the ganglion cells, which comprise the optic nerve. The vast majority of ganglion cell axons project to the lateral geniculate nucleus (LGN), which in turn sends most of its axons to the primary visual cortex.

A. Selectivity

The retina does not encode and transmit all of the information available in the retinal image; it is selective to a limited range of wavelengths, luminances around the mean, and temporal frequencies. Furthermore, over most of the visual field, the retina encodes a limited range of the available spatial frequencies (only in the fovea is the full range encoded). Within these broad limits, there are subpopulations of neurons that are even more selective, subdividing the information into narrower ranges.

1. Space

The optics of the eye create a high-quality retinal image over much of the visual field. However, it is not possible to encode all of the information available in the entire image, given the limited resources of the visual system. To solve this problem,
the image is encoded with high resolution in the fovea and then the resolution decreases with eccentricity; eye movements are utilized to direct the fovea to points of interest. At each eccentricity, the spatial information is relayed from the retina to the visual cortex via separate and parallel neural systems, each performing a slightly different analysis.

The photoreceptors sample the retinal image at discrete locations. The cones, which are specialized for daytime light levels, are concentrated in the central retina, but present throughout the visual field. The rods, which are specialized for nighttime light levels, are more evenly distributed throughout the retina, but are absent in the fovea.

There are approximately 120 cones per degree of visual angle in the center of the fovea; this sampling density is sufficient to represent the full range of spatial frequencies transmitted by the optics (for a review see Williams, 1988). However, the sampling density of the cones decreases rapidly with eccentricity, reaching approximately 20 cones per degree, 10° away from the fovea. The optics of the eye degrade very gradually with eccentricity, and thus the peripheral cone sampling density is not sufficient to represent the full range of spatial frequencies present in the image. Further, the diameter of the cones, and hence their attenuation of high spatial frequencies, increases with eccentricity.

The responses of the cones are relayed to the cortex along a pair of parallel pathways that are first established at the level of the bipolar cell. In the central retina, the dendrites of midget bipolar cells make synaptic connections with a single cone, and the dendrites of diffuse bipolar cells make synaptic connections with a small cluster of cones. As eccentricity increases, the dendritic field diameter for both classes increases such that the ratio of the diameters between them remains approximately constant. Midget and diffuse bipolar cells appear to make synaptic connections with midget and diffuse ganglion cells. The midget ganglion cells project to the parvocellular (P) layers of the LGN, the diffuse ganglion cells to the magnocellular (M) layers. Sampling density (cells/deg) is about three times greater for the cells in the P pathway than in the M pathway, and this ratio remains approximately constant over eccentricity. The decrease in the density of M and P cells is more rapid than the decrease in the density of cones; thus, in the peripheral retina there is a further loss of high spatial frequency information, beyond that imposed by the cones.

Single neuron electrophysiological measurements of ganglion cells and LGN cells have shown that the receptive field center sizes of neurons in the P and M pathways increase with eccentricity in a fashion similar to the diameters of the dendrites of the bipolar cells. Further, the ratio of the receptive field center sizes remains approximately constant (3:1) with eccentricity. In the central retina, the receptive field center size of P cells is primarily determined by the optical point spread function, because a single P cell appears to be driven by a single cone. The diameter of the surround is approximately four to six times the size of the center, and the strength of the surround is approximately half that of the center, at all eccentricities, for both P and M cells (Croner & Kaplan, 1995; Derrington & Lennie, 1984).
Neurons in the P and M pathways are selective for spatial position and spatial frequency. The P and M pathways only encode and transmit spatial frequency information from a highly localized region of the visual field. The size of the center determines the highest spatial frequency that can be resolved (the spatial resolution), and the size of the surround determines the attenuation at lower spatial frequencies. Because the ratio of the surround size to the center size is large, and the ratio of the surround strength to the center strength is small, both P and M neurons are selective to a broad range of spatial frequencies. Because the receptive fields of M cells are larger than P cells, at any given eccentricity, the M cells are selective to an overlapping but lower band of spatial frequencies. Because the receptive field sizes of both P and M cells increase with eccentricity, their spatial frequency tuning curves shift to lower frequencies with eccentricity. In sum, for each location in space, the information transmitted to the cortex is subdivided into two spatial frequency bands by the P and M pathways.

The P and M layers of the LGN are separated by layers of very small cells, the koniocellular (K) layers (Kaas, Huerta, Weber, & Harting, 1978; for a review see Casagrande, 1994). These layers receive inputs from small ganglion cells including the bistratified ganglion cells that appear to carry chromatic information from the S (blue) cones (Dacey, 1996; for a review see Rodieck, 1998).

The responses of the rods are transmitted to the midget and diffuse ganglion cells (and hence into the P and M pathways) through a specialized sequence of neurons: the rod bipolar followed by the AII amacrine; rod responses may also be transmitted via small ganglion cells to the K layers of the LGN (for reviews see Rodieck, 1998; Sterling, 1998). Although the sampling density of the rods in the periphery is nearly as great as that of the cones in the central fovea, there is substantial spatial pooling by the bipolar and amacrine cells. The combination of dense sampling and spatial pooling optimizes sensitivity to low light levels, at the expense of decreased spatial resolution. However, this is a reasonable trade-off because, at low light levels, high spatial resolution information is not available in the stimulus due to photon noise.

2. Contrast and Luminance

Objects of interest in natural visual scenes generally consist of surfaces that are illuminated by primary light sources, such as the sun. These surfaces reflect a certain percentage of the incident light, typically 10 to 90%. Thus, to obtain useful information about the environment, the visual system must be sensitive to small reflectance differences within this range. These small reflectance differences produce small differences in luminance around a mean level. Further, a specific reflectance difference produces a specific luminance contrast independent of the mean luminance. Thus, the crucial information that must be encoded by the retina and LGN is the spatial distribution of local contrast.

During a 24-h period, the intensity of natural light outdoors varies over an enormous range, approximately 12 orders of magnitude, and thus the visual system must
be able to encode small contrasts over an extraordinarily wide range of mean luminances. This poses a serious encoding problem, given the limited response range of neurons. The visual system partially solves this problem by encoding contrast information in two separate and parallel neural pathways, the rod (scotopic) and cone (photopic) systems. In addition, within each pathway, there are a variety of different adaptation mechanisms for adjusting luminance sensitivity to compensate for the variations in mean luminance.

As light level increases, the response of a photoreceptor increases linearly and then gradually saturates. In a cone, absorption of approximately 2000 photons, during a brief presentation, will produce a response equal to half the maximum; whereas in a rod, absorption of approximately 30 photons will produce a response equal to half the maximum. The temporal integration time for a cone is approximately 50 ms, and for a rod it is approximately 150 ms. The rods and the cones show only a small amount of light adaptation, except at very high light levels where photopigment depletion occurs (for reviews see Hood, 1998; Walraven, Enroth-Cugell, Hood, MacLeod, & Schnapf, 1990). These facts, combined with the known changes in pupil size with background luminance, imply that rods respond linearly up to nearly 1 cd/m² (a cloudy sky just after sunset) and cones respond linearly up to nearly 1,000 cd/m² (a cloudy sky at noon). Thus, over most of the light levels encountered in the natural environment, one of the two systems will be operating in an approximately linear range: over the scotopic range the rods respond linearly, and over most of the photopic range the cones respond linearly.

To increase sensitivity to small changes in reflectance around a given mean luminance, the bipolar and ganglion cells amplify the photoreceptor responses. (Purpura, Tranchina, Kaplan, & Shapley, 1990; Virsu & Lee, 1983). The slopes of the intensity response functions (on linear axes) for bipolar and ganglion cells are steeper than the are for the photoreceptors, and they saturate at lower intensities. For example, under dark-adapted conditions, foveal ganglion cells (both P and M) reach half their maximum response for brief presentations that produce a small number of photons absorbed by a single cone. Similarly, under dark adapted conditions, peripheral ganglion cells reach half their maximum response for brief presentations that produce less than one photon absorbed per rod. This amplification is produced through a variety of different mechanisms, which include both electrochemical mechanisms and spatial summation mechanisms. For example, there is a great deal of spatial pooling of photoreceptor responses in the peripheral ganglion cells and bipolar cells (especially over the rods).

The high amplification of photoreceptor responses seen in the ganglion cells results in saturation at relatively low intensities. Thus, light adaptation mechanisms are required in order for small changes in reflectance to be signaled at higher mean luminance levels. In general, as mean luminance increases, the adaptation mechanisms (both multiplicative and subtractive) adjust the response characteristics of post-receptor neurons so that the neurons become less sensitive to absolute intensity differences but more sensitive to relative intensity ratios (luminance contrasts).
At higher mean luminance levels the sensitivity of the neurons to intensity ratios is constant, and there is little or no change in the maintained activity with mean luminance; thus the response to the mean luminance is largely removed. In other words, at higher mean luminances post-receptor neurons respond primarily to the contrast of the stimulus. Figure 5 shows the typical responses of M and P cells as a function of the contrast of sine wave gratings. As can be seen, M cells have higher contrast gain and larger maximum response (Kaplan & Shapley, 1986; Sclar, Maunsell, & Lennie, 1990).

Diffuse and midget bipolar cells come in two varieties, on-cells and off-cells. The responses of on-bipolars increase to increments of light on the center (white), and the responses of off-bipolars increase to decrements of light in the center (black). On-cells and off-cells make synaptic contacts with the very same cones. The separation of contrast information into on-and-off pathways, which continues up to the
cortex, has a number of potential benefits. First, having two independent neurons sample in exactly the same retinal location increases sensitivity. Second, separate on and off pathways make it possible to encode increases and decreases around the mean luminance level with low maintained activity. Low maintained activity results in substantial savings in metabolic resources, and also allows the full dynamic range of a neuron to be devoted to half of the intensity range. These advantages of on and off pathways are particularly evident in the visual cortex, where the number of neurons increases dramatically and the maintained discharges are quite small.

The amount of information that can be transmitted along the visual pathway is limited by the amount of noise (random variability) in the neural responses. The noise in the neural responses is a combination of noise within the stimulus (e.g., photon noise) and noise within the nervous system (for a review see Pelli, 1990). As the intensity of the light increases, the variability in the number of photons in the retinal image increases in direct proportion to the intensity (the variance is equal to the mean). Near scotopic threshold, a small number of absorbed photons can produce a reliable response in a ganglion cell (Barlow, 1981), and thus under these stimulus conditions the effect of photon noise must be substantial. Under photopic conditions, measurements of the intensity response functions of ganglion cells have shown that although the mean of the response increases with the intensity, the variance of the response remains nearly constant (Croner, Purpura, & Kaplan, 1993)—the variance increases in proportion to the mean response with a slope of only 0.2 (see Figure 5).² This suggests that under most photopic conditions the neural noise is considerably larger than the photon noise, and thus it dominates, or hides, the effect of the photon noise.

3. Time

In order to react quickly to events that occur in the environment, neurons in the visual system must be able to respond quickly to changes in intensity or wavelength. The response latency of the visual system is limited by the time interval over which light is integrated: the shorter the interval, the shorter the response latency. Two of the important factors that determine the temporal integration interval are light adaptation and compensation for neural and photon noise.

As described earlier, the light adaptation mechanisms are utilized to keep neural responses within a dynamic range across variations in mean luminance. One of the mechanisms that the visual system utilizes is to shorten the temporal integration interval as mean luminance increases. The effects of this can be seen in the temporal contrast sensitivity of ganglion cells. As light level increases, the high temporal frequency resolution increases.

The amount of noise (both neural noise and photon noise) sets strong constraints on the temporal integration interval. As light level decreases, the signal-to-noise

² This result holds even when the analysis is restricted to increases above the maintained activity (unpublished observations).
level in the retinal image decreases. Thus, to the extent that neural responses are dominated by photon noise, light must be summed over longer periods of time to obtain a reliable representation. In comparison to the photopic system, the scotopic system has a longer integration time, which results in slower reaction times and lower temporal frequency resolution. For example, the integration time for a rod is approximately three times longer than the integration time for a cone, the peak response occurs approximately 100 ms later, and the highest temporal frequency that can be resolved is approximately three times lower. This difference in the photopic and scotopic systems is probably due at least in part to the fact that the relative importance of photon noise (in comparison to neural noise) increases as light level decreases.

4. Color

In general, surfaces of objects in natural visual scenes differ simultaneously in both the overall amount of light that is reflected, and the relative amount of light that is reflected at different wavelengths. Thus, surface boundaries can be detected on the basis of either luminance or chromatic differences. The luminance contrast is usually quite a bit larger than the chromatic contrast.

As analysis of the responses of photoreceptors to natural scenes has shown that the amount of information available for contour detection from the luminance differences is generally much greater than the amount of information from chromatic differences (Geisler, 1995). Although the chromatic contrasts in the natural environment can be substantial, the chromatic contrasts are greatly reduced after the photoreceptors, due to the overlap of the absorption spectra of the photopigments. The first two rows of Table 1 show the average chromatic and luminance contrasts that can be found in a random sample of natural surfaces at the cornea and then after the photoreceptors. The third row shows that the effective increase in total contrast due to chromatic information is only 5% after the photoreceptors. Nonetheless, chromatic information is crucial for a variety of tasks (other than contour detection), such as texture segregation, grouping, visual search, and so on. Further, spatial discriminations that are based solely upon chromatic differences can be quite good when the chromatic contrasts are high (for a review see De Valois, 1994).

3 The overlap of absorption spectra of the long and middle wavelength cones reduces the possibility of undersampling the retinal image, and it also reduces the effects of chromatic aberration.
The major goal of color vision is to help identify objects and materials on the basis of the spectral reflectance of surfaces. Unfortunately, the distribution of wavelengths that reaches the eye is the product of the spectral distribution of the surface reflectance and the light source. Furthermore, the eye contains only three classes of cone, each selective to a different band of wavelengths. Nonetheless, the visual system is often able to determine the approximate spectral reflectance of the surfaces because the spectral distributions of natural surface reflectances can be characterized quite well with only three parameters, and because in any given scene the light source remains relatively constant over all of the different surfaces. Although the light source is not known, it can be estimated by comparison of the photoreceptor responses from the different surfaces. (For example, retinal adaptation mechanisms help to compensate for the light source by averaging the reflected light over many fixations). Thus, three classes of photoreceptor are often sufficient to estimate approximately the spectral reflectance of the surface. (For a general discussion of the issues see Wandell, 1995.)

The neurons in the P pathway carry both luminance and chromatic information. Most P cells are chromatically opponent across space, with their centers dominated by one range of wavelengths and their surrounds by another. In comparison, the M cells carry primarily luminance information. Most M cells are not chromatically opponent across space; their centers and surrounds are dominated by the same range of wavelengths. Recent evidence suggests that the koniocellular pathway plays an important role in color vision (Dacey, 1996; De Valois & De Valois, Chapter 4, this volume; Rodieck, 1998).

B. Performance

All of the visual information extracted by the eye is represented in the responses of the retinal ganglion cells, which number approximately one million in each eye. However, this same information is represented in the responses of 200 to 500 million neurons at the level of the primary visual cortex. Thus, a reasonable hypothesis would be that much of the information contained in the ganglion cell responses is reliably represented in the primary visual cortex, and hence that much of the information lost during visual processing is lost in those anatomical and physiological mechanisms lying between the cornea and the optic nerve. In other words, it would seem plausible that optical and retinal mechanisms are the major factors limiting visual performance in many tasks. The evidence suggests that, indeed, optical and retinal mechanisms are dominant for some basic visual tasks, such as pattern detection against uniform backgrounds. However, the evidence also suggests that cortical mechanisms become major limiting factors in other basic visual tasks, such as pattern discrimination and identification.

As described earlier, a Bayesian ideal-observer model can serve as a useful linking hypothesis for evaluating how anatomical and physiological mechanisms (as well as the physical properties of the stimulus) limit behavioral performance (e.g.,
An ideal observer analysis consists of three components: (a) a precise physical description of the stimuli; (b) a description of how the stimuli are modified by the relevant anatomical and physiological mechanisms; and (c) an optimal decision rule for the specific task under consideration. Such an ideal observer specifies the best possible performance that could be obtained given the stimuli, the task, and the mechanisms included in the analysis. The difference in performance between the human and ideal observers is a precise measure of the information lost in the nervous system following the anatomical and physiological mechanisms included in the ideal-observer analysis. The degree to which human performance is predicted by the ideal observer shows the degree to which overall visual performance is limited by the physical properties of the stimuli and by the anatomical and physiological mechanisms included in the ideal observer.

Ideal-observer analyses have suggested that optical and retinal factors are largely responsible for limiting the detectability of sine wave stimuli along the dimensions of spatial frequency, eccentricity, and wavelength (Arnow & Geisler, 1996; Banks et al., 1987; Banks et al., 1991; Sekiguchi et al., 1993). For example, Arnow and Geisler (1996) compared the detection performance of an ideal observer, which incorporated optical and retinal factors, to the detection performance of humans, for the dimensions of spatial frequency, eccentricity, and target size. The optics were represented using the point-spread function reported by Campbell and Gubisch (1966) for a 3-mm pupil. The retinal factors were represented by current estimates of the properties of retinal ganglion cells. The sampling density of the ganglion cells was taken from the measurements of Curcio and Allen (1990) for the adult human retina. Based upon electrophysiological studies in the monkey (Croner & Kaplan, 1995; Derrington & Lennie, 1984), the center diameter of the receptive field was made equal to the spacing between the ganglion cells; the surround diameter was six times the diameter of the center; and the surround strength was 75% of the center strength. Only the P cells were included in the analysis because of evidence from lesion studies that M cells contribute little to sine wave grating thresholds in monkeys except for low spatial frequencies presented at high temporal frequencies (Merigan & Maunsell, 1993).4

The symbols in Figure 6 (connected by dashed lines) show the contrast threshold data reported by Robson and Graham (1981), and the solid curves show the performance of the ideal observer. The symbols connected by dashed lines plot contrast sensitivity (1/contrast threshold) as a function of retinal eccentricity for five different spatial frequencies. The grating target width decreased with spatial frequency such that the target contained four cycles, and the length was set equal to

---

4 Two relevant properties of the neural responses in the cortex were also included in the analysis. First, near detection threshold cortical cell responses grow with contrast according to a power function with an exponent that averages 2.5 in the monkey. Thus, an exponent of 2.5 was applied to the final output. Second, the variance of cortical cell responses increases in proportion to the mean response with a proportionality constant of 1.5 (the average for neurons in primary visual cortex). Thus, the variance of the final output was made equal to the mean times 1.5.
FIGURE 6 Contrast sensitivity as a function of eccentricity and spatial frequency for a briefly presented target with a width and height of four cycles. The symbols connected by dashed lines show the measured contrast sensitivity for a single observer (JR). The solid curves show the ideal-observer predictions incorporating optical and retinal factors. (Data from Robson & Graham, 1981.)

The width. The symbols in Figure 7 plot contrast sensitivity as a function of target size for three different spatial frequencies; the left panel is for targets located in the fovea, and the right panel is for targets located in the periphery. The solid curves show the predictions of the ideal observer.

FIGURE 7 Contrast sensitivity as a function of target size and spatial frequency for briefly presented targets in the fovea and in the periphery (42 periods from the fovea). The symbols show the measured contrast sensitivity for a single observer. The solid curves show the ideal-observer predictions incorporating optical and retinal factors. (Data from Robson & Graham, 1981.)
As noted above, the performance of the ideal observer was based upon current anatomical and physiological measurements with only one free scaling parameter, and thus the close correspondence between the human and ideal performance suggests that the visibility window for the dimensions of spatial frequency and eccentricity is largely determined by optical and retinal factors.

VI. PRIMARY VISUAL CORTEX

The output neurons of the LGN project to the primary visual cortex. The number of neurons in the primary visual cortex is on the order of 200 to 500 times greater than the number of neurons projecting from the LGN. Thus, it would be possible (according to sampling theory) to have 200 to 500 complete, yet different, independent representations of all of the information contained in the LGN output. In spite of this incredible potential for representing all of the information, some information is lost due to various nonlinear mechanisms, such as response expansion and contrast gain control. Although these nonlinear mechanisms limit suprathreshold discrimination performance, they nevertheless enhance and maintain the stimulus selectivity of cortical cells. We will argue that these nonlinearities play an essential role in higher level processes, such as object recognition.

In addition to the nonlinearities, it is crucial to consider the variability of the neural responses because noise characteristics also play an important role in determining performance.

A. Selectivity

Single neuron electrophysiology has shown that neurons in the primary visual cortex generally respond to a narrower range of stimuli than neurons in the retina and LGN (De Valois, Albrecht, & Thorell, 1982; De Valois, Yund, & Hepler, 1982; Hubel & Wiesel, 1962, 1968; Movshon, Thompson, & Tolhurst, 1978a,b,c). Among other things, it has been shown that they are often simultaneously selective along the dimensions of spatial position, orientation, spatial frequency, contrast, temporal frequency, direction of motion, and color. Selectivity along these fundamental dimensions makes it possible for the visual cortex to represent the position, size, and orientation of the local image structure produced by surfaces and objects in the environment. Because most cortical neurons are simultaneously selective along many dimensions, a typical cortical neuron is generally much less active than a neuron in the retina or the LGN. For example, during normal saccadic inspection of a

5 The P, M, and K cell projections terminate in separate target regions; however, the clear segregation maintained in the retina, the LGN, and these cortical regions is more difficult to discern in the subsequent neural circuitry (for a review see, Callaway, 1998). Nonetheless, there is some evidence for functionally different pathways in the primary visual cortex and in extrastriate visual cortical areas that are linked to these pathways originating in the retina and LGN (Felleman & Van Essen, 1991; Merigan & Maunsell, 1993; Van Essen & DeYoe, 1994).
complex natural scene, a large fraction of the ganglion cells and LGN cells would respond during each fixation because of their broad tuning in space and time, whereas only a very small fraction of cortical cells would respond during a single fixation.

The selectivity of neurons in the visual pathway has often been characterized in terms of linear mechanisms which perform simple addition and subtraction of the light falling on the receptive field (Enroth-Cugell & Robson, 1966; Hartline, 1974; Rodieck, 1965). The receptive field map indicates the magnitude of the excitatory and the inhibitory responses of the cell to a spot of light placed at each location within the receptive field. To the extent that the neuron is linear, it is then possible to predict the neuron’s response to arbitrary spatial patterns of light from this map. This is because an arbitrary spatial pattern can be decomposed into small spots of light and then the total response of the neuron can be found by adding up the responses to each of the small spots according to the weights given by the receptive field map. These linear models have been able to account quantitatively for many of the response properties of visual neurons.

However, there are a number of response properties of neurons in the primary visual cortex that can only be explained by nonlinear mechanisms. For example, it is well established that the neurons which Hubel and Wiesel classified as “complex cells” cannot be described by simple addition and subtraction of the light falling on the receptive field (Skottun et al., 1991). For these cells, it is not possible to predict the responses to arbitrary spatial patterns from the receptive field map. In general, complex cells show excitatory responses to small spots of light everywhere within the receptive field, and thus one would predict that the best stimulus would be a uniform distribution of light covering the receptive field. However, this pattern produces little or no response. The optimal stimulus is typically an oriented bar that is much narrower than the width of the receptive field map; further, the bar will generally produce a large response no matter where it is positioned within the receptive field.

In comparison to complex cells, the neurons classified by Hubel and Wiesel as “simple cells” are better characterized by linear mechanisms, in that it is possible to qualitatively predict the optimal stimulus based upon the receptive field map. Nevertheless, simple cells display several fundamental nonlinear response properties. Further, these nonlinear response properties are also observed in complex cells. First, the responses of cortical neurons as a function of stimulus contrast are nonlinear (Albrecht & Hamilton, 1982): at low contrasts the responses increase with a power function exponent greater than one, and at high contrasts the responses approach an asymptotic maximum (that is, the responses saturate at a maximum value). Second, the response saturation generally occurs at the same contrast for all stimuli (both optimal and nonoptimal); as a consequence, the asymptotic response maximum is different for optimal and nonoptimal stimuli, even though the power function exponent remains the same (Albrecht & Hamilton, 1982). Third, most cortical cells have little or no maintained spontaneous discharge and thus the effects of inhibitory stimuli cannot be directly observed (i.e., the responses are rectified).
These nonlinear properties have a profound effect in determining the selectivity of simple and complex cortical cells.

1. Contrast

As noted above, a specific reflectance difference in the natural environment produces a constant luminance contrast independent of the level of the illumination. Thus, the crucial information that must be encoded is the spatial distribution of local contrast. Cortical neurons are selective to local contrast; this selectivity is the culmination of a process begun in the retina.

a. Nonlinearities

The response of a typical cortical cell as function of stimulus contrast is plotted in linear coordinates in Figure 8A and in logarithmic coordinates in Figure 8B. As the stimulus contrast increases, the response increases rapidly in the range of 0 to 10% and then reaches a maximum saturated response beyond 15%. The contrast at which the response reaches half of the maximum value, the "half-saturation contrast," is approximately 7% for this cell. Across the population of cells as a whole, the half-saturation contrast varies over the full range of possible values, with some cells saturating at a contrast as low as 5% and others saturating beyond 50%, with a median value of approximately 30%.

Figure 9A shows the responses of a cortical neuron as a function of stimulus contrast measured for five different spatial frequencies. Note that the responses saturate at different levels depending upon the spatial frequency of the stimulus. For example, the response saturates at 10 spikes/second for the optimal spatial frequency of 4.0 cycles/degree but saturates at 3 spikes/second for the nonoptimal spatial frequency of 1.5 cycles/degree. On the other hand, the half-saturation contrast remains approximately constant for the different spatial frequencies. These response
properties are not what would be expected from simple response saturation or compression of the type often found in sensory neurons and routinely proposed in psychophysical models. Simple response saturation often reflects some neurophysiological limit, such as the maximum level of depolarization, or the maximum spike rate. For example, the response saturation observed in photoreceptors is quite different from that observed in cortical cells. Figure 9B illustrates the responses of a photoreceptor as a function of intensity for several different wavelengths of light. Note that, unlike cortical cells, the responses saturate at the same level for all of the different wavelengths and that the half-saturation intensity changes with wavelength.

The nonlinear saturation observed in the photoreceptors is determined by a limit in the maximum response (a hyperpolarization limit); this type of saturation, "response saturation," has negative consequences for color selectivity, as well as color identification and discrimination. At high intensities, nonoptimal wavelengths (those away from the peak of the spectral-sensitivity function) produce the same maximum response as near optimal wavelengths. In comparison, the nonlinear saturation observed in cortical neurons is not determined by a limit in the maximum response rate, but instead is determined by the local contrast. This type of saturation, "contrast saturation," has a number of beneficial consequences for stimulus selectivity, as well as for identification and discrimination performance, because only near optimal stimuli can produce the maximum response. With contrast saturation, the selectivities along all the various stimulus dimensions remain invariant with contrast.

Interestingly, note that the responses shown in Figures 8 and 9A do not increase in a linear fashion even at low contrasts, but instead increase in a nonlinear accelerating fashion. As illustrated by the dashed line in Figure 8B, a linear relationship appears as a straight line with a slope of 1.0, when plotted in log-log coordinates (a power function with an exponent, or power, equal to 1.0). However, the measured responses follow a line with a slope closer to 3.0 (a power function with an exponent equal to 3.0). Across cells, the slope, or power function exponent, varies over...
a wide range, from less than 1.0 to greater than 5.0, with a median value of 2.5 (Albrecht & Hamilton, 1982; Geisler & Albrecht, 1997; Sclar, Maunsell, & Lennie, 1990). This accelerating nonlinearity enhances stimulus selectivity and eases the structural requirements for producing high degrees of selectivity. For example, high degrees of orientation tuning can be obtained with relatively short receptive fields. This accelerating nonlinearity also has beneficial consequences for discrimination and identification performance because responses to near optimal stimuli are enhanced relative to the responses to nonoptimal stimuli. 

In sum, measurements of the contrast response function at different spatial frequencies (and other stimulus dimensions) have revealed two important nonlinear characteristics of cortical neurons: a saturation controlled by the local contrast that manages to preserve stimulus selectivity, and an accelerating response (revealed at low contrasts) which enhances stimulus selectivity. There is considerable evidence that the saturation is produced by a contrast normalization network (Albrecht & Geisler, 1991, 1994; Albrecht & Hamilton, 1982; Bonds, 1991; Carandini & Heeger, 1994; Geisler & Albrecht, 1997; Heeger, 1992b; Li & Creutzfeldt, 1984; Robson, 1991; Sclar & Freeman, 1982; Tolhurst & Heeger, 1997) and that the accelerating response is due to a final expansive response exponent (Albrecht & Geisler, 1991, 1994; Albrecht & Hamilton, 1982; DeAngelis, Ohzawa, & Freeman, 1993; Geisler & Albrecht, 1997; Heeger, 1992a; McLean & Palmer, 1994; Sclar et al., 1990).

b. Variability

Visual performance is limited by the amount of noise (random variability) in the neural responses. Unlike the retina and LGN, the variability of cortical neuron responses increases substantially as the magnitude of the responses increases. Figure 10A illustrates the change in both the mean and the variability of the response as a function of contrast for a typical cortical neuron; the measurements are shown for an optimal and a nonoptimal spatial frequency. Note that as the response increases the variability increases in a parallel fashion. Further, the magnitude of the variability is not determined by the magnitude of the contrast, per se, but rather by the magnitude of the response; thus, the variability is smaller for a nonoptimal stimulus, even at very high contrasts. Systematic measurements from many different laboratories are consistent with the hypothesis (represented by the solid curves in Figure 10A) that the variance of the response is directly proportional to the mean of the response (Geisler & Albrecht, 1995, 1997; Snowden, Treue, & Andersen, 1992; Softky & Koch, 1993; Tolhurst, Movshon, & Dean, 1983; Vogels, Spileers, & Orban, 1989).

2. Spatial Frequency

The selectivity of cortical cells to stimulus size has been quantitatively investigated by varying the spatial frequency of sinusoidal grating stimuli (for general reviews

---

6 Current evidence suggests that this accelerating power function exponent could be a consequence of a nonlinear relationship between the intracellular voltage and the number of action potentials (see Figure 1, McCormick, Connors, Lighthall, & Prince, 1985).
Measured Responses & Descriptive Functions

FIGURE 10  Response means and standard deviations of representative neurons recorded from the primate visual cortex, along six stimulus dimensions for stimulus durations of 200 ms. (Taken from Geisler and Albrecht, 1997.)

see De Valois & De Valois, 1990; Palmer, Jones, & Stepnoski, 1991). Figure 10C shows the mean and the variability of the response of a typical cell as a function of spatial frequency; the measurements are shown for a low contrast and a saturating contrast. This neuron responds best to a spatial frequency of 4.0 cpd. The optimal (critical) spatial frequency varies widely across neurons, even for those neurons sampling the same location in visual space; the physiological range of critical frequencies is roughly consistent with the behavioral range of detectable frequencies.

As is typical of cortical cells, the neuron shown in Figure 10C responds to only a limited range of spatial frequencies (approximately 2.5–5.0 cpd at half-height). This range of spatial frequencies corresponds to a bandwidth of 1.2 octaves. There is considerable heterogeneity in the bandwidth across cells, from less than 0.7 octaves to greater than 3.0 octaves. The average bandwidth reported in the literature is approximately 1.5 octaves. This average bandwidth is smaller than the average bandwidth of ganglion cells or LGN cells (3 to 5 octaves) and corresponds to only a small
portion of the overall behavioral range (for a general review see Shapley & Lennie, 1985).

The spatial frequency tuning of cortical cells remains relatively invariant when measured at different contrasts. Remarkably, even at saturating contrasts the critical frequency and the bandwidth do not change. For example, the solid curves in the upper panel of Figure 10C have the same critical frequency and bandwidth; they only differ by a scale factor. As described above (see section VI.A.1.a), this contrast-invariant tuning is consistent with a saturation that is determined by the local contrast (a contrast normalization mechanism).

The spatial response properties of cortical cells have been characterized by measuring spatial frequency tuning functions and receptive field maps. If the spatial response properties were the result of simple summation of excitation and inhibition then the two characterizations would be equivalent (one would be the Fourier transform of the other). Cells with broad spatial frequency tuning would be expected to have receptive field maps with only a few excitatory and inhibitory regions, whereas cells with narrow tuning would be expected to have maps with many excitatory and inhibitory regions. Although this is approximately true for some cells, for many cells the receptive field maps have fewer excitatory and inhibitory regions than would be expected from their spatial frequency tuning.

This mismatch could potentially be accounted for by the expansive nonlinearity revealed in measurements of the contrast response function (see section VI.A.1.a). An expansive nonlinearity simultaneously narrows the spatial frequency tuning and reduces the number of excitatory and inhibitory regions in the receptive field map. The solid curves in Figure 11 show the receptive field map (A) and the spatial frequency tuning (B) of a hypothetical linear cortical neuron which performs simple

![FIGURE 11](image-url)

FIGURE 11 The effect of an expansive response exponent on the receptive field map and the spatial frequency tuning function. The solid curves show the receptive field map (A) and the corresponding spatial frequency tuning function expected given a linear receptive field (B). The dotted curves show the effect of an expansive response exponent of 2.5 on the receptive field map and tuning function. The dashed curve in A shows the receptive field predicted (dotted curve) in B under the assumption of linearity. The difference between the dashed and dotted curves in A shows the expected mismatch between the receptive field map measured in the space domain and the one predicted from measurements of the tuning function in the frequency domain.
summation of excitation and inhibition. The dotted curves show the receptive field map and the spatial frequency tuning function that would be measured given an expansive exponent of 2.5. If the spatial frequency tuning indicated by the dotted curve in 11B were measured in an experiment, then the receptive field map, predicted given linear summation of excitation and inhibition, would be the dashed curve in 11A. This figure demonstrates that because of the expansive nonlinearity, high degrees of spatial frequency selectivity can be obtained with a structurally simple receptive field.

Note that the variability of the response, as a function of spatial frequency, for the cell shown in 10C, mirrors the mean of the response. This is a consequence of the fact that the variance increases in direct proportion to the mean; the solid curves through the mean and variance data were generated under the hypothesis that the variance is proportional to the mean.

3. Orientation

Figure 10D shows the mean and the variability of the response of a typical cell as a function of orientation; the measurements are shown for two directions of motion. As Hubel and Wiesel (1962, 1968) first demonstrated, the optimal (critical) orientation differs from cell to cell, and is distributed across the full range (De Valois, Albrecht, & Thorell, 1982). The orientation bandwidth for this cell is 22°. The orientation bandwidth also differs from cell to cell; the average bandwidth is 40°. Like spatial frequency tuning, orientation tuning remains relatively invariant with contrast; this is true even at contrasts which produce response saturation (Sclar & Freeman, 1982). The degree of orientation tuning is determined by the elongation of the receptive field; the greater the length, the smaller the orientation bandwidth. However, for many cells, the receptive field maps are shorter than what would be expected from their orientation tuning. Once again, this mismatch could potentially be accounted for by an expansive nonlinearity (see section VI.A.1.a), which narrows the measured orientation tuning without requiring a longer receptive field. The solid curves through the mean and variance data were generated under the hypothesis that the variance is proportional to the mean.

4. Position

The selectivity of cortical cells to spatial position has been quantitatively investigated by varying the phase, or position, of spatial frequency stimuli (De Valois, Albrecht, & Thorell, 1982; Movshon et al., 1978c). Figure 10B shows the mean and the variability of the response of a typical cell as a function of the spatial position of a grating pattern that was turned on and off, reversing in contrast. The optimal position differs from cell to cell, such that the population of cells as a whole covers the entire visual field. The width of the position tuning function at half-height, for the cell shown in Figure 10B, is approximately 4 min of arc. The width varies depending upon the cell's receptive field as well as the spatial frequency of the
stimulus. Like spatial frequency and orientation, the spatial position tuning remains invariant with contrast (Albrecht & Geisler, 1991). If the spatial response properties were the result of simple summation of excitation and inhibition, then the spatial position tuning functions should be exactly sinusoidal in shape (Enroth-Cugell & Robson, 1966; Hochstein & Shapley, 1976). Although the functions appear approximately sinusoidal, they are in fact narrower than would be expected, and are more similar to a sine wave taken to an exponent greater than one. Again, this could potentially be accounted for by an expansive nonlinearity and is, in fact, converging evidence for the existence of such a nonlinearity. Comparable to other stimulus dimensions, the effect of the expansive nonlinearity is to enhance selectivity, in this case position selectivity. The solid curves through the mean and variance data were generated under the hypothesis that the variance is proportional to the mean.

5. Temporal Frequency

The temporal integration properties of cortical cells have been quantitatively investigated by varying the temporal frequency of sine wave grating patterns (Foster, Gaska, Nagler, & Pollen, 1985; Hamilton, Albrecht, & Geisler, 1989; Hawken, Shapley, & Grosof, 1996; Movshon et al., 1978b). Figure 10E shows the mean and the variability of the response of a typical cell as a function of the temporal frequency of a drifting sine wave grating with the optimal spatial frequency and orientation; the measurements are shown for a low contrast and a saturating contrast. This neuron responds best to a temporal frequency of 8.0 Hz. The optimal (critical) temporal frequency varies across neurons. The physiological range of temporal frequency tuning measured in the cortex is roughly consistent with the behavioral range of detectable frequencies. In general, temporal frequency tuning is relatively broad in comparison to spatial frequency tuning, with nearly half of the cells showing little or no attenuation to frequencies as low as 0.25 Hz. For those cells that do show low-frequency attenuation, the average bandwidth is approximately 3.0 octaves at half-height. The bandwidth for the cell in Figure 10E is 3.1 octaves. The solid curves through the mean and variance data were generated under the hypothesis that the variance is proportional to the mean.7

6. Direction of Motion

Many of the cells in the primary visual cortex are selective for the direction of stimulus motion (Hubel & Wiesel, 1962, 1968). M and P cells in the retina and LGN

7 Unlike the other dimensions, the tuning function for temporal frequency changes shape somewhat with contrast. As contrast increases, the critical frequency increases, and the temporal phase of the response decreases (Albrecht, 1995; Carandini & Heeger, 1994; Reid, Victor, & Shapley, 1992). Similar effects are observed in the retina (Shapley & Victor, 1978) and LGN (Sclar, 1987). These effects are consistent with what could be expected from a dynamic contrast normalization mechanism. Also, unlike the other dimensions, the variance proportionality constant is affected by temporal frequency: the value of the constant increases at lower temporal frequencies (see Appendix B, Geisler & Albrecht, 1997).
are not direction selective; the visual cortex is the first level in which direction-selective neurons are found. Figure 10F shows the responses of a representative single neuron as a function of contrast for a sine wave grating moving in the optimal and the nonoptimal directions. In the optimal direction the average maximum response was 7.5 spikes (in 200 ms); in the nonoptimal direction the average maximum was 2.5 spikes. The response in the preferred direction is approximately 2.5 times larger than the response in the nonpreferred direction, and this ratio remains the same at all contrasts, even in the saturated region. The solid curves through the mean and the variance data were generated under the hypothesis that the variance is proportional to the mean.

The degree of direction selectivity varies from cell to cell. For approximately 30% of simple cells and complex cells the magnitude of the response in the preferred direction is more than twice the magnitude of the response in the nonpreferred direction (De Valois, Yund, & Hepler, 1982). The average direction selectivity for the population as a whole (one minus the ratio of the nonpreferred to preferred response) is slightly less than 0.5.

Barlow and Levick (1965) demonstrated that direction selectivity could be produced by neural summation of inputs that are displaced in space and delayed in time. Others have described direction selectivity in terms of “spatiotemporal” receptive fields, where one axis represents space and the other axis represents time (e.g., Adelson & Bergen, 1985; Watson & Ahumada, 1985). They demonstrated that direction selectivity can be obtained through strictly linear summation of excitation and inhibition over the spatiotemporal receptive field. The degree of direction selectivity is determined by the degree to which the spatiotemporal receptive field has a single dominant orientation off the temporal axis. The dominant orientation determines the preferred direction of motion.

It has been shown that the spatiotemporal orientation of the receptive fields measured in the visual cortex correctly predicts the preferred direction of motion. However, there is a mismatch between the spatiotemporal receptive field and the degree of direction selectivity, which is greatly reduced when the response exponent is taken into consideration (DeAngelis et al., 1993; McLean & Palmer, 1989; McLean, Raab, & Palmer, 1994). Although high degrees of direction selectivity require rather precise wiring in space and time, cortical cells apparently achieve high degrees of direction selectivity by combining a less precise receptive field structure with an expansive response exponent (Albrecht & Geisler, 1991, 1994).

7. Surround Suppression and Facilitation

The responses of cortical neurons are often affected by stimuli which fall outside of the classical receptive field (e.g., De Valois, Thorell, & Albrecht, 1985; DeAngelis, Freeman, & Ohzawa, 1994; Gilbert & Wiesel, 1990; Levitt & Lund, 1997; Li & Li, 1994; Sengpiel, Baddeley, Freeman, Harrad, & Blakemore, 1998; Sillito, Grieve, Jones, Cudiero, & Davis, 1995; Zipser, Lamme, & Schiller, 1996). The classical
receptive field is defined operationally; typically, a small patch of grating (of the optimal spatial frequency, orientation, and direction of motion) is centered on the receptive field and then expanded in size until the response of the neuron increases no further. Stimuli presented outside of the classical receptive field produce little or no response when presented alone. However, when presented in conjunction with stimuli inside the classical receptive field, surround stimulation can sometimes suppress or facilitate responses. The magnitude and sign of the effect vary from cell to cell and depend upon a variety of different stimulus factors (e.g., contrast, orientation, spatial frequency). In general, the magnitude of the reduction in response that is produced by the suppressive effect varies across cells from 0 to 100%, with an average value of approximately 30%. Facilitatory effects have been observed less frequently and are generally smaller in magnitude. There have been many different suggestions for the role of these surround effects in cortical functioning (e.g., gain control, texture segregation, etc.). Nonetheless, at this point in time, it appears as though the responses of V1 cortical cells are dominated by the effects of stimulation within the classical receptive field and modulated by the effects of stimulation within the surround.

8. Continuous and Heterogeneous Distribution of Properties

The neurons illustrated in Figure 10 are representative of neurons in the primate visual cortex. However, there is a great deal of heterogeneity from cell to cell: the cells vary widely in their particular preferences along all of the stimulus dimensions. For example, each cell has a different preferred orientation, spatial frequency, temporal frequency, direction of motion, and spatial phase. The bandwidths of the cells also vary along all of the stimulus dimensions. In addition, there is a great deal of heterogeneity from cell to cell in the contrast response function, the maximum firing rate, the spontaneous firing rate, the noise characteristics, the color tuning, the AC/DC index (for the simple–complex dimension) and so on.

Throughout the history of vision science there has been a tendency to organize the complexity of the visual system by segregating and classifying neural mechanisms into a small manageable number of discrete types: four types of photoreceptors, three types of opponent mechanisms (two color and one luminance), two spatial phase pathways (on and off pathways), P and M pathways, simple and complex cells, discrete spatial frequency tunnel channels, sustained and transient temporal channels, direction-selective and nondirection-selective cells, and so on. Although such simple taxonomies may be appropriate for classifying the various mechanisms in the retina and LGN, they seem less appropriate at the level of the visual cortex, where the properties of the neurons are heterogeneous and distributed continuously.

However, even though the properties of cortical neurons are heterogeneous and distributed continuously, the neurons are very regular in a certain way. For example, their tuning functions and contrast response functions generally have a
characteristic shape. Thus, in general, it is possible to summarize the stimulus–
response relationships of each cortical cell using a small set of relatively simple
descriptive functions, where each descriptive function has only a few free param-
eters. For example, Geisler and Albrecht (1997) found that a set of simple descriptive
functions could account for approximately 90% of the variation in both the
means and standard deviations of cortical cell responses (across a large population
of neurons in both monkey and cat), for the dimensions of contrast, spatial fre-
quency, spatial position, temporal frequency, orientation, and direction of motion.
This suggests that a better way to organize the complexity of cortical neurons might
be based upon the frequency distributions of the various cell properties.

9. Contrast–Gain Exponent Model

The response properties of cortical cells described in the sections above can be sum-
marized with the model illustrated in Figure 12 (see also, Albrecht & Geisler, 1991;
Geisler & Albrecht, 1997; Heeger, 1991, 1992a,b). The model consists of four com-
ponents: (a) a linear filter, which establishes the neuron’s stimulus selectivity, (b) a
contrast normalization mechanism, which maintains selectivity in spite of response
saturation, (c) an expansive nonlinearity, which enhances selectivity, and (d) a noise
source, which makes the variance proportional to the mean. This figure shows the
hypothesized processing, as a function of contrast, for two different sinusoidal stim-
uli; one that is optimal for the cell (the solid curves) and one that is nonoptimal (the
dashed curves). The nonoptimal stimulus could be nonoptimal along any dimen-
sion (e.g., spatial frequency, orientation, position, direction of motion, etc.). The
four upper boxes represent the four major stages of the model. The five lower boxes
illustrate contrast response functions measured before and after each stage.

Measured at the input, the amplitude of the sinusoidal stimulus increases in
proportion to contrast, by definition. If amplitude and contrast are plotted on log–
log coordinates, as they are here, the slope is 1.0. Next, the contrast normalization mechanism (contrast gain) scales the input amplitude by a factor that decreases with increasing contrast. This is a fast-acting gain control which is set by the local spatiotemporal contrast. Thus, the gain does not depend upon the response rate of the cell. As a consequence, both the optimal and nonoptimal stimulus are attenuated equally. According to this model, response saturation occurs because at higher contrasts the decrease in gain cancels the increase in input amplitude.

Next, the gain-adjusted signals are passed through a linear filter (summation of excitation and inhibition), which gives the cell its fundamental selectivities. The response to the nonoptimal stimulus, Stim 2, is attenuated more than that to the optimal stimulus, Stim 1. Next, the response exponent takes the output of the linear filter, or filters (after half-wave rectification) to an exponent greater than 1.0. Note that the exponent does not eliminate response saturation, but it does increase the stimulus selectivity of the neuron (notice the bigger difference in response to optimal vs. nonoptimal stimuli). Finally, a multiplicative noise source causes the variance of the response to be proportional to the mean of the response. It is important to emphasize that this model is meant to provide a functional description of single cortical neuron responses and should not be taken as a hypothesis about the detailed anatomy and physiology. For example, contrast normalization is probably occurring at many levels, starting in the retina (Albrecht, 1995; Albrecht & Geisler, 1991; Sclar, 1987; Shapley & Victor, 1978). This model is consistent with the descriptive functions which have been used to summarize the responses of cortical neurons and hence it can account for 90% of the variability in both the means and standard deviations of cortical cell responses across a wide array of stimulus dimensions.

B. Performance

There is considerable psychophysical evidence for multiple pathways or channels in the visual system that are selective for the stimulus dimensions of size (spatial frequency), orientation, direction of motion, and binocular disparity (for reviews see Braddick, Campbell, & Atkinson, 1978; De Valois & De Valois, 1990; Graham, 1989; Regan, 1991; Sekuler, Pantele, & Levinson, 1978; Wilson, Levi, Maffei, Rovamo, & De Valois, 1990). Much of this evidence is based upon masking and adaptation studies, which demonstrate selective threshold elevations for stimuli that are similar to the masker or adapting pattern. For example, adapting to a specific spatial frequency raises the threshold for detecting that particular spatial frequency and similar spatial frequencies, but it has little effect on the detection of dissimilar spatial frequencies. Neurophysiological evidence has demonstrated that the primary visual cortex is the first level in the visual pathway where neurons exhibit high degrees of selectivity along these stimulus dimensions (see section VI.A). Thus, the response properties of neurons in the primary visual cortex are particularly important to consider in trying to understand the relationship between neural mechanisms and visual performance, which involve these stimulus dimensions.
To compare the performance of single neurons to behavior it is necessary to measure the responses of individual neurons along the same stimulus dimensions that are used to measure behavioral performance. The performance of the neuron can then be assessed by applying an optimal decision rule to the measured responses (which is equivalent to assuming that subsequent brain mechanisms are able to make use of all of the information in the responses). Using this approach it has been demonstrated (a) that the sensitivity of single neurons sometimes approaches that of the organism, (b) that the shape of the neural performance function sometimes resembles that of the organism, but (c) that no single neuron matches the performance along the entire stimulus range to which the organism is sensitive (see for example, Barlow, Kaushal, Hawken, & Parker, 1987; Bradley, Skottun, Ohzawa, Sclar, & Freeman, 1985; Geisler & Albrecht, 1997; Geisler, Albrecht, Salvi, & Saunders, 1991; Hawken & Parker, 1990; Newsome, Britten, & Movshon, 1989; Parker & Newsome, 1998; Tolhurst et al., 1983).

The standard method of measuring single-neuron performance is to essentially run a psychophysical experiment on the individual neuron. However, this method is quite time consuming, and thus it greatly restricts both the number of neurons that can be measured and the number of stimulus conditions that can be run on an individual neuron. For example, discrimination performance is often measured using a two-alternative forced choice task, where a minimum of 100 stimulus presentations is required for each point on the discrimination function. Thus, only a handful of discrimination points can be measured during the typical time available for recording from a given neuron.

An alternative method, the descriptive-function method, is to measure the response of the neuron (both the mean and the variance) along the stimulus dimension of interest and then summarize the responses as a whole with a descriptive function. With this method, entire performance functions can be measured along several different stimulus dimensions during the time available for recording from a given neuron (Geisler & Albrecht, 1995, 1997). Furthermore, because the descriptive function method makes it practical to measure the performance of large populations of individual neurons, it becomes possible to estimate the performance of a population of cells as a whole and compare this population performance to behavioral performance. Control experiments have shown that the discrimination performance obtained using the descriptive function method is equivalent to the discrimination performance obtained using the standard method (see Appendix D, Geisler & Albrecht, 1997).

When making comparisons of neural and behavioral performance it is important to take into consideration the fact that behavioral discriminations typically occur within the time frame of a single fixation (approximately 200 ms), or less: psychophysical estimates of temporal integration intervals typically range from 50 to 200 ms. Thus, it is important to measure neural performance using comparable stimulus durations.
In what follows, we first describe the detection and discrimination performance of single neurons for brief stimulus presentations. Next, we describe the detection and discrimination performance of the primary visual cortex as a whole based upon measurements from large populations of single neurons. Finally, we consider the relationship of the population performance of cortical neurons to psychophysical models of performance.

1. Single Neuron Performance
   
   a. Discrimination Performance

   In order to encode surface reflectances the visual system must be able to discriminate contrasts. Figure 13A shows the contrast discrimination performance (contrast threshold as a function of the background, or base, contrast) for a typical neuron in the primary visual cortex. This discrimination function was obtained by applying an optimal decision rule (from signal-detection theory) to the descriptive functions that were fitted to the responses in Figure 10A. The trial duration was 200 ms. Note that the neuron can only discriminate contrasts over a narrow range of background contrasts (0–12%); the absolute detection threshold was 7% contrast; the threshold decreased to a minimum value of 4% and increased thereafter. Similar to behavioral

   ![Discrimination Performance: Representative Cells](image)

   **FIGURE 13** Discrimination performance of single neurons in monkey visual cortex.
contrast discrimination functions, this cell shows a "dipper effect" (the thresholds initially decrease as a function of contrast). However, behavioral contrast discrimination functions extend across the full range of contrasts (0–100%). Most neurons in the primary visual cortex show the dipper effect and can only discriminate over a limited range of background contrasts; those cells that can discriminate over a wide range are very insensitive (see for example, Barlow et al., 1987; Bradley et al., 1985; Geisler & Albrecht, 1997; Geisler et al., 1991).

In order to encode different shapes and sizes, the visual system must be able to discriminate differences in spatial frequency. Figure 13C shows the spatial frequency discrimination performance (frequency threshold as a function of the base frequency) for a typical neuron. This discrimination function was obtained by applying an optimal decision rule to the descriptive functions that were fitted to the responses in Figure 10C. Note that there are two minima on either side of the characteristic frequency (at 2.5 cpd and 6 cpd). The minima occur where small changes in frequency produce big changes in response; that is, where the slope of the selectivity function is steep. These discrimination functions were determined for both frequency increments (dashed curves) and decrements (solid curves); for ease of viewing, only the lower of the two thresholds at each frequency is plotted. For some cells, the minimum of the frequency discrimination function approaches that of behavioral performance. However, the shape of the behavioral discrimination function is very different; it is a straight line that corresponds to Weber's law.

In order to encode the location, shape, and motion of objects we must be able to discriminate spatial position, orientation, temporal frequency, and direction of motion. The discrimination performance of representative neurons along these stimulus dimensions are also shown in Figure 13.

In sum, although there is heterogeneity in the discrimination functions from cell to cell, there are general trends: (a) the neural discrimination functions often share some properties with behavioral discrimination functions; (b) for most neurons the discrimination functions only cover a fraction of the behavioral range; and (c) those few cells whose discrimination functions cover most of the behavioral range are very insensitive, relative to behavioral performance.

The discrimination performance illustrated in Figure 13 is for stimulus durations of 200 ms, a duration similar to the duration of fixations during saccadic inspection. If the stimulus duration were longer than 200 ms, one would expect performance to improve. However, longer stimulus durations would not be representative of the normal functioning of the visual system because the behavioral temporal integration intervals for similar discrimination tasks are less than 200 ms. As mentioned above, psychophysical estimates of temporal integration intervals typically range from 50 to 200 ms. If the stimulus duration were shorter than 200 ms, then one would expect performance to decrease. Interestingly, the discrimination performance of cortical neurons does not decrease dramatically when the stimulus duration decreases from 200 to 50 ms, because the variance proportionality constant decreases in the shorter intervals (Frazor, Albrecht, Geisler, & Crane, 1998).
Identification Performance

In order to perform object recognition it would be useful if neurons in the early visual system were able to identify local image features along various stimulus dimensions. This would permit subsequent brain mechanisms to begin the process of segregating objects from their context on the basis of local feature similarity. Although cortical cells are simultaneously selective along several different stimulus dimensions, they may not be capable of identifying local image features because they are rather broadly tuned, they have low firing rates, and they have high response variability. However, quantitative measurements of the identification performance indicates that individual cortical neurons can reliably signal the presence of specific local image features.

The identification task can be conceptualized as follows. One must identify which stimulus was presented on a trial, when the stimulus is randomly selected from a wide range of possibilities. Suppose as an observer (or as a subsequent brain mechanism), that the only information available on each trial is the total number of action potentials from a given neuron (plus some implicit knowledge of the neuron's tuning functions and variability). What could be known about the stimulus given a particular spike count during a single trial? This question can be answered using the descriptive functions for the mean and variance to determine a certainty function (an a posteriori probability density function), which is the probability of each possible stimulus given that particular spike count. From the certainty function it is possible to determine a 95% confidence region in stimulus space. If the 95% confidence region is small it means that the neuron's identification performance is good (i.e., the stimulus can be localized to a small region of stimulus space).

Figure 14 shows representative certainty functions for the dimensions of contrast, spatial position, spatial frequency, orientation, temporal frequency, and direction of motion, given that the stimulus is random and free to vary along the entire axis (i.e., a uniform prior probability distribution for the stimulus). These probability distributions were obtained from the descriptive functions shown in Figure 10, for the case where a maximum response has occurred in a 200-ms trial (e.g., a count of seven spikes for the dimension of contrast). The arrows along the stimulus dimensions indicate the 95% confidence region.

The certainty functions in Figure 14 are for single dimensions. In the natural environment, cortical neurons are confronted with simultaneous uncertainty along many stimulus dimensions. Interestingly, the certainty function along any given dimension is little affected by uncertainty along other stimulus dimensions. For example, the certainty functions for spatial frequency and orientation are unaffected even when the contrast is random and free to vary.

In general, the following statements are true:

1. Identification performance improves with increasing response, reaching fairly impressive levels at 5–10 spikes in a 200-ms interval. For these response levels, the width of the 95% confidence region along a particular stimulus dimension is
approximately equal to the half-height bandwidth of the cell's tuning along that dimension.

2. At low spike rates, identification performance becomes quite poor (the confidence regions become large and complex).

3. Because many cortical neurons saturate at low contrasts (i.e., between 5 and 20% contrast) without losing their tuning, good identification performance is often reached at very low contrasts.

The remarkable ability of a cortical neuron to respond uniquely to a relatively specific local image feature is a consequence of the linear filtering combined with the nonlinear response expansion and the contrast normalization. The linear filtering and the response expansion are essential for establishing the selectivity; however, the contrast normalization plays the crucial role of transforming cortical neurons so that they behave less like linear filters and more like feature detectors. Because of the special kind of saturation that is introduced by contrast normalization, a cortical cell will produce a large response to a very low contrast image feature of the correct size, orientation, and position, but will produce a weaker response to a high-contrast image feature that differs only slightly from the preferred stimulus along any dimension. This is very different from what would be expected from a linear filter.
2. Neural Population Discrimination Performance

Although the discrimination performance of individual neurons is similar to the behavioral discrimination performance in some respects, the performance of any single neuron is generally restricted to a much smaller stimulus range. Behavioral performance across the entire range of any stimulus dimension must involve the pooled activity of more than one neuron. Therefore, to understand behavioral performance, it is necessary to go beyond an analysis of the single neuron and to consider the performance of neural populations.

One method for evaluating the discrimination performance of neural populations is to combine the measurements from individual neurons. In order to do this, it is first of all necessary to measure the responses of a large representative sample of neurons along the stimulus dimensions of interest. (As shown earlier, the measured responses along the stimulus dimensions can be summarized using simple descriptive functions.) It is then necessary to combine the measured responses of each neuron using an appropriate pooling rule (i.e., a linking hypothesis). There are many possible pooling rules, and thus one sensible strategy to consider both "minimum pooling," where the performance at each stimulus level is determined by the most sensitive neuron, and "optimum pooling," where the performance at each stimulus level is determined by optimally combining the responses of all the neurons in the sample. These two pooling rules represent the end points of a continuum of possible pooling rules and hence set the boundaries of behavioral performance that might be expected based upon the responses of the neurons in the sample.

Figure 15 shows the population discrimination performance for a sample of neurons recorded from monkey cortex, along the stimulus dimensions of contrast and spatial frequency, using these two pooling rules. Consider first the minimum pooling rule. Each solid symbol in the upper panels shows the minimum of the discrimination function for one of the neurons in the population (see arrows in Figure 13A, 13C). As can be seen, for the dimension of contrast, the best performance points are widely scattered at each background contrast, and they span most of the contrast axis. All points that fall on the vertical axis at 0% contrast represent cells for which the best discrimination threshold was the absolute threshold; all other points represent cells for which the threshold initially decreased with contrast (i.e., they showed the dipper effect). It is interesting to note that as contrast increases, the thresholds of the most sensitive cells increase monotonically. Specifically, those cells with the lowest thresholds are bounded by an envelope that is approximately constant at lower contrasts and linear (in log-log coordinates) at higher contrasts. For the discrimination of spatial frequency, the best performance points are also widely scattered at each base frequency, and they span most of the frequency axis. As frequency increases, \( \Delta f \) initially decreases and then becomes approximately constant (Weber's law). The open symbols in Figure 15 plot behavioral discrimination functions for monkeys and humans; each function is for a different published study or
FIGURE 15 Comparison of neural population discrimination performance in monkeys with behavioral discrimination performance in monkeys and humans.

subject. Interestingly, the general shape of the behavioral functions is quite similar to the shape of the envelope of the most sensitive cells. However, the behavioral thresholds are somewhat lower.

Consider next the optimum pooling rule. The thick curves in the lower panels of Figure 15 show the discrimination functions that result from pooling all of the discrimination information from all of the cells in an optimal fashion. As can be seen, the shapes of the population discrimination functions are very similar to the shapes of the behavioral discrimination functions. These curves were obtained as follows. First, we used the descriptive functions for each neuron to find the mean responses and standard deviations to the base stimulus (contrast or frequency) and

8 Like the behavioral performance, the contrast discrimination functions for most individual cortical neurons show a dipper effect; however, the contrast discrimination function for the neural population does not. This is a consequence of the fact that the minima of the contrast discrimination functions for the individual neurons are widely scattered along the contrast axis.
to the base-plus-increment. Second, we used these means and standard deviations to obtain the signal-to-noise ratio, $d'$, for each cell. Third, the $d'$ summation formula from Bayesian decision theory was used to find the signal-to-noise ratio for the entire population. Fourth, the threshold was obtained by varying the increment until the population $d'$ equaled 1.0 (i.e., 75% correct discrimination). Finally, we shifted the population function vertically, using an efficiency parameter (which was the same for contrast and spatial frequency), to allow comparison of the shapes of the behavioral and neural population discrimination functions.

These two methods of comparing behavioral and neural performance represent two extremes in psychophysical linking hypotheses. At one extreme, behavioral performance is determined by assuming no pooling across neurons—the threshold for each stimulus condition is determined by the most sensitive cell (Barlow, 1972, 1995; De Valois, Abramov, & Mead, 1967; Talbot, Darian-Smith, Kornhuber, & Mountcastle, 1968). At the other extreme, behavioral performance is determined by assuming optimal pooling of the responses across all the neurons (Geisler, 1984, 1989; Watson, 1983). The results show that both pooling rules account for the data about equally well, and presumably, pooling rules between these two extremes could also account for the data reasonably well. Because of this, it may prove difficult to distinguish between different decision models (pooling rules) for discrimination tasks, even when both behavioral and neurophysiological data are available (Parker & Newsome, 1998). Nonetheless, the fact that the neural population performance matches the shape of the behavioral psychophysical performance regardless of the pooling rule suggests that the mechanisms up to and including area VI are largely responsible for limiting contrast and spatial frequency discrimination performance.

Finally, note that the heterogeneity of the contrast response functions and the spatial frequency tuning functions in cortical neurons appears to be the critical factor in determining the behavioral discrimination functions. Different cells are most sensitive to different regions along the stimulus dimensions and hence at each point along the stimulus dimension, different cells are determining the overall behavioral performance. Although population performance has not yet been determined for other stimulus dimensions, this principle, in which behavioral performance is dependent upon the continuous and heterogeneous distribution of cell properties, may prove to be general.

3. Psychophysical Models of Performance

a. Traditional Models

There are many aspects of spatial pattern discrimination that cannot be explained by the response properties of neurons in the retina. For example, in pattern adaptation experiments, a sine wave grating adaptor only elevates thresholds for spatial frequencies and orientations that are similar to that of the adapting grating. In pattern masking experiments, a sine wave grating masker only elevates thresholds for
spatial frequencies and orientations similar to that of the masking grating. Furthermore, retinal properties cannot explain interocular pattern masking and adaptation, or the complex effects of maskers with multiple frequency components. Finally, it is unlikely that either the contrast discrimination function or the spatial frequency discrimination function can be explained by retinal mechanisms.

Early models of spatial vision combined linear spatial frequency and orientation tuned channels with a simple pooling mechanism, such as selecting the most sensitive channel (Campbell & Robson, 1968) or probability summation (Brindley, 1960; Graham, Robson, & Nachmias, 1978; Wilson & Bergen, 1979). These models are able to qualitatively account for pattern adaptation and pattern masking. More recent models incorporated a final nonlinear response function prior to the pooling mechanism (Foley & Legge, 1981; Legge & Foley, 1980; Wilson, 1980; Wilson & Gelb, 1984). These models are able to quantitatively account for contrast discrimination, frequency discrimination, and pattern masking under limited conditions. The most recent models have replaced the final nonlinear response function with nonlinear mechanisms similar to the contrast normalization and response expansion found in the primate visual cortex (Foley & Boynton, 1994; Teo & Heeger, 1994; Thomas & Olzak, 1997; Watson & Solomon, 1997; Wilson & Humanski, 1993). These models are able to quantitatively account for a wider array of pattern masking effects. However, there remain a number of effects that cannot be explained, such as the strong masking effects that can occur when two weak maskers are presented simultaneously (see, for example, Derrington & Henning, 1989).

Most of these recent models of spatial vision consist of a small number of discrete channels, each tuned to a different range of spatial frequency and orientation, along with some nonlinear response mechanism. The parameters for these models are generally estimated by fitting psychophysical data. The sensitivities of the channels are estimated by fitting contrast sensitivity functions; the bandwidths of the channels and the nonlinearities are estimated by fitting masking data. The models are best able to account for the masking data when the bandwidths of the spatial frequency and orientation channels are similar to the average bandwidth of neurons in the primary visual cortex. However, for most of the models the estimated parameters of the nonlinearity are different from the average values for neurons in the primary visual cortex.

b. Neuron Sampling Models

The traditional models were designed to account for psychophysical data, and hence there is no requirement that they be consistent with all aspects of cortical neurophysiology. Nonetheless, there are important ways in which they are not consistent, and it may be valuable to consider models that more closely approximate the properties of the visual cortex. First, there is little neurophysiological evidence for a small number of discrete channels. Cortical neurons appear to be continuously variable in preferred spatial frequency, preferred orientation, spatial frequency bandwidth,
and orientation bandwidth. Second, the spatial vision models have generally assumed a single nonlinear contrast response function. In fact, cortical neurons appear to be continuously variable in their contrast response functions, just as they are in their spatial frequency and orientation tuning functions. These discrepancies with the physiology could be reduced by expanding the number of spatial channels and nonlinearities. Unfortunately, the cost is reduced parsimony and more free parameters. Each new channel or nonlinearity would add more parameters to already complex models.

However, even though cortical neurons are heterogeneous in their response properties, they are very regular in a different way. As we have seen (Section VI.A), it is possible to summarize the stimulus–response relationships of each cortical cell using a small set of relatively simple descriptive functions, where each descriptive function has only a few free parameters. This suggests that it would be useful to consider quantitative models of spatial vision that are defined in terms of probability distributions of descriptive function parameters (i.e., cell properties) rather than in terms of discrete channels or pathways. Specifically, the early visual system might be modeled as a large collection of neurons whose spatial-frequency tuning, orientation tuning, contrast response functions, and noise characteristics are randomly sampled from frequency distributions. We term this class of models neuron sampling models (Geisler & Albrecht, 1997). In a neuron sampling model, the frequency distributions for the descriptive function parameters replace the discrete channels as the fundamental construct of the model.

Neuron sampling models consist of three major components: (a) a functional model of single cortical neuron responses, which generates a predicted response mean and variance (e.g., the contrast-gain exponent model shown in Figure 12), (b) a set of frequency distributions, or histograms, for each key parameter in the functional model (taken from empirical measurements of the neurons or estimated from fitting psychophysical data), and (c) a pooling/decision rule (e.g., optimal pooling). To generate predictions for any stimulus of interest, a large population of model neurons is created by randomly sampling a value for each of the key parameters. (Any measured correlations that might exist between parameters can be incorporated in the sampling process.) The receptive field spacing (and average size) at each eccentricity is proportional to the cortical magnification at that eccentricity. The mean and variance of each neuron’s response is then determined for the given stimulus. Finally, a pooling/decision rule is applied to the population as a whole.

This class of model can generate predictions for simple patterns such as sine wave gratings as well as complex natural images.

VII. IMPLICATIONS FOR OBJECT RECOGNITION AND SCENE INTERPRETATION

The overall goal of the visual system is to provide an accurate three-dimensional description of the environment over time which includes the identification of objects and events. The retina contributes to this overall goal by (a) measuring the
intensity and wavelength information in the retinal image, (b) encoding small contrast variations independent of the ambient light level, (c) enhancing local image contours, and (d) compressing the information to a manageable size. The primary visual cortex contributes to this overall goal by representing the visual scene in terms of local image features/attributes, including (a) local orientation, (b) local size, (c) local spatial phase, and (d) local motion. This local representation is crucial for identification of objects and events by subsequent cortical regions.

Following the representation in terms of local image features/attributes, the next crucial step toward the overall goal is organizing the local image features into larger coherent groups, which form the basis for segregating individual objects from their context (i.e., solving the problem of “Object Segregation”). Once the features are organized into coherent wholes and the objects are segregated, object recognition becomes feasible. (For a recent review of this literature see Ullman, 1996.)

The linear and nonlinear receptive field properties of primary visual cortex neurons produce responses that are particularly well suited for grouping and segregation by subsequent brain mechanisms. The spatiotemporal receptive field shape makes each neuron selective to a specific local image feature. The expansive response exponent enhances this selectivity. The contrast normalization ensures that each neuron will only respond when a particular image feature is present—only a near optimal stimulus can produce a near maximum response. Finally, in the case of complex cells, the rectification nonlinearities produce phase-invariant responses.

These linear and nonlinear response properties create a population of cells with three particularly useful characteristics for object segregation. First, there are many different neurons selective to many different local image features (i.e., there is a wide range of narrow selectivities across the population of neurons as a whole). Second, each neuron only responds when a particular feature is present in the image (i.e., each neuron has a unique stimulus-response relationship). Third, there is a subpopulation of neurons (the complex cells) that respond invariantly to local spatial phase (i.e., they respond equivalently to the same stimulus anywhere within the receptive field). These three characteristics make it possible to find global similarities and differences within and between image regions along the dimensions of orientation, size, motion, contrast, and so on, which ultimately permits simple grouping (similarity grouping). Furthermore, these properties also make it possible to find smooth transitions in local similarities along the dimensions of orientation, size, motion, and so on, which ultimately permits contour integration and region integration.

9 The local image feature to which a cortical cell responds is not a single spatiotemporal pattern (such as a bar, or an edge, or sine wave grating) but is instead a collection, or set, of similar local spatiotemporal patterns. This collection of patterns defines the local image feature to which the cell responds. An appropriate method for quantifying the local image feature is to measure the confidence region in stimulus space when the stimulus is random and free to vary: specifically, the 95% confidence region in the space of all possible local two-dimensional spatiotemporal patterns that might occur in the natural environment (see section VI.B.1.b).
References


