4.1

Computational Models of Early Human Vision

Lawrence K. Cormack
The University of Texas at Austin

"The nature of things, hidden in darkness, is revealed only by analogizing. This is achieved in such a way that by means of simpler machines, more easily accessible to the senses, we lay bare the more intricate." — *Marcello Malpighi, 1675*

1 Introduction

1.1 Aim and Scope

The author of a short chapter on computational models of human vision is faced with an *embarras de richesse*. One wishes to make a choice between breadth and depth, but even this is virtually impossible within a reasonable space constraint. It is hoped that this chapter will serve as a brief overview for engineers interested in processing done by the early levels of the human visual system. We will focus on the representation of luminance information at three stages, the optics and initial sampling, the representation at the output of the eyeball itself, and the representation at primary visual cortex. With apologies, I have allowed us a very brief foray into the historical roots of the quantitative analysis of vision, which I hope may be of interest to some readers.

1.2 A Brief History

The first known quantitative treatment of image formation in the eyeball by Alhazan predated the Renaissance by four centuries. In 1604, Kepler codified the fundamental laws of physiologic optics including the then-controversial inverted retinal image, which was then verified by direct observation of the image *in situ* by Pater Scheiner in 1619 and later (and more famously) by Rene Descartes. Over the next two centuries there was little advancement in the study of vision and visual perception *per se* with the exception of Newton's formulation of laws of color mixture. However, Newton's seemingly innocuous suggestion that "the Rays to speak properly are not coloured" [1] anticipated the core feature of modern quantitative models of visual perception: the computation of higher *perceptual* constructs (e.g., color) based on the activity of peripheral receptors differentially sensitive to a *physical* dimension (e.g., wavelength).[1]

In 1801, Thomas Young proposed that the eye contained but three classes of photoreceptor, each of which responded with a sensitivity that varied over a broad spectral range [2].

[1] Newton was pointing out that colors must arise in the brain, because a given color can arise from many wavelength distributions, and some colors can *only* arise from multiple wavelengths. The purples for example, and even unique red (a red that observers judge as tinged with neither orange nor violet), are colors that cannot be experienced by viewing a monochromatic light.

This theory, including its extensions by Helmholtz, was arguably the first modern computational theory of visual perception. The Young/Helmholtz theory *explicitly* proposed that the properties of objects in the world are not sampled directly, but that certain properties of light are encoded by the nervous system, and that the resulting neural activity was transformed and combined by the nervous system to result in perception. Moreover, the neural activity was assumed to be quantifiable in nature, and thus the output of the visual system could be precisely predicted by a mathematic model. In the case of color, it could be firmly stated that sensation "may always be represented as simply a function of three variables" [3]. While not a complete theory of color perception, this has been borne out for a wide range of experimental conditions.

Coincident with the migration of trichromatic theory from England to Central Europe, some astronomical data made the same journey, and this resulted in the first *applied* model of visual processing. The data were observations of stellar transit times from the Greenwich Observatory taken in 1796. There was a half-second discrepancy between the observations by Maskelyne (the director) and Kinnebrook (his assistant), and for this Kinnebrook lost his job. The observations caught the notice of Bessel in Prussia at a time when the theory of variability was being given a great deal of attention due to the work of Laplace, Gauss, and others. Unable to believe that such a large, systematic error could be due to sloppy astronomy, Bessel developed a linear model of observers' reaction times to visual stimuli (i.e., stellar transits) relative to one another. These models, which Bessel called "personal equations" could then be used to correct the data for the individual making the observations.

It was no accident that the 19th century saw the genesis of models of visual behavior, for it was at that time that several necessary factors came together. First, it was realized that an understanding of the eyeball itself begged rather than yielded an explanation of vision.

Second, the brain had be to viewed as explainable, that is, viewed in a mechanistic fashion. While this was not entirely new to the 19th century, the measurement of the conduction velocity of a neural impulse by Helmholtz in 1850 probably did more than any other single experiment to demonstrate that the senses did not give rise to immediate, qualitative (and therefore incalculable) impressions, but rather transformed and conveyed information by means that were ultimately quantifiable.

Third, the stimulus had to be understood to some degree. To make tangible progress in modelling the *early* levels of the visual system it was necessary to think, not in terms of objects and meaningful structures in the environment, but of light, of wavelength, of intensity, and its spatial and temporal derivatives. The enormous progress in optics in the nineteenth century created a climate in which vision could be thought of quantitatively; light was not understood, but its veils of magic were quickly falling away.

Finally, theories of vision would have to constrained and testable in a quantitative manner. Experiments would have to be done in which observers made well-defined responses to well-controlled stimuli in order to establish quantitative input-output relationships for the visual system, which could then in turn be modeled. This approach, called *psychophysics*, was born with the publication of *Elemente der Psychophysik* by Gustav Fechner in 1860.

With the historical backdrop painted, we can now proceed to a selective survey of quantitative treatments of early human visual processing.

1.3 A Short Overview

Figure 1 shows a schematic overview of the major structures of the early visual system and some of the functions they perform. We start with the visual world, which varies with space, time, and wavelength, and has an amplitude spectrum roughly proportional to $1/f$, where f is the spatial frequency of luminance variation. The first major operations by the visual system are passive: lowpass filtering by the optics and sampling by the receptor mosaic, and both of these operations and the relationship between them vary with eccentricity.

The retina of the eyeball filters the image further. The photoreceptors themselves filter along the dimensions of time and wavelength, and both vary with receptor type. The output ccells of the retina, the retinal ganglion cells, synapse onto the lateral geniculate nucleus of the thalamus (known as the LGN). We will consider the LGN primarily as a relay station to cortex, and the properties of retinal ganglion cells and LGN cells will be treated as largely interchangeable.

LGN cells come in two major types in primates, Magnocellular ("M") and Parvocellular ("P"); the terminology was adopted for morphologic reasons, but important functional properties distinguish the cell types. To grossly simplify, M cells are tuned to low luminance spatial frequencies, high temporal frequencies, and are insensitive to variation in wavelength. In contrast, P cells are tuned to high luminance spatial frequencies, low temporal frequencies, and encode wavelength information. These two cell types work independently and in parallel, emphasizing different aspects of the same visual stimuli. In the 2D Fourier plane, both are circularly symmetric bandpass filters.

In the primary visual cortex, several properties emerge. Cells now encode three important stimulus properties that involve displacement of stimulus energy across an extensive dimension.

Cells become tuned to direction of motion (displacement across time), binocular disparity (displacement across eyeballs) and orientation (displacement across space). Because of

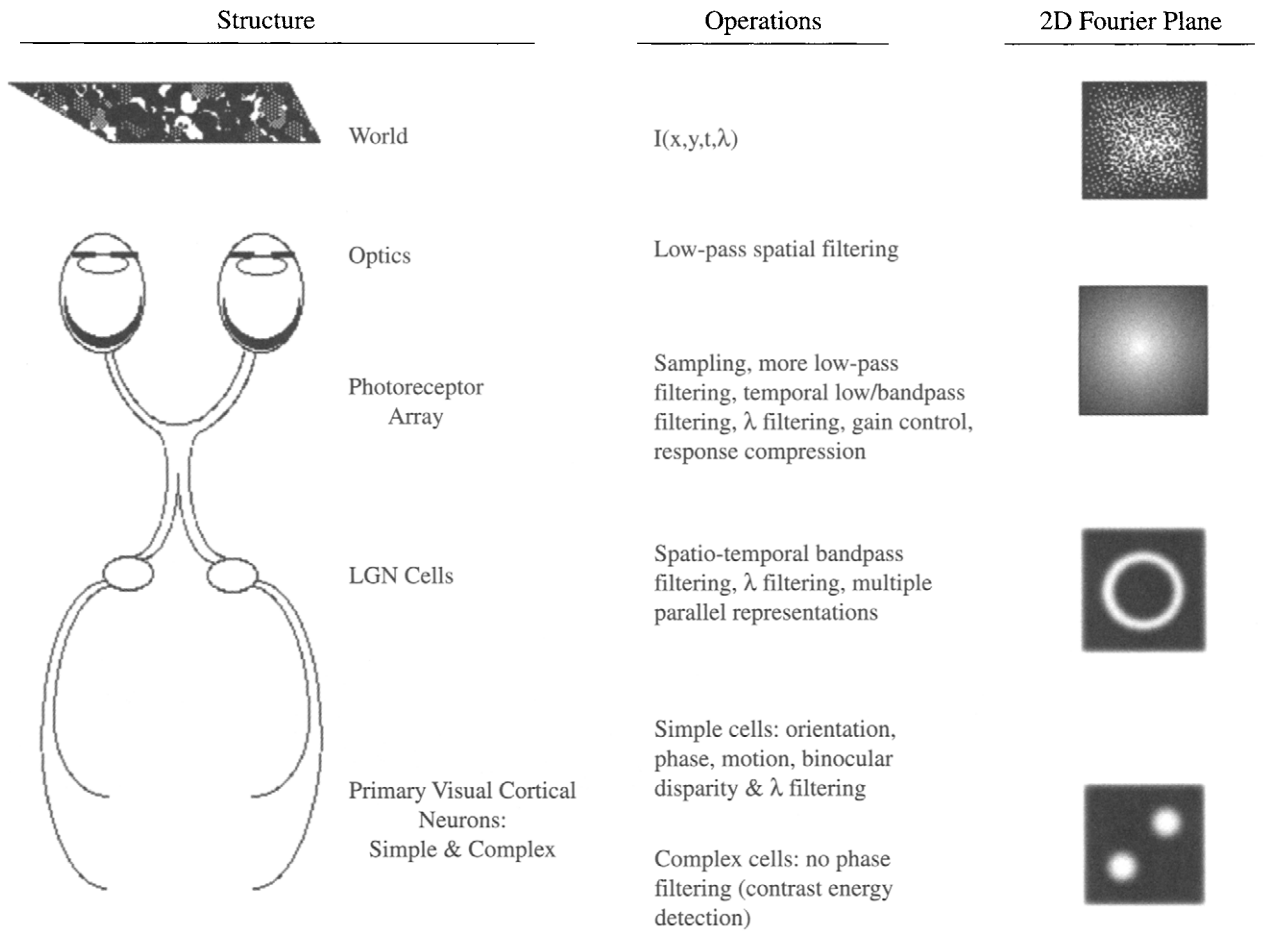


FIGURE 1 Schematic overview of the processing done by the early visual system. On the left are some of the major structures to be discussed. The middle column lists some of the major operations done at the associated structure. The right-hand column shows the 2D Fourier representation of the world, the retinal image, and the sensitivities typical of a ganglion and cortical cell.

the orientation tuning, they can be represented as Gaussian blobs on the spatial Fourier plane, as shown in Fig. 1.

A new dichotomy also emerges, that between so-called "simple" and "complex" cells. Simple cells behave much as wavelet-like linear filters, although they demonstrate some response nonlinearities critical to their function. The complex cells are more difficult to model, as their sensitivity shows no obvious spatial structure.

We will now explore the properties of each of these functional divisions, and their consequences, in turn.

2 The Front End

A scientist in biologic vision is likely to refer to anything between the front of the cornea and the area on which he or she is working as "the front end." Herein, we use the term to refer to the optics and sampling of the visual system and thus take advantage of the natural division between optical and neural events.

2.1 Optics

The optics of the eyeball are characterized by its 2D spatial impulse response function, the point-spread function [4]:

$$h(r) = 0.952e^{-2.59|r|^{1.36}} + 0.048e^{-2.43|r|^{1.74}} \quad (1)$$

in which r is the radial distance in minutes of arc from the center of the image.

This function, plotted in Fig. 2, (or its Fourier transform, the modulation-transfer function), completely characterizes the optics of the eye within the central visual field (completely encompassing the foveola, the central region in which our vision is most acute). The optics do deteriorate more substantially in the far periphery, so a spatially variant point-spread function

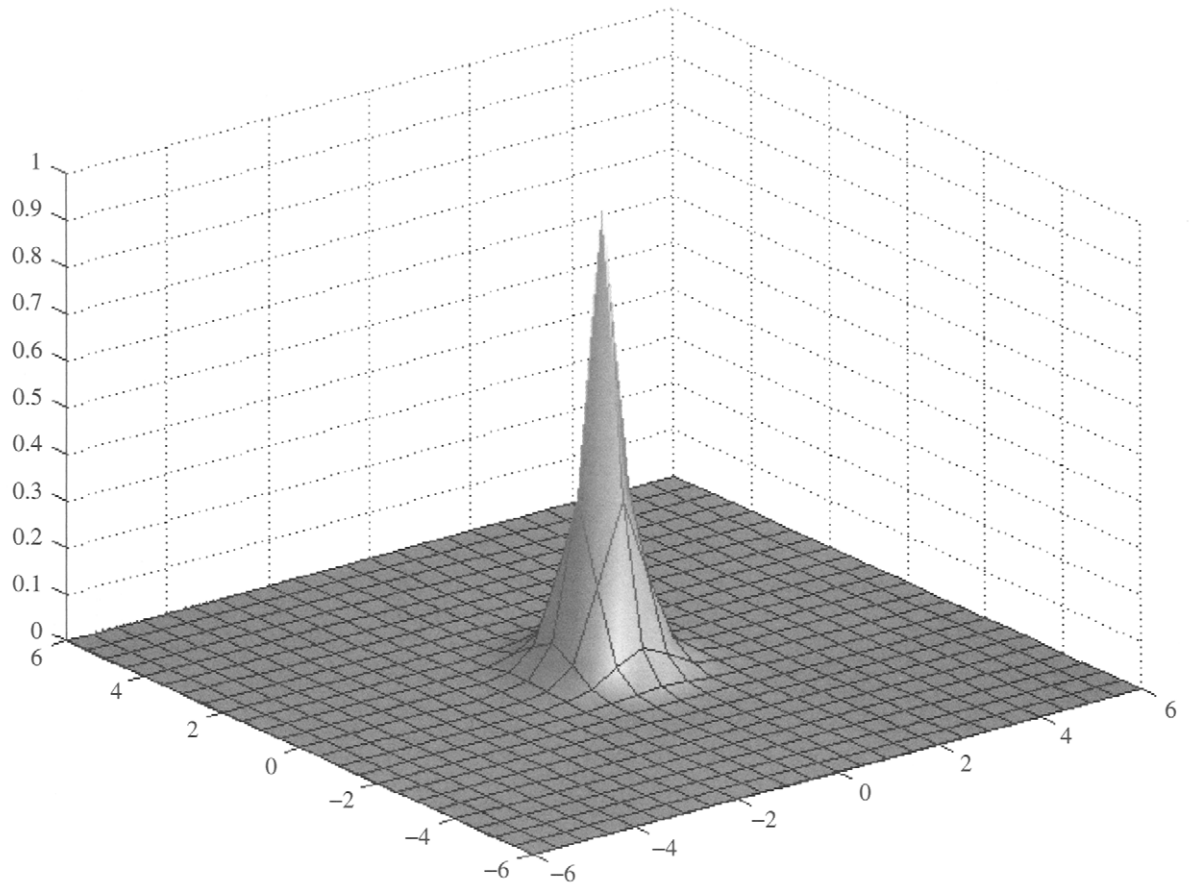


FIGURE 2 The point spread function of the human eyeball. The x and y axes are in minutes of arc, and the z axis is in arbitrary units. The spacing of the grid lines is equal to the spacing of the photoreceptors in the central visual field of the human eyeball, which is approximately 30 seconds of arc.

is actually required to fully characterize image formation in the human eyeball. For most purposes, however, the point-spread function may be simply convolved with an input image, viz.:

$$i(x, y) = I(x, y) * h(x, y) \tag{2}$$

to compute the central retinal image for an arbitrary stimulus, and thus derive the starting point of vision.

2.2 Sampling

While sampling by the retina is a complex spatio-temporal neural event, it is often useful to consider it to be a passive event governed only by the geometry of the receptor grid and the stationary probability of a single receptor absorbing a photon. In the human retina, there are two parallel sampling grids to consider, one comprising the rod photoreceptors and operating in dim light, and the other comprising the cone photoreceptors (on which we concentrate) and operating in moderate to bright light. Shown in Fig. 3 are images of the cone sampling grid one degree from the center of the fovea taken in two living, human eyes using aberration-correcting adaptive optics (similar to those used for correcting atmospheric distortions for terrestrial telescopes) [5]. The short-, medium-, and long-wavelength sensitive cones have been colored blue, green, and red, respectively. At the central fovea, the average inter-receptor distance is about 2.5 μm, which is about 30 seconds of arc in the human eyeball. Locally, the lattice is roughly hexagonal, but it is irregular over large areas and seems to become less regular as eccentricity increases. Theoretical performance has been compared in various visual tasks using both actual foveal receptor lattices taken from anatomic studies of the macaque[2] retina and idealized hexagonal lattices of the same receptor diameter, and little difference was found [6].

While the use of a regular hexagonal lattice is convenient for calculations in the space domain, it is often more efficient

[2] The macaque is an old-world monkey, *macaca fascicularis*, commonly used in vision research because of the great similarity between the macaque and human visual systems.

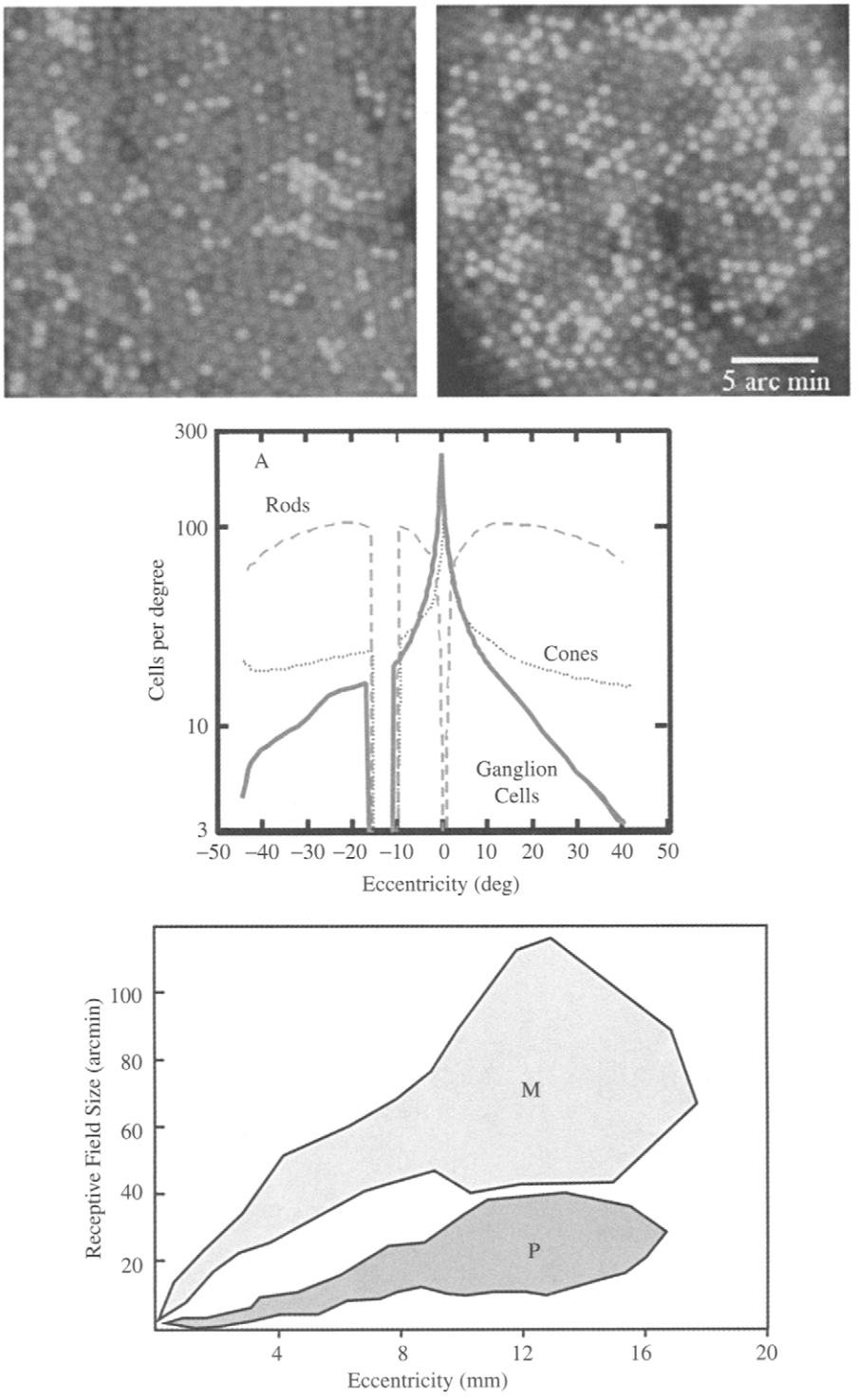


FIGURE 3 The upper panel shows the retinal sampling grid near the center of the visual field of two living human eyeballs. The different cone types are color-coded (from Roorda and Williams, 1999, reprinted with permission). The middle panel shows the density of various cell types in the human retina. The rods and cones are the photoreceptors that do the actual sampling in dim and bright light, respectively. The ganglion cells pool the photoreceptor responses and transmit information out of the eyeball (from Geisler and Banks, 1995). The lower panel shows the dendritic field size (assumed to be roughly equal to the receptive field size) of the two main types of ganglion cell in the human retina (redrawn from Dacy, 1993). The gray shaded region shows the parasol (or M) cells, and the green region shows the midget (or P) cells. The two cell types seems to independently and completely tile the visual world. The functional properties of the two cell types are summarized in Table 1 (see color insert).

TABLE 1 A summary of some of the important properties of the two major cell types providing input to the visual cortex

Property	P cells	M cells	Comments
Percent of cells	80	10	The remainder project to subcortical streams
Receptive field size	Relatively small, single cone center in fovea, increases with eccentricity (See Fig. 3)	Relatively large, about 3x larger than P cells at any given eccentricity	RF modelled well by a difference-of-Gaussians
Contrast sensitvity	Poor (factor of 8–10 lower than for M cells), driven by high contrast	Good, saturation at high contrasts	
Contrast gain	Low	High (about 6x higher)	Possible gain control in M cells
Spatial frequency response	Peak and high-frequency cutoff at relatively high spatial frequency	Peak and high-frequency cutoff at relatively low spatial frequency	Unclear dichotomy: physiologic differences tend to be less pronouced than predicted by anatomy
Temporal frequency response	Lowpass, falloff at 20–30 Hz	Bandpass, peaking at or above 20 Hz	
Spatial linearity	Almost all have linear summation	Most have linear summation, some show marked non-linearities	Estimated proportion of nonlinear neurons depends on how the distinction is made
Wavelength opponency	Yes	No	
Conduction velocity	Slow (6 m/sec)	Fast (15 m/sec)	

to work in the frequency domain. In the central retina, one can take the effective sampling frequency to be $\sqrt{3/2}$ times the average inter-receptor distance (due to the hexagonal lattice), and then treat the system as sampling with an equivalent 2D comb (sampling) function. In the peripheral retina, where the optics of the eye pass frequencies above the theoretical sampling limits of the retina, it is possible that the irregular nature of the array helps prevent some of the effects of aliasing. However, visual discriminations in the periphery can be made above the nyquist frequency via the detection of aliasing [7], so a 2D comb function of appropriate sampling density can probably suffice for representing the peripheral retina under some conditions.

The photoreceptor density as a function of eccentricity for the rod and cone receptor types in the human eye is shown in Fig. 3b. The cone lattice is *foveated,* peaking in density at a central location and dropping off rapidly away from this point. Also shown is the variation in the density of retinal ganglion cells that transmit the information out of the eyeball. The ganglion cells effectively sample the photoreceptor array in *receptive fields,* whose size also varies with eccentricity. This variation for the two main types of ganglion cells (which will be discussed below) is shown in Fig. 3. The ganglion cell distribution is also foveated, and it falls off more rapidly than cone density, indicating that ganglion cell receptive fields in the periphery summate over a larger number of receptors, thus sacrificing spatial resolution. This is reflected in measurements of visual acuity as a function of eccentricity, which fall in accord with the ganglion cell data [7].

The other main factor to consider is the probability of given receptor absorbing a photon, which is governed by the area of the effective aperture of the photoreceptor and the probability that a photon entering the aperture will be absorbed. This latter probability is obtained from Beer's Law, which gives the ratio of radiant flux reaching the back of the receptor outer segment to that entering the front [9]:

$$\upsilon(\lambda) = 10^{-lc\varepsilon(\lambda)} \tag{3}$$

in which l is the length of the receptor outer segment, c is the concentration of unbleached photopigment, and $\varepsilon(\lambda)$ is the absorption spectrum of the photopigment.

For many modelling tasks, it is most convenient to express the stimulus in terms of $n(\lambda)$, the number of quanta per second as a function of wavelength. This is given by [10]:

$$n(\lambda) = 2.24 \times 10^3 A \frac{L(\lambda)}{V(\lambda)} t(\lambda)\lambda \tag{4}$$

in which A is area of the entrance pupil, $L(\lambda)$ is the spectral luminance distribution of the stimulus, $V(\lambda)$ is the standard spectral sensitivity of human observers, and $t(\lambda)$ is the transmittance of the ocular media. Values of these functions are tabulated in [9].

Thus, for any receptor, the number of absorptions per second, N, is given approximately by:

$$N = \int a(1 - \upsilon(\lambda))n(\lambda)d\lambda \tag{5}$$

in which a is the receptor aperture.

These equations are of fundamental import because they describe the data that the visual system collects about the world. Any comprehensive model of the visual system must ultimately use these data as input. In addition, since these equations specify the information available to the visual system, they allow us to specify how well a particular visual task could be done in principle. This specification is done using a special type of model called an *ideal observer.*

2.3 The Ideal Observer

An ideal observer is a mathematic model that performs a given task as well as possible given the information in the stimulus. It is included in this section because it was traditionally used to assess the visual system in terms of quantum efficiency, f, which is the ratio of the number of quanta theoretically required to do a task to the number actually required [e.g., 11]. It is therefore more natural to introduce the topic in terms of optics. However the ideal observer has been used to assess the information loss at various neurophysiologic sites in the visual system [6, 12], the only requirement being that the information present at a given site can be quantitatively expressed.

The ideal observer performs a given task optimally (in the Bayesian sense), and it thus provides an *absolute* theoretical limit on performance in any given task (it thus gives to psychophysics and neuroscience what absolute zero gives to thermodynamics: a fundamental baseline). For example, the smallest offset between a pair of abutting lines (such as on a vernier scale on a pair of calipers) that a human observer can reliably discriminate (75% correct, say) from a stimulus with no offset is very small indeed — a few *seconds* of arc. Recalling from above that foveal cone diameters and receptor spacing are on the order of a half a *minute* of arc, such performance seems almost unbelievable. But why? With what do we make a comparison? The ideal observer gives us the answer by defining what the best possible performance is. In our example, a human observer would be less than 1% efficient as measured at the level of the photoreceptors, meaning that the human observer would require on the order of 10^3 more quanta to achieve the same level of discrimination performance. In this light, human performance ceases to appear quite so amazing, and attention can be directed towards determining how and where the information loss is occurring.

An ideal observer consists of two main parts, a model of the visual system and a Bayesian classifier. The latter is usually expressed as a likelihood ratio:

$$l(s) = \frac{P(s \mid a)}{P(s \mid b)} \tag{6}$$

in which the numerator and denominator are the conditional probabilities of making observations given that the stimulus was actually a or b, respectively. If the likelihood ratio, or more commonly its logarithm, exceeds a certain amount, stimulus a is judged to have occurred. For a simple discrimination, s would be a vector containing observed quantum catches in a set of photoreceptors, and the probability of this observation being made given hypotheses a and b would be calculated using the Poisson distribution of light and the factors described above in Optics and Sampling.

The beauty of the ideal observer is that it can be used to parse the visual system into layers, and examine the information loss at each layer. Thus, it becomes a tool by which we can learn which patterns of behavior result from the physics of the stimulus and the structure of the early visual system, and which patterns of behavior result from nonoptimal strategies or algorithms employed by the human visual system. For example, there exists an asymmetry in visual search in which a patch of low-frequency texture in a background of high-frequency texture is much easier to find than when the figure and ground are reversed. It is intuitive to think that if only low-level factors were limiting performance, detecting A on a background of B should be equivalent to detecting B on a background of A (by almost any measure, the contrast of A on B would equal that of B on A). However, an ideal-observer analysis proves this intuition false, and an ideal-observer based model of visual search produces the aforementioned search asymmetry [13].

3 Early Filtering and Parallel Pathways

In this section, we discuss the nature of the information that serves as the input to visual cortex. This information is contained in the responses of the retinal ganglion cells (the output of the eyeball) and the LGN.[3] Arguably, this is the last stage which can be comfortably modeled as a strictly data-driven system in which neural responses are independent of activity from other cells in the same or subsequent layers.

3.1 Spatio-Temporal Filtering

One difficulty with modelling neural responses in the visual system, particularly for someone new to reading the physiology literature, is that people have an affinity for dichotomies. This is especially evident from a survey of the work on retinogeniculate processing. Neurons have been dichotomized a number of dimensions. In most studies, only one or perhaps two of these dimensions are addressed, which leaves the relationships between the various dimensions somewhat unclear.

With that caveat in mind, the receptive field shown in Fig. 4 is fairly typical of that encountered in retinal ganglion cells or cells of the lateral geniculate nucleus. The upper panel shows the hypothetical cell's sensitivity as a function of spatial position. The receptive field profile shown is a difference-of-gaussians, which agrees well with physiologic recordings of the majority of ganglion cell receptive field profiles [14, 15], and is given by:

$$DOG(x,y) = a_1 e^{\left(\frac{x^2+y^2}{s_1^2}\right)} - a_2 e^{\left(\frac{x^2+y^2}{s_2^2}\right)} \tag{7}$$

in which a_1 and a_2 normalize the areas, and s_1 and s_2 are space constants in a ratio of about 1 : 1.6. Their exact values will vary as a function of eccentricity as per Fig. 3.

This representation is fairly typical of that seen in the early work on ganglion cells [e.g., 16], in which the peak response of a neuron to a small stimulus at a given location in the receptive field was recorded, but the location in *time* of this peak response was somewhat indefinite. Thus, a receptive field profile as shown represents a slice in time of the neuron's response some tens of milliseconds after stimulation and, further, the slice of time represented in one spatial location isn't necessarily the same as that represented in another (although for the majority of ganglion cells, the discrepancy would not be too large).

Since the receptive field is spatially symmetric, we can get a more complete picture by looking at a plot of one spatial dimension against time. Such an x-t plot is shown in the lower panel of Fig. 4, in which the x-dimension is in arcmin and the t-dimension is in msec. The response is space-time separable; the value at any given point is simply the value of the spatial impulse response at that spatial location scaled by the value of the temporal impulse response at that point in time. Thus, the response is given by

$$r(x,t) = DOG(x) \cdot [h(t)] \tag{8}$$

in which $h(t)$ is a biphasic temporal impulse response function. This response function, $h(t)$ was constructed by subtracting two cascaded lowpass filters of different order [cf. 17]. These low-pass filters are constructed by successive autocorrelation of an impulse response function of the form:

$$h(t) = H(t)e^{-t/\tau} \tag{9}$$

[3] Thus we regrettably omit a discussion of the response properties of the photoreceptors per se and of the circuitry of the retina. These are fascinating topics — the retina is a marvelous computational structure — and interested readers are referred to [44].

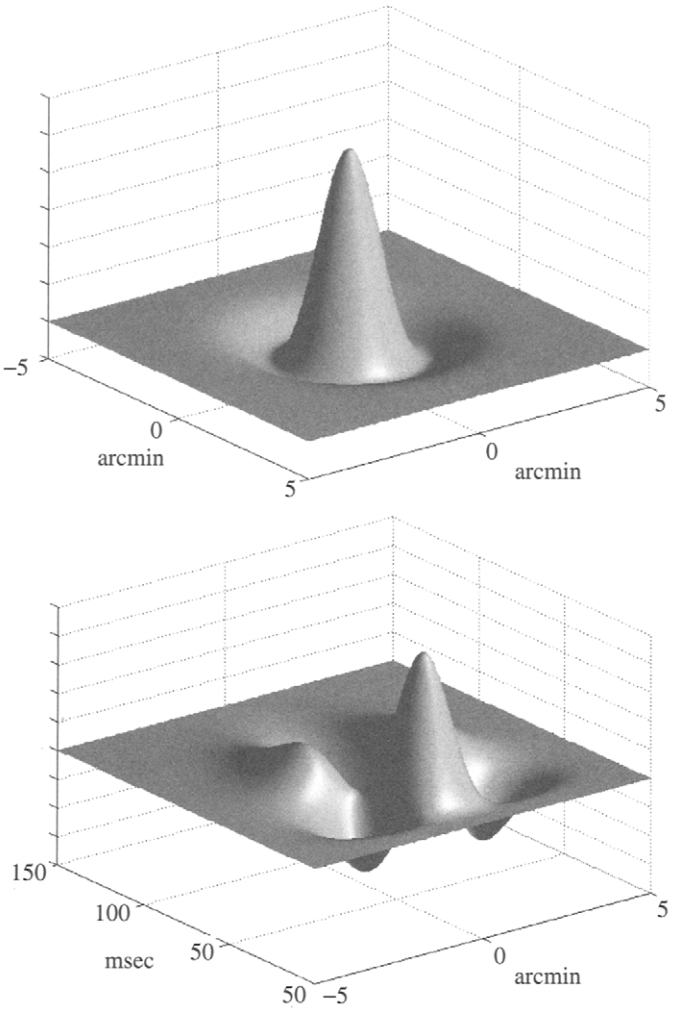


FIGURE 4 The upper panel shows receptive field profile of a retinal ganglion cell modeled as a difference-of-gaussians. The x and y axes are in minutes of arc, so this cell would be typical of an M cell near the center of the retina, or a P cell at an eccentricity of 10 to 15 degrees (see Fig. 2). The lower panel shows a space-time plot of the same receptive field, illustrating its biphasic temporal impulse response. The x-axis is in minutes of arc, and the y-axis is in milliseconds.

in which $H(t)$ is the Heaviside unit step:

$$H(t) = \begin{cases} 1, & t \geq 0 \\ 0, & t < 0 \end{cases} \tag{10}$$

A succession of n autocorrelations gives

$$h_n(t) = \frac{H(t)(t/\tau)^n e^{-t/\tau}}{\tau n!} \tag{11}$$

which is a monophasic (lowpass) filter of order n. A difference of two filters of different orders produces the biphasic bandpass response function, and the characteristics of the filter can be adjusted by using component filters of various orders.

The most important implication of this receptive field structure, obvious from the figure, is that the cell is bandpass in both spatial and temporal frequency. As such, the cell discards information about absolute luminance and emphasizes change across space (likely to denote the edge of an object) or change across time (likely to denote the motion of an object). Also obvious from the receptive field structure is that the cell is not selective for orientation (the direction of the spatial change) or the direction of motion.

The cell depicted in the figure is representative in terms of its qualitative characteristics, but the projection from retina to cortex comprises on the order of 10^6 such cells which vary in their specific spatio-temporal tuning properties. Rather than being continuously distributed, however, the cells seem to form functional subgroups that operate on the input image in parallel.

3.2 Early Parallel Representations

The early visual system carries multiple representations of the visual scene. The earliest example of this is at the level of the photoreceptors, where the image can be sampled by rods, cones, or both (at intermediate light levels). An odd aspect of the rod pathway is that it ceases to exist as a separate entity at the output of the retina; there is no such thing as a "rod retinal ganglion cell." This is an interesting example of a need for a separate sensor system for certain conditions combined with a need for neural economy. The pattern analyzing mechanisms in primary visual cortex and beyond are used for both rod and cone signals with (apparently) no information about which system is providing the input.

Physiologically, the most obvious example of separate, parallel projections from the retina to the cortex is the presence of the so-called ON- and OFF-pathways. All photoreceptors have the same sign of response. In the central primate retina, however, each photoreceptor makes direct contact with at least two *bipolar* cells — cells intermediate between the receptors and the ganglion cells — one of which preserves the sign of the photoreceptor response, and the other of which inverts it. Each of these bipolar cells in turn serves as the excitatory center of a ganglion cell receptive field, thus forming two parallel pathways: an ON pathway which responds to increases in light in the receptive field center, and an OFF pathway which responds to decreases in light in the receptive field center. Each system forms an independent tiling of the retina, resulting in two complete, parallel neural images being transmitted to the brain.

Another fundamental dichotomy is between midget (or "P" for reasons to become clear in a moment) and parasol (or "M") ganglion cells. Like the ON-OFF subsystems, the midget and parasol ganglion receptive fields perform a separate and parallel tiling of the retina. On average, the receptive fields of parasol ganglion cells are about a factor of 3 larger than those of midget ganglion cells at any given eccentricity (as shown in Fig. 3), so the two systems can be thought of as operating in parallel at different spatial scales. This separation is strictly maintained in the projection to the LGN, which is layered like a wedding cake. The midget cells project exclusively to what are termed the *parvocellular* layers of the LGN (the dorsal-most four layers), and the parasol cells project exclusively to the *magnocellular* layers (the ventral-most two layers). Because of this separation and the important physiologic distinctions that exist, visual scientist now generally speak in terms of the Parvocellular (or "P") pathway, and the magnocellular (or "M") pathway.

There is a reliable difference in the temporal frequency response between the cells of the M and P pathways [18]. In general, the parvocellular cells peak at a lower temporal frequency than parvocellular cells (<10 Hz vs. 10–20 Hz), have a lower high-frequency cutoff (approx. 20 Hz vs. approx. 60 Hz), and shallower low-frequency rolloff (with many P cells showing a DC response). The temporal frequency response envelopes of both cell types can be functionally modelled as a difference of exponentials in the frequency domain.

Another prevalent distinction is based upon linear vs. nonlinear summation within a cell's receptive field. Two major classes of retinal ganglion cell have been described in the cat, termed *X* and *Y* cells, based on the presence or absence of a null phase when stimulated with a sinusoidal grating [16]. The response of a cell such as shown in Fig. 4 will obviously depend strongly on the spatial phase of the stimulus. For such a cell, a spatial phase of a grating can be found such that the grating can be exchanged with a blank field of equal mean luminance with no effect on the output of the cell. These *X* cells compose the majority. For other cells, termed *Y* cells, no such null-phase can be found indicating that something other than linear summation across space occurs.

In the primate, nonlinear spatial summation is much less prevalent at the level of the LGN (although nonlinear cells do exist, and are more prevalent in M cells than in P cells [18]). It may be that nonlinear processing, which is very important, has largely shifted to the cortex in primates, just as have other important functions such as motion processing, which occurs much earlier in the visual systems of more phylogenically challanged species.

At this point, there is a great body of evidence suggesting that the M-P distinction is a fundamental one in primates, and that most of the above dichotomies are either an epiphenomenon of it, or at least best understood in terms of it. Table 1 provides a fairly comprehensive albeit qualitative overview of what we could term the Magnocellular and Parvocellular "geniculate-transforms" that serve as the input to the cortex [19]. If, in fact, work on the visual cortex continues to show effects such as malleability of receptive fields, it may be that developing models of geniculate function will actually increase in importance, because it may be last stage at which we can confidently rely on a relatively linear transform-type model. Attempts in this direction have been made [20, 21] but most modelling efforts seem to have been concentrated on either cortical cells or psychophysical behavior (i.e., modelling the output of the human as a whole, e.g., contrast threshold, in response to some stimulus manipulation).

4 The Primary Visual Cortex and Fundamental Properties of Vision

4.1 Neurons of the Primary Visual Cortex

The most striking feature of neurons in the visual cortex is the presence of several emergent properties. We begin to see,

for example, orientation tuning, binocularity, and selectivity for the direction of motion. The distinction between the magnocellular and parvocellular pathways remains — they synapse at different input layers in the visual cortex — but interactions between them begin to occur.

Perhaps the most obvious and fundamental physiologic distinction in cortex is between so-called simple and complex cells [22, 23]. This terminology was adopted — prior to wide application of linear systems analysis in vision — because in the case of the simple cells mapping the receptive field was straightforward and, once the receptive field was mapped, the response of the cell to a variety of patterns could be intuitively predicted (just as with ganglion cells). Complex cells, on the other hand, were more complex (see below). The simple/complex distinction seems to have no obvious relationship with the magnocellular/parvocellular distinction, but seems to be a manifestation of a computational scheme used within both processing streams.

The spatial receptive field of a generic simple cell is shown in the upper panel of Fig. 5. The cell is modeled as a Gabor function, in which sensitivity is given by:

$$s(x,y) = a \cdot e^{-\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)} \cdot \sin(2\pi\omega x + \phi) \qquad (12)$$

As the axes are in arcmin, the cell is most sensitive to horizontal Fourier energy at about 3 cyc/deg. In this case, the cell is odd-symmetric. While it would be elegant if cells always even or odd symmetric it seems that phase is continuously represented in cortex [24, 25], although this certainly does not preclude the use of pairs of cells in quadrature phase in subsequent processing.

As in Fig. 4, the lower panel shows the spatio-temporal receptive field of the model cell: the cell's sensitivity at $y=0$ plotted as a function of x and t. Notice that this model cell is spatiotemporally inseparable; it effectively changes its spatial phase over time. This makes it oriented in space-time and thus directionally selective [26, 27]. In this case, the optimal stimulus would be a 3 cyc/deg grating drifting at approximately 5 deg/sec. Many, but not all, cortical cells are directionally selective (see below).

Cells in the primary visual cortex can be thought of as banks of spatiotemporal filters that tile the visual world on several dimensions and, in so doing, determine the envelope of information to which we have access. We can get a feel for this envelope by looking the distribution of cell tuning along various dimensions. This is done in Fig. 6 using data from cells in macaque primary visual cortex reported in Geisler and Albrecht [28]. In the upper row, the response of an average cell is shown as a function of the spatial frequency of a counterphasing grating (left column), the temporal frequency of same stimulus at optimal spatial frequency (middle column), or the orientation of a drifting grating of optimal spatio-temporal frequency (right column). The middle and lower rows show the normalized frequency distributions of the parameters of the tuning functions for the population of cells surveyed ($n=71$).[4]

At this point, we can sketch a sort of standard model of the spatial response properties of simple and complex cortical cells [e.g., 28, 29]. The basic elements of such a model are illustrated in Fig. 7 (upper panel). The model comprises four basic components, the first of which is a contrast gain control which causes a response saturation to occur (see below). Typically, it takes the form of:

$$r(c) = \frac{c^n}{c^n + c_{50}^n} \qquad (13)$$

in which c is the image contrast, c_{50} is the contrast at which half the maximum response is obtained, and n is the response exponent, which averages about 2.5 for macaque cortical cells.

Next is the sampling of the image by a Gabor or gabor-like receptive field, which is a linear spatial summation:

$$f(x,y) = \sum c(x,y)h(x,y) \qquad (14)$$

in which $h(x,y)$ is the spatial receptive field profile, and $c(x,y)$ is the effective contrast of the pixel at (x,y), i.e. the departure of the pixel value from the average pixel value in the image.

The third stage is a half-wave rectification (unlike ganglion cells, cortical cells have a low maintained discharge and thus can signal in only one direction) and an expansive non-linearity, which serves to enhance the response disparity between optimal and non-optimal stimuli. Finally, Poisson noise is incorporated, which provides a good empirical description of the response variability of cortical cells. The variance of the response of a cortical cell is proportional to the mean response with an average constant of proportionality of about 1.7.

A model complex cell is adequately constructed by summing (or averaging) the output of two quadrature pairs of simple cells with opposite sign as shown in Fig. 7 (lower panel) [e.g., 29]. Whether complex cells are actually constructed out of simple cells this way in primary visual cortex is not known; they could be constructed directly from LGN input. For modelling purposes, using simple cells to construct them is simply convenient. The important aspect is that their response is phase-independent, and they thus behave as detectors of local contrast energy.

The contrast response of cortical cells deserves a little additional discussion. At first glance, the saturating contrast response function described above seems to be a rather

[4] While these distributions are based on real data, they are schematized using a gaussian assumption, which is not strictly valid. They do, however, convey a reasonable portrayal of the variability of the various parameters.

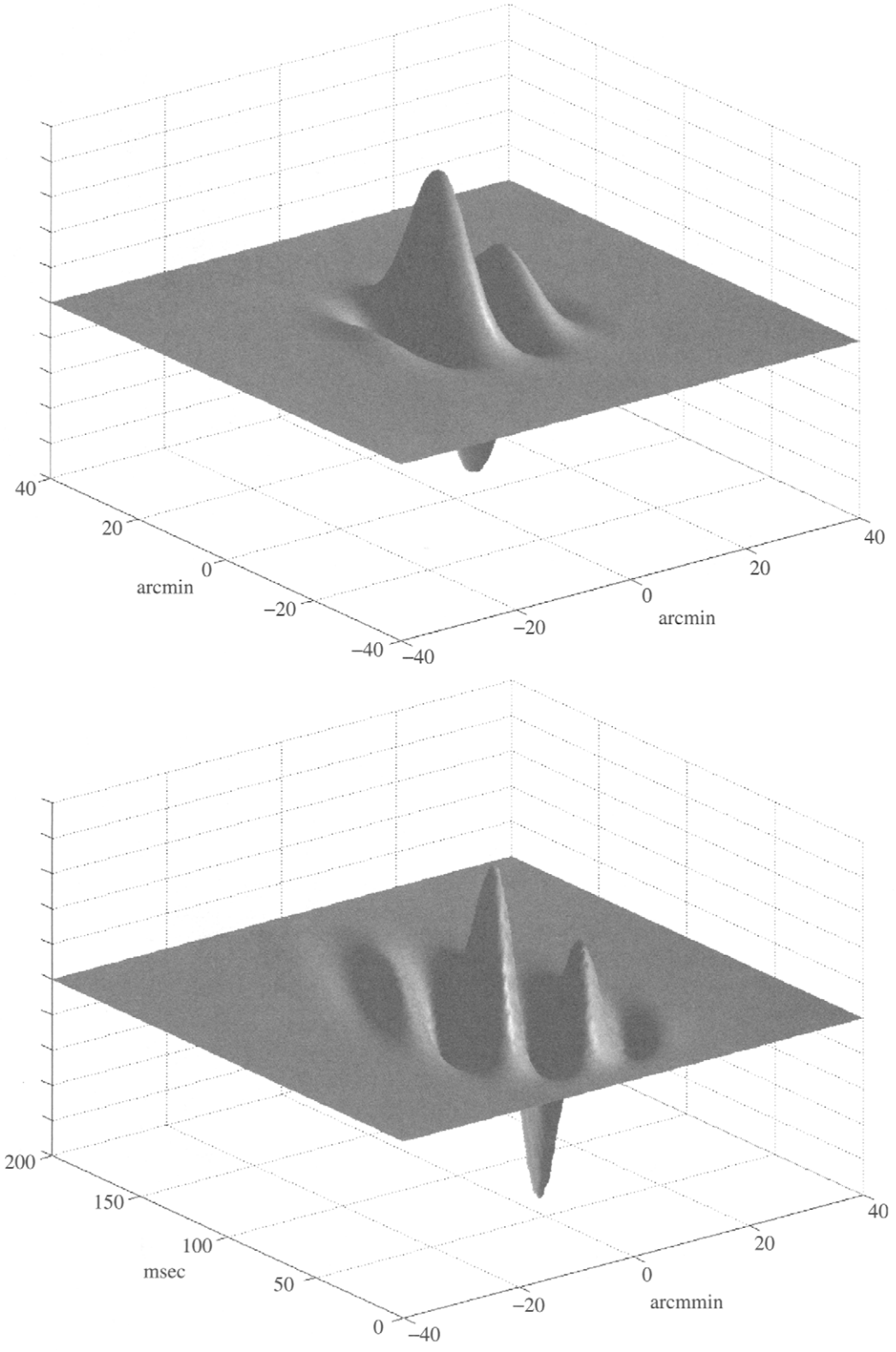


FIGURE 5 A receptive field profile of a cortical simple cell modeled as gabor function. The upper panel shows the spatial receptive field profile with the x and y axes in minutes of arc, and the z axis in arbitrary units of sensitivity. The lower panel shows a space-time plot of the same receptive field with the x axis in minutes of arc and the y axis in msec. The receptive field is space-time inseparable and the cell would be sensitive to rightward motion.

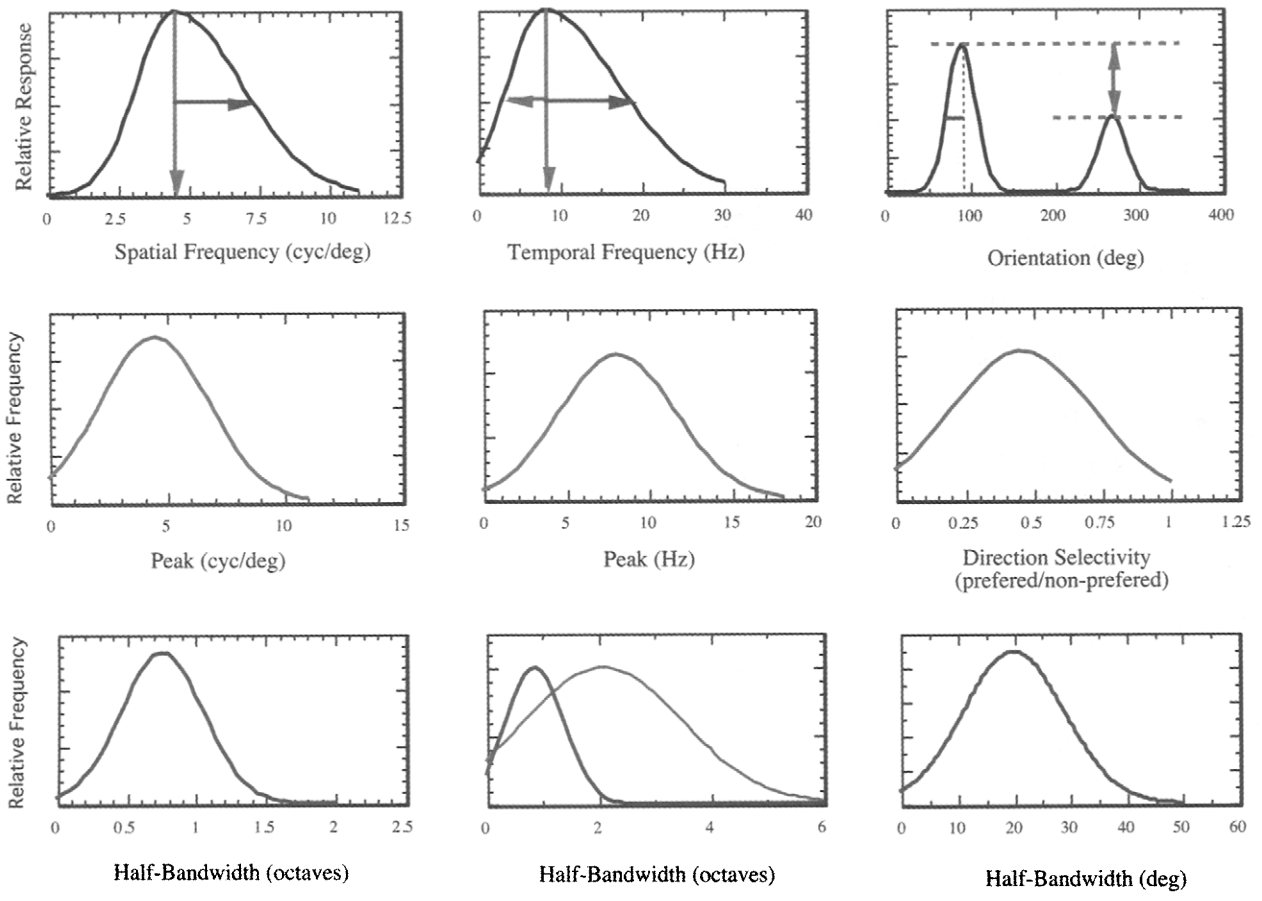


FIGURE 6 Left column: the upper panel shows a spatial frequency tuning profile typical of cell such as shown in Fig. 5. The middle and lower panels show distribution estimates of the two parameters of peak sensitivity (middle) and half-bandwidth in octaves (lower) for cells in macaque visual cortex. Middle column: same as the left column, but showing the temporal frequency response. As the response is asymmetric in octave bandwidth, the lower figure shows separate distributions for the upper and lower half-bandwidths (blue and green, respectively). Right column: the upper panel shows the response of a typical cortical cell to the orientation of a drifting sinusoidal grating. The estimate of half-bandwidth for macaque cortical cells is shown in the middle panel. The ratio of responses between the optimal direction and its reciprocal is taken as an index of directional selectivity; the estimated distribution of this ratio is plotted in the lower panel (the index cannot exceed unity by definition) (see color insert).

mundane response limit, perhaps imposed by metabolic constraints. However, a subtle but key feature is that the response of a given cortical neuron saturates at the same *contrast,* regardless of overall response level (as opposed to saturating at some given *response* level as might be expected given a metabolic limit). This is important because neurons have a multidimensional sensitivity manifold but a unidimensional output. Thus, if the output of a neuron increases from 10 to 20 spikes per second, say, then any number of things could have occurred to cause this. The contrast may have increased, the spatial frequency may have shifted to a more optimal one, etc., or any combination of such factors may have occurred. There is no way to identify which may have occurred from the output of the neuron.

But consider the effect of the contrast saturation on the output of the neuron for both an optimal and a non-optimal stimulus. Since the optimal stimulus is much more effective at driving the neuron, the saturation will occur at a higher response rate for the optimal stimulus. This partially defeats the response ambiguity: because of the contrast saturation, only an optimal stimulus is capable of driving the neuron to its maximum output. Thus, if a neuron is firing at or near its maximum output, the stimulus is specified fairly precisely. Moreover, the expansive nonlinearity magnifies this by enhancing small differences in output. Thus, 95% confidence regions for cortical neurons on, for example, the contrast/spatial frequency plane are much narrower than the spatial frequency tuning curves themselves [30]. This suggests that it is important to rethink the manner in which subsequent levels of the visual system may use the information conveyed by neurons in primary visual cortex. Over the last two and half decades, linear system analysis has dominated the thinking in vision science. It has been assumed that the act of perception would involve a large-scale comparison of

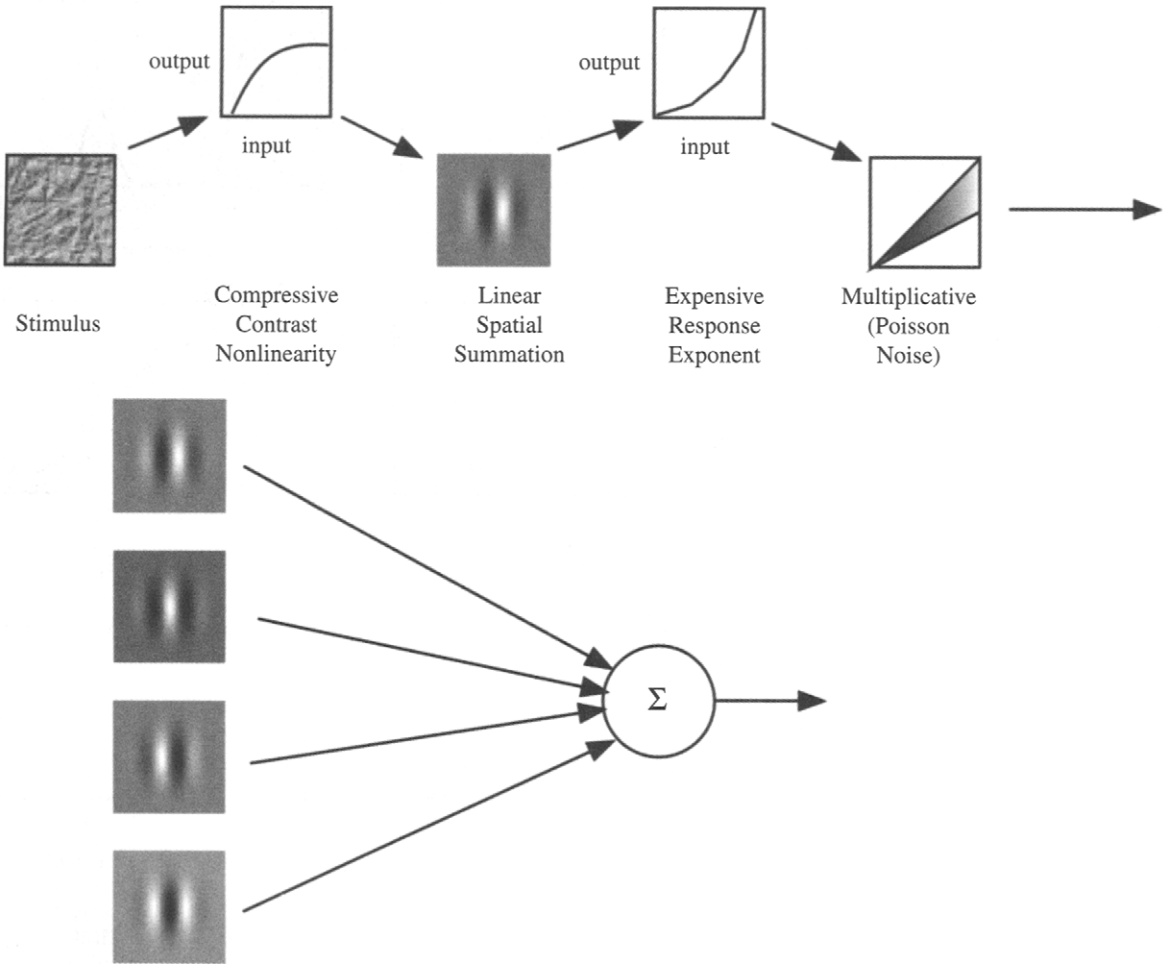


FIGURE 7 An overview of a model neuron similar to that proposed by Heeger and colleagues (1991, 1996) and Geisler and Albrecht (1997). An early contrast saturation precedes linear spatial summation across the gabor-like receptive field; the contrast saturation insures that only optimal stimuli can maximally stimulate the cell (see text). An expansive nonlinearity such as half-squaring enhances small differences in output. Multiplicative noise is then added; the variance of cortical cell output is proportional to the mean response (with the constant of proportionality about 1.7), so the signal-to-noise ratio grows as the square root of output. The lower panel illustrates the construction of a phase-independent (i.e., energy detecting) complex cell from simple cell outputs.

the outputs of many linear filters, outputs which would individually be very ambiguous. While such across-filter comparison is certainly necessary, it may be that the filters of primary visual cortex behave much more like "feature detectors" than we have been assuming. Moreover, it seems likely that the features that are being encoded represent, in some sense, an optimal set given the statistics of images arising from the natural environment [30–32].

Receptive profiles in cortex (such as shown in Fig. 5) probably bring to mind techniques such as the wavelet transform or Laplacian pyramid. Not surprisingly, then, most models of the neural image in primary visual cortex share the property of encoding the image in parallel at multiple spatial scales, and several such models have been developed. One model that is computationally very efficient and easy to implement is the cortex transform [34]. The cortex model is not, nor was it meant to be, a full model of the cortical representation. For example, response nonlinearities, the importance of which were discussed above, are omitted. It does, however, produce a simulated neural image that shares many of the properties of the simple cell representation in primary visual cortex. Models such as this have enormous value in that they give vision scientists a sort of testbed that can be used to investigate other aspects of visual function, e.g., possible interactions between the different frequency and orientation bands, in subsequent visual processes such as the computation of depth from stereopsis.

4.2 Motion and Cortical Cells

As mentioned previously, ganglion cell receptive fields are space-time separable. The resulting symmetry around a

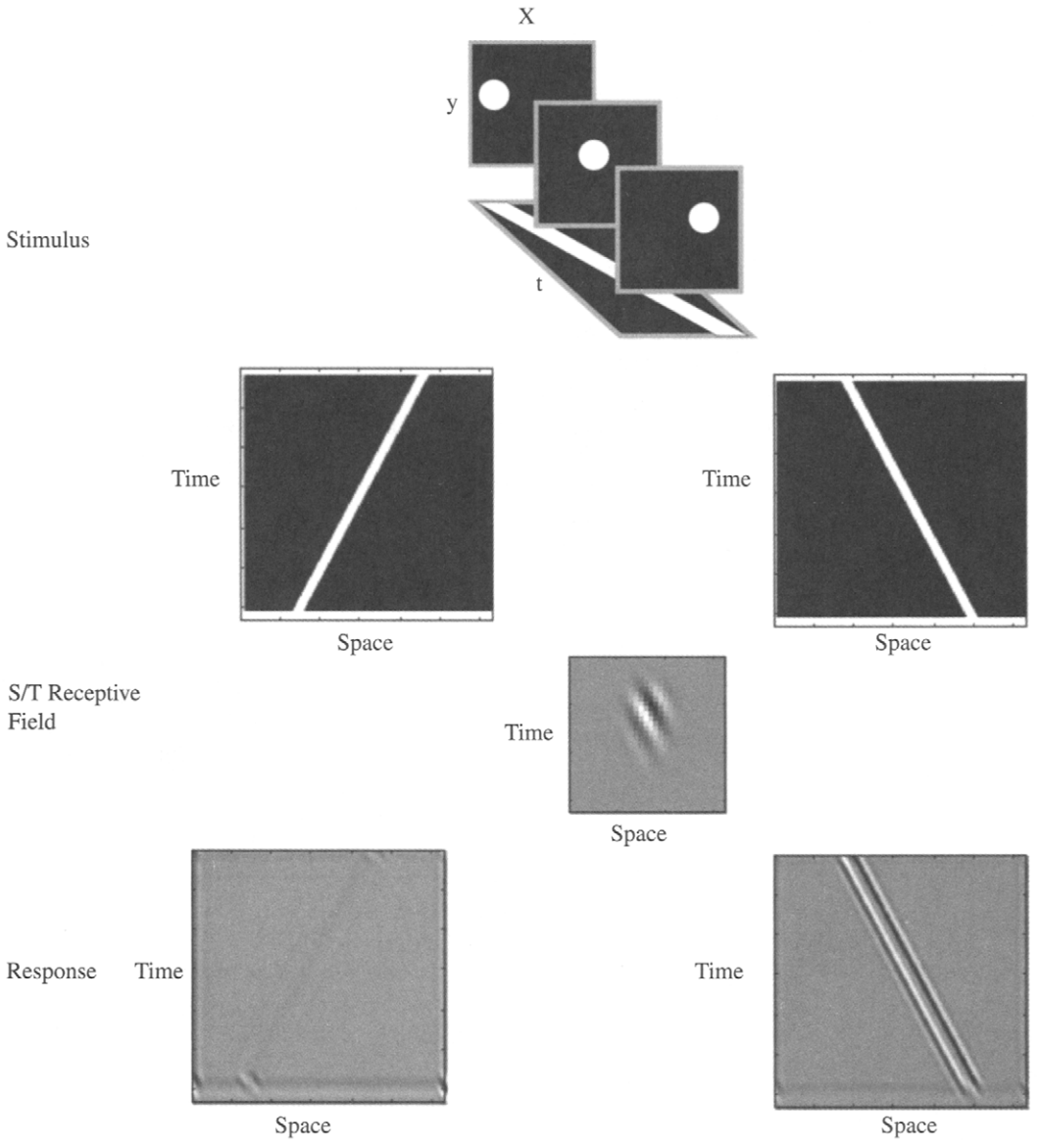


FIGURE 8 Three *x-y* slices are shown of a spot moving from left to right, and directly below is the continuous *x-t* representation: a diagonal bar. Below this are the space-time representations of a leftward and rightward moving bar, the receptive field of a directionally selective cortical cell (shown enlarged for clarity), and the response of the cell to the leftward and rightward stimuli.

constant-space axis (Fig. 4, lower panel) makes them incapable of coding the direction of motion. Many cortical cells, on the other hand, are directionally-selective.

In the analysis of motion, a representation in space-time is often most convenient. Figure 8 (top) shows three frames of a moving spot. The continuous space-time representation is shown projected onto the *x-t* plane below, and is simply an oriented bar in space-time. The next row of the figure shows the space-time representation of both a rightward and leftward moving bar. The third row of the figure shows a space time receptive field of a typical cortical cell as was also shown in Fig. 5 (for clarity, it is shown enlarged relative to the stimulus). Such space-time inseparable receptive fields are easily constructed from ganglion cell inputs by summing pairs of space-separable receptive fields (such as those shown in Fig. 4) which are in quadrature in both the space and time domains [26, 27]. The orientation of the receptive field in space-time gives it a fairly well defined velocity tuning.

The bottom row of the figure shows the response of such cells to the stimuli shown in the second row obtained by convolution. In these panels, each column represents the output of a cell as a function of time (row), and each cell has a receptive field centered at the spatial location represented

by its column. Clearly, each cell produces vigorous output modulation in response to motion in the preferred direction (with a relative time delay proportional to its spatial position, obviously), and almost no output in response to motion in the opposite direction.

For most purposes, it would be desirable to sense "motion energy." That is, one desires units that would respond to motion in one direction regardless of the sign of contrast or the phase of the stimulus. Indeed, such motion energy units may be thought of as the spatiotemporal equivalent of the complex cells described above. Similar to the construction of complex cells, such energy detectors are easily formed by, for example, summing the squared output of quadrature pairs of "simple" velocity sensitive units. Such a model captures many of the basic attributes of human motion perception, as well as some common motion illusions [26].

Motion sensing is vital. If nothing else, a primitive organism asks its visual system to sense moving things, even if it is only the change in a shadow which triggers a sea scallop to close. It is perhaps not surprising, then, that there seems to be a specialized cortical pathway, an extension of the magnocellular pathway an earlier levels, for analyzing motion in the visual field. A review of the physiology and anatomy of this pathway is clearly beyond the scope of this chapter. One aspect of the pathway worth mentioning here , however, is the behavior of neurons in an area of cortex known as MT, which receives input from primary visual cortex (it also receives input from other areas, but for our purposes, we can consider only its V1 inputs).

Consider a "plaid" stimulus, as illustrated in Fig. 9 (upper) composed of two drifting gratings differing in orientation by 90 degrees — one drifting up and to the right and the other up and to the left. When viewing such a stimulus, a human observer sees an array of alternating dark and light areas — the intersections of the plaid — drifting upward. The response of cells such as pictured in Fig. 5, however, would be quite different. Such cells would respond in a straightforward way according to the Fourier energy in the pattern, and would thus signal a pair of motion vectors corresponding to the individual grating components of the stimulus. Obviously, then, the human visual system incorporates some mechanism that is capable of combining motion estimates from filters such as the cells in primary visual cortex to yield estimates of motion for more complex structures. These mechanisms, corresponding to cells in area MT, can be parsimoniously modeled by combining complex cell outputs in manner similar to that by which complex cells can be constructed from simple cell outputs [35, 36]. These cells effectively perform a local sum over the set of cells tuned to the appropriate orientation and spatio-temporal frequency combinations consistent with a real object moving in a given direction at a given rate. In effect, then, these cells are a neural implementation of the intersection-of-constraints solution to the aperture problem of edge (or grating) motion [37]. This problem is illustrated in Fig. 9 (middle and lower). In Fig. 9 (lower), an object is shown moving to the right with some velocity. Various edges along the object will stimulate receptive fields with the appropriate orientation. Clearly, these individual cells have no way of encoding the true motion of the *object*. All they can sense is the motion of the *edge*, be it almost orthogonal to the motion of the object at a relatively low speed, or in the direction of the object at a relatively high speed. The set of motion vectors generated by the edges, however, must satisfy the intersection of motion constraint as illustrated in Fig. 9 (lower). The endpoints of the motion vectors generated by the moving edges lie on a pair of lines which intersect at the true motion of the object. Thus, a cell summing (or averaging) the outputs of receptive fields of the appropriate orientation and spatiotemporal frequency (i.e., speed) combinations will effectively be tuned to a particular velocity and largely independent of the structure moving at that velocity.

4.3 Stereopsis and Cortical Cells

Stereopsis refers to the computation of depth from the image displacements which result from the horizontal separation of the eyeballs. Computationally, stereopsis is closely related to motion, the former involving displacements across viewpoint rather than across time. For this reason, the development of models in the two domains has much in common. Early models tended to focus on local correlations between the images, and excitatory and/or inhibitory interactions in order to filter our false matches (spurious correlations).

As with motion, however, neurophysiologic and psychophysic findings [e.g., 38] have served to concentrate efforts on models based on receptive field structures similar to those found in Fig. 5. Of course, this is not incompatible with disparity domain interactions, but ambiguity is more commonly eliminated via interactions between spatial scales.

The primary visual cortex is the first place along the visual system in which information from the two eyes converges on single cells; as such, it represents the beginning of the binocular visual processing stream. Traditionally, it has been assumed that in order to encode horizontal disparities these binocular cells received monocular inputs from cells that had a different receptive field *locations* in the two eyes, thus being maximally stimulated by an object off the plane of fixation. It is now clear, however, that binocular simple cells in the primary visual cortex often have receptive fields like that shown in Fig. 5, but with different *phases* between the two eyes [39].[5] The relative phase relation between the receptive fields in the two eyes is distributed uniformly (not in quadrature pairs) for cells tuned to vertical

[5] Many recent studies have not measured the absolute receptive position in the two eyes, as it is very difficult to do. Thus, the notion that absolute monocular receptive field position plays a role in stereopsis cannot be rejected.

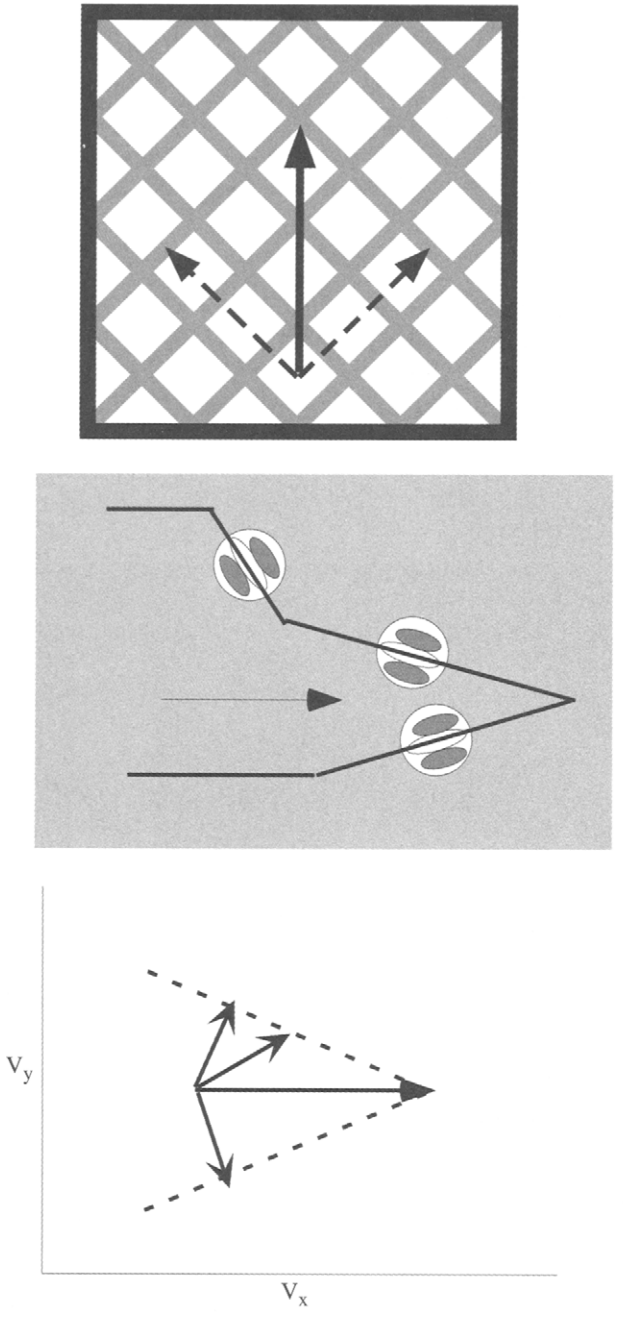


FIGURE 9 Top: Two gratings drifting obliquely (dashed arrows) generate a percept of a plaid pattern moving upward (solid arrow). Middle: an illustration of the aperture problem and the ambiguity of motion sensitive cells in primary visual cortex. Each cell is unable to distinguish a contour moving rapidly to the right from a contour moving more slowly perpendicular to its orientation. Lower: the intersection of constraints which allows cells that integrate over units such as in the middle panel to resolve the motion ambiguity.

orientations while there is little phase difference for cells tuned to horizontal orientations, indicating that these phase differences are almost certainly involved in stereopsis. Just as in motion, however, these simple cells have many undesirable properties, such as phase sensitivity and phase ambiguity a phase disparity $k\pi$ being indistinguishable from a phase disparities of $2nk\pi$, n an integer.

To obviate the former difficulty, an obvious solution would be to build a binocular version of the complex cell by summing across simple cells with the same disparity tuning but various monocular phase tunings [e.g., 40]. Such construction is analogous to the construction phase-independent, motion-sensitive complex cells discussed earlier, except that the displacement of interest is across eyeballs instead of time.

This has been shown to occur in cortical cells and, in fact, these cells show more precise disparity tuning than 2D position tuning [41].

Yet, because these cells are tuned to a certain phase disparity of a given *spatial frequency*, there remains an ambiguity concerning the absolute disparity of a stimulus. This can be seen in Fig. 10, which plots the output (as brightness) of a hypothetical collection of cells tuned to various values of phase disparity, orientation, and spatial frequency. The tuning of the cell is given by its position in the volume (in the upper panel orientation is ignored, and only a single spatial frequency/disparity surface is shown). In the upper panel, note that the output of cells tuned to a single spatial frequency contains multiple peaks along the dimension of disparity, indicating the phase ambiguity of the output. It has been suggested that this ambiguity could be resolved by units that sum the outputs of disparity units across spatial frequency and orientation [e.g., 40]. Such units would solve the phase ambiguity in a manner very analogous to the intersection of constraints solution to motion ambiguity described above. In the case of disparity, as a broadband stimulus is shifted along the disparity axis, it yields a sinusoidal variation in output at all spatial frequencies, but the frequency of modulation is proportional to the spatial frequency to which the cells are tuned. The resolution to the ambiguity lies in the fact that there is only *one* disparity at which peak output is obtained at *all* spatial frequencies, and that is the true disparity of the stimulus. This is shown in Fig. 10 by the white ridge running down the spatial frequency — disparity plane in the upper panel.

The pattern of outputs of cells tuned to a single spatial frequency but to a variety of orientations as a function of disparity is shown on the floor of the lower panel of Fig. 10. Summing across cells tuned to different orientations will also disambiguate disparity information because a Fourier component at an oblique orientation will behave as a vertical component with a *horizontal* frequency proportional to the cosine of the angle of its orientation from the vertical.

The lower panel of Fig. 10 is best thought of as a volume of cells whose sensitivity is given by their position in the volume (for visualization convenience, the phase information is repeated for the higher spatial frequencies, so the phase tuning is giving by the position on the disparity axis modulo 2π). The combined spatial frequency and disparity information results in a surface of maximum activity at the true disparity of a broadband stimulus, so a cell which sums across surfaces in this space will encode for physical disparity independent of spatial frequency and orientation.

Very recent work indicates that cells in MT might perform just such a task [42]. Recall from above that cells in MT decouple velocity information from the spatial frequency and orientation sensitivity of motion selective cells. DeAngelis et al. [42] have discovered a patterned arrangment of disparity sensitive cells in the same area, and have demonstrated their consequence in perceptual judgements. Given the conceptually identical nature of the ambiguities to be resolved the domains of motion and disparity, it would seem likely that the disparity-sensitive cells in MT perform role in stereopsis analogous to that which the velocity-sensitive cells play in motion perception.

5 Concluding Remarks

Models are wonderful tools and have an indispensable role in vision science. Neuroscientists must reverse-engineer the brain, and for this the methods of engineering are required. But the tools themselves can lead to biases (when all you have is a hammer, everything looks like a nail). There is always a danger of carrying too much theory, often implicitly, into an analysis of the visual system. This is particularly true in the case of modelling, because a model must have a quantitative output and thus must be specified, whether intentionally or not, at what Marr called the level of computational theory [13]. In short, a computational model, like any theory, framework, or way of thinking, makes a good servant but a bad master.

Yet without quantitative models, it would be very hard to compare psychophysics (human behavior) and physiology in deep or meaningful ways. This may seem like a strong statement, but there are subtle flaws in simple comparisons between the results of human experiments and single-cell response profiles. Consider an example taken from [43]. The experiment was designed to reveal the underlying mechanisms of disparity processing. A "mechanism" is assumed to comprise a group of neurons with similar tuning properties (peak location and bandwidth) on the dimension of interest working in parallel to encode that dimension. The tuning of the mechanism then reflects the tuning of the underlying neurons. This experiment used the typical psychophysical technique of *adaptation*. In this technique, one first measures the sensitivity of human observers along a dimension; in this case, we measured the sensitivity to the interocular correlation of binocular white noise signals as a function of binocular disparity. Following this, the subjects adapted to a signal at a given disparity. This adaptation fatigues the neurons sensitive to this disparity and therefore reduces the sensitivity of any mechanism comprising these neurons. Re-testing sensitivity, we found that it is systematically elevated in the region of the adaptation, and a difference between the pre- and post-adaptation sensitivity yields a "tuning profile" of the adaptation, for which a peak location, bandwidth, etc. can be defined.

But what *is* this tuning profile? In these types of experiments, it is tempting to assume that it directly reflects the sensitivity profile of an underlying mechanism, but this would be a dangerous and generally wrong assumption. The tuning profile actually reflects the combined outputs

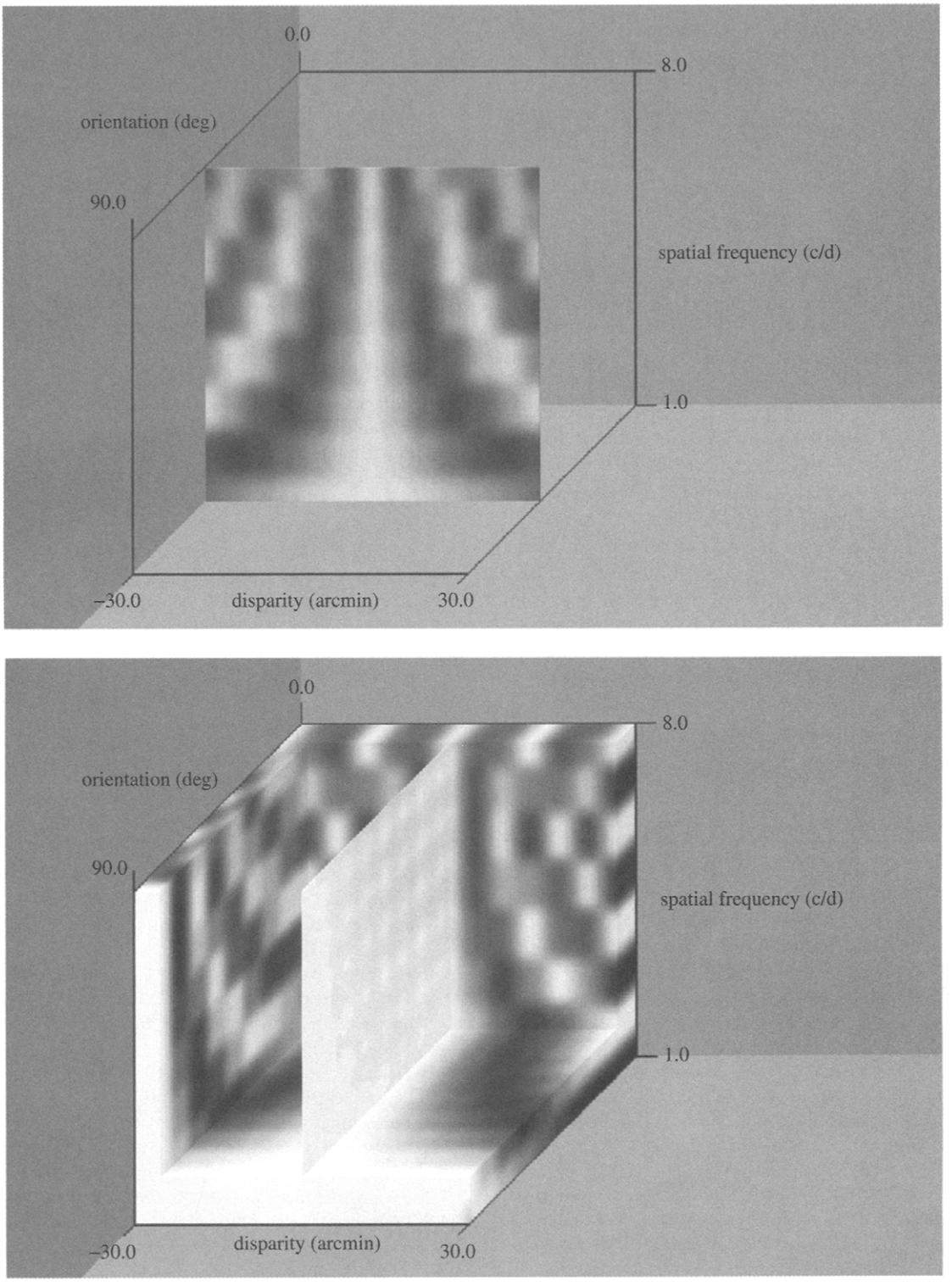


FIGURE 10 The upper panel shows the output of cortical cells on the spatial-frequency/disparity plane. The output of any one cell uniquely specifies only a phase disparity, but summation across spatial frequencies at the appropriate phase-disparities uniquely recovers absolute disparity. The lower panel adds orientation to this representation.

of numerous mechanisms in response to the adaptation. The degree to which the tuning profile itself resembles any one of the individual underlying mechanisms depends on a number of factors involving the nature of the mechanisms themselves, their interaction, and how they are combined at subsequent levels to determine overall sensitivity.

If one can't get a direct glimpse of the underlying mechanisms using psychophysics, how does one reveal them? This is where computational models assert their value. We constructed various models incorporating different numbers of mechanisms, different mechanism characteristics, and different methods of combining the outputs of mechanisms. We found that with a small number of disparity-sensitive mechanisms (e.g., three, as had been proposed by earlier theories of disparity processing) we were unable to simulate our psychophysical data. With a larger number of mechanisms, however, we able to reproduce our data rather precisely, and the model became much less sensitive to the manner in which the outputs of the mechanisms were combined.

So while we are unable to get a *direct* glimpse at underlying mechanisms using psychophysics, models can guide us in determining what kinds of mechanisms can and cannot be used to produce sets of psychophysical data. As more physiologic data become available, more precise models of the neurons themselves can be constructed, and these can be used, in turn, within models of psychophysical behavior. It is thus that models sew together psychophysics and physiology, and I would argue that without them the link could never be but tenuously established.

References

[1] Newton, I. (1931) Opticks. G. Bell & Sons, London.

[2] Young, T. (1802) On the theory of light and colour. Philosophical Transactions of the Royal Society. 73, 12–48.

[3] Helmholtz, H. V. (1962) Treatise on Physiological Optics. Dover Publications, New York.

[4] Westheimer, G. (1986) The eye as an optical instrument. In K. R. Boff, L. Kaufman, and J. P. Thomas (Eds.), Handbook of Human Perception and Performance. Wiley and Sons, New York.

[5] Roorda, A. and Williams, D. R. (1999) The arrangement of the three cone classes in the living human eye. Nature, 397, 520–522.

[6] Geisler, W. S. (1989) Sequential ideal-observer analysis of visual discriminations. Psychological Review, 96, 267–314.

[7] Williams, D. R. and Coletta, N. J. (1987) Cone spacing and the visual resolution limit. Journal of the Optical Society of America, A, 4, 1514–1523.

[8] Merigan, W. H. and Katz, L. M. (1990) Spatial resolution across the macaque retina.Vision Research, 30, 985–991.

[9] Wyszecki, G. and Stiles, W. S. (1982) Color Vision. Wiley and Sons, New York.

[10] Geisler, W. S. and Banks, M. S. (1995) Visual Performance. In (M. Bass Ed.) Handbook of Optics. McGraw-Hill, New York.

[11] Barlow, H. B. (1962) Measurements of the quantum efficiency of discrimination in human scotopic vision. Journal of Physiology, 150, 169–188.

[12] Pelli, D. G. (1990) The quantum efficiency of vision. In C. Blakemore (Ed.), Vision: Coding and Efficiency. Cambridge University Press, Cambridge.

[13] Geisler, W. S. and Chou, K. (1995) Separation of low-level and high-level factors in complex tasks: visual search. Psychological Review, 102, 356–378.

[14] Marr, D. (1982) Vision. Freeman and Co., New York.

[15] Rodieck, R. W. (1965) Quantitative analysis of cat retinal ganglion cell response to visual stimuli. Vision Research, 5, 583–601.

[16] Enroth-Cugel, C. and Robson, J. G. (1966) The contrast sensitivity of retinal ganglion cells in the cat. Journal of Physiology, 187, 517–522.

[17] Watson, A. B. (1986) Temporal Sensitivity. In K. R. Boff, L. Kauffman, and J. P. Thomas (Eds.), Handbook of Perception and Human Performance. Wiley and Sons, New York.

[18] Derrington, A. M. and Lennie, P. (1984) Spatial and temporal contrast sensitivities of neurons in the lateral geniculate nucleus of macaque. Journal of Physiology, 357, 2219–240.

[19] Lennie, P. (1993) Roles of M and P pathways. In R. Shapley and D. M. K. Lam (Eds.), Contrast Sensitivity. The MIT Press, Cambridge.

[20] Troy, J. B. (1993) Modeling the receptive fields of mammalian retinal ganglion cells. In R. Shapley and D. M. K. Lam (Eds.), Contrast Sensitivity. The MIT Press, Cambridge.

[21] Donner, K. and Hemila, S. (1996) Modelling the spatio-temporal modulation response of ganglion cells with difference-of-Gaussians receptive fields: Relation to photoreceptor response kinetics. Visual Neuroscience, 13, 173–186.

[22] Hubel, D. H. and Weisel, T. N. (1962) Receptive fields, binocular interaction and functional archetecture in the cat's visual cortex. Journal of Physiology, 160, 106–154.

[23] Skottun, B. C., DeValois, R. L., Grosof, D. H., Movshon, J. A., Albrecht, D. G., and Bonds, A. B. (1991) Classifying simple and complex cells on the basis of response modulation. Vision Research, 31, 1079–1086.

[24] Hamilton, D. B., Albrecht, D. G., and Geisler, W. S. (1989) Visual cortical receptive fields in monkey and cat: spatial and temporal phase transfer function. Vision Research, 29, 1285–1308.

[25] Field, D. J. and Tolhurst, D. J. (1986) The structure and symmetry of simple-cell receptive-field profiles in the cat's visual cortex. Proceedings of the Royal Society of London, 228, 379–400.

[26] Adelson, E. H. and Bergen, J. R. (1985) Spatiotemporal energy models for the perception of motion. Journal of the Optical Society of America, A, 2, 284–299.

[27] Watson, A. B. and Ahumada, A. J. (1985) Spatiotemporal energy models for the perception of motion. Journal of the Optical Society of America, A, 2, 322–341.

[28] Geisler, W. S. and Albrecht, D. G. (1997) Visual cortex neurons in monkeys and cats: Detection, discrimination, and stimulus certainty. Visual Neuroscience, 14, 897–919.

[29] Heeger, D. J. (1991) Nonlinear model of neural responses in cat visual cortex. In M. S. Landy and J. A. Movshon (Eds.), Computational Models of Visual Processing. The MIT Press, Cambridge.

[30] Geisler, W. S. and Albrecht, D. G. (1995) Bayesian analysis f identification performance in monkey visual cortex: nonlinear mechanisms and stimulus certainty. Vision Research, 35, 2723–2730.

[31] Field, D. J. (1987) Relations between the statistics of natural images and the response properties of cortical cells Journal of the Optical Society of America, A, 4, 2379–2394.

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

[33] van Hateren, J. H. and van der Schaaf, A. (1998) Independent component filters of natural images compared with simple cells in primary visual cortex. Proc. R. Soc. Lond. B, 265, 359–366.

[34] Watson, A. B. (1987) The cortex transform: Rapid computation of simulated neural images. Computer vision, graphics, and image processing, 39, 311–327.

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

[36] Heeger, D. J., Simoncelli, E. P. and Movshon, J. A. (1996) Computational models of cortical visual processing. Proceedings of the National Academy of Science, 93, 623–627.

[37] Adelson, E. H. and Movshon, J. A. (1982) Phenomenal coherence of visual moving pattens. Nature, 300, 523–525.

[38] DeAngelis, G. C., Ohzawa, I., and Freeman, R. D. (1991) Depth is encoded in the visual cortex by a specialized receptive field structure. Nature, 352, 156–159.

[39] Ohzawa, I., DeAngelis, G. C., and Freeman, R. D. (1996) Encoding of binocular disparity by simple cells in the cat's visual cortex. Journal of Neurophysiology, 75, 1779–1805.

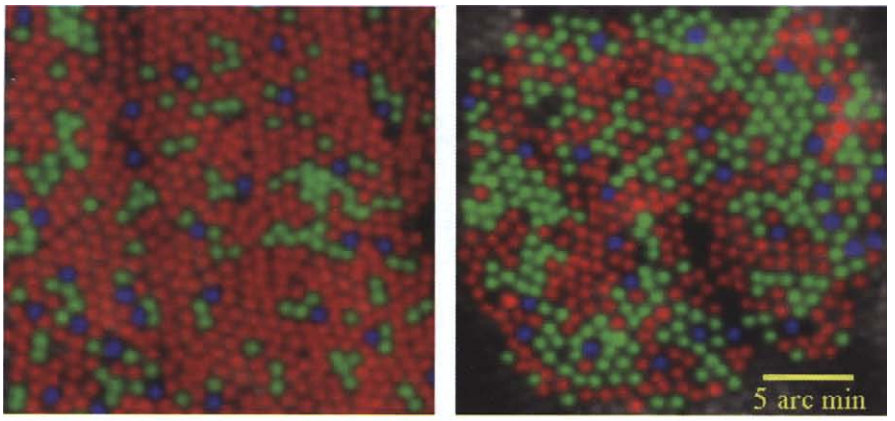
[40] Fleet, D. J., Wagner, H., and Heeger, D. J. (1996) Neural encoding of binocular disparity: Energy models, position shifts, and phase shifts. Vision Research, 36, 1839–1858.

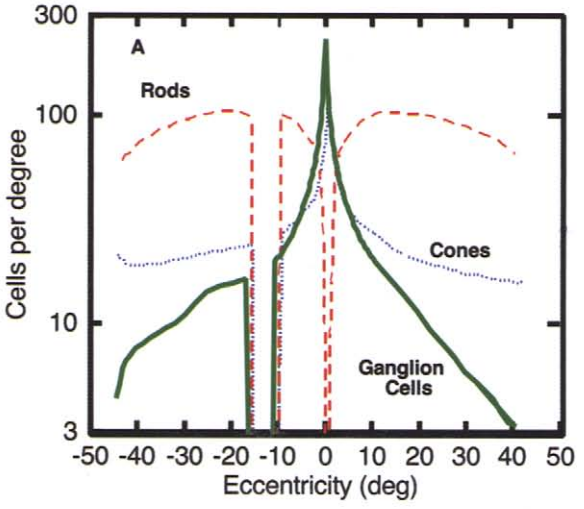
[41] Ohzawa, I., DeAngelis, G. C., and Freeman, R. D. (1997) Encoding of binocular disparity by complex cells in the cat's visual cortex. Journal of Neurophysiology, 76, 2879–2909.

[42] DeAngelis, G. C., Cumming, B. G., and Newsome, W. T. (1998) Cortical area MT and the perception of stereoscopic depth. Nature, 394, 677–680.

[43] Stevenson, S. B., Cormack, L. K., Schor, C. M., and Tyler, C. W. (1992) Disparity-tuned mechanisms of human stereopsis. Vision Research, 32, 1685–1689.

[44] Rodiek, R. W. (1998) The First Steps in Seeing. Sinauer Associates, Inc.: Sunderland.





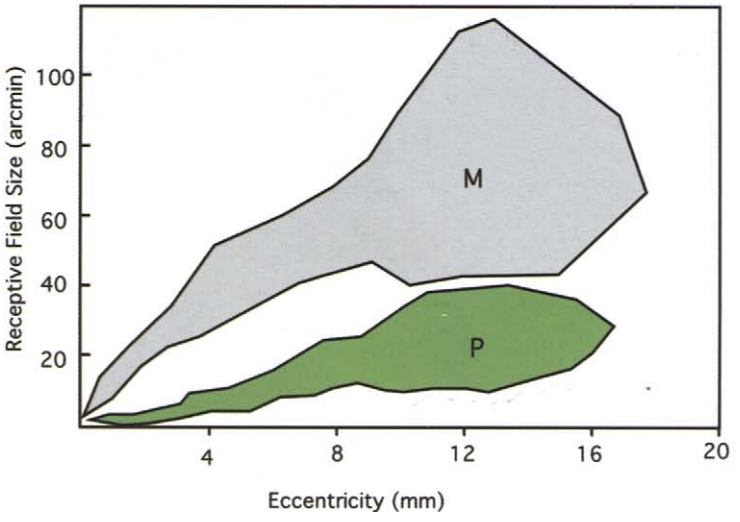


FIGURE 4.1.3 The upper panel shows the retinal sampling grid near the center of the visual field of two living human eyeballs. The different cone types are color-coded (from Roorda and Williams, 1999, reprinted with permission). The middle panel shows the density of various cell types in the human retina. The rods and cones are the photoreceptors that do the actual sampling in dim and bright light, respectively. The ganglion cells pool the photoreceptor responses and transmit information out of the eyeball (from Geisler and Banks, 1995). The lower panel shows the dendritic field size (assumed to be roughly equal to the receptive field size) of the two main types of ganglion cell in the human retina (redrawn from Dacy, 1993). The gray shaded region shows the parasol (or M) cells, and the green region shows the midget (or P) cells. The two cell types seem to independently and completely tile the visual world. The functional properties of the two cell types are summarized in Table 1.

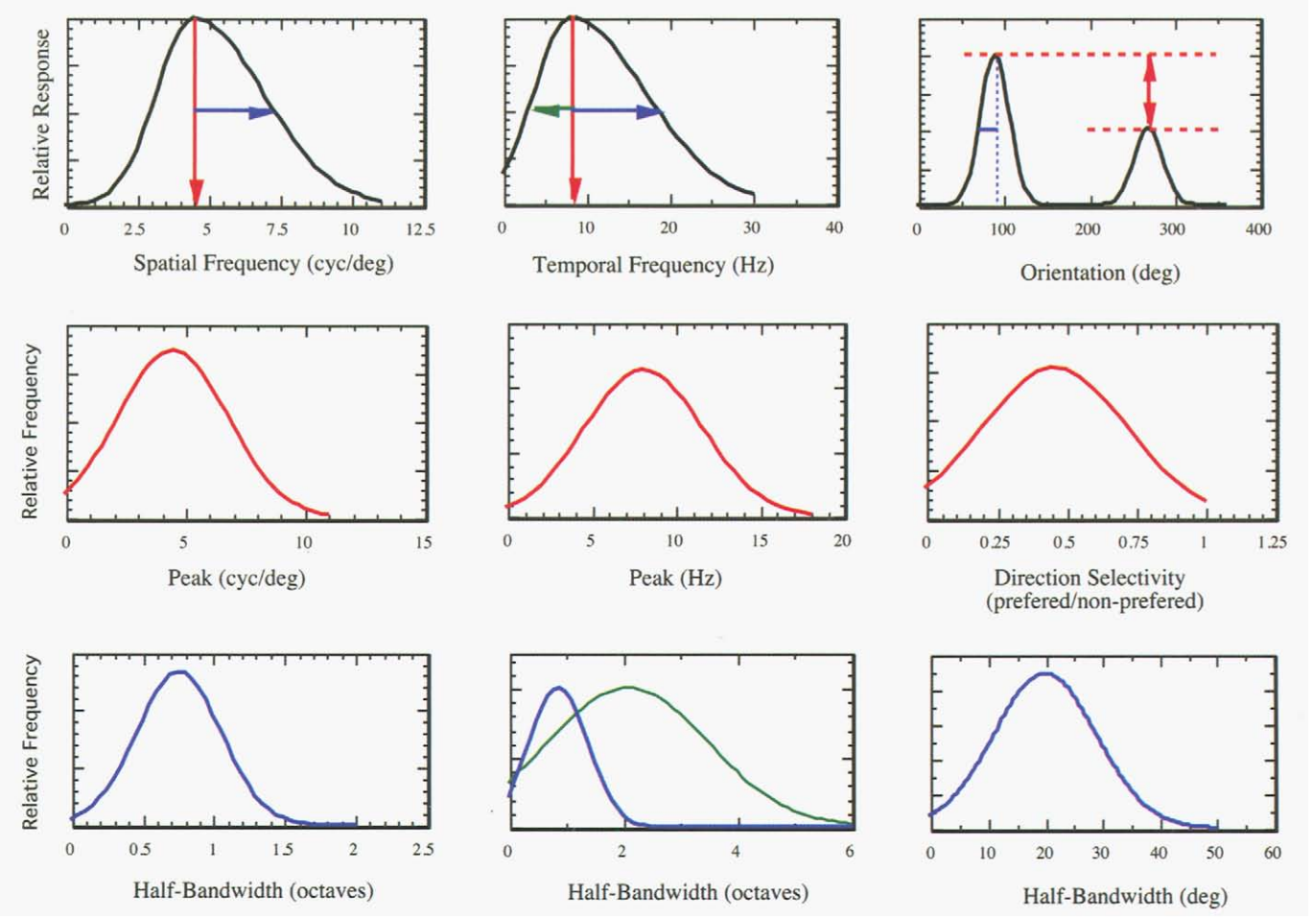


FIGURE 4.1.6 Left column: the upper panel shows a spatial frequency tuning profile typical of cell such as shown in Fig. 5. The middle and lower panels show distribution estimates of the two parameters of peak sensitivity (middle) and half-bandwidth in octaves (lower) for cells in macaque visual cortex. Middle column: same as the left column, but showing the temporal frequency response. As the response is asymmetric in octave bandwidth, the lower figure shows separate distributions for the upper and lower half-bandwidths (blue and green, respectively). Right column: the upper panel shows the response of a typical cortical cell to the orientation of a drifting sinusoidal grating. The estimate of half-bandwidth for macaque cortical cells is shown in the middle panel. The ratio of responses between the optimal direction and its reciprocal is taken as an index of directional selectivity; the estimated distribution of this ratio is plotted in the lower panel (the index cannot exceed unity by definition).

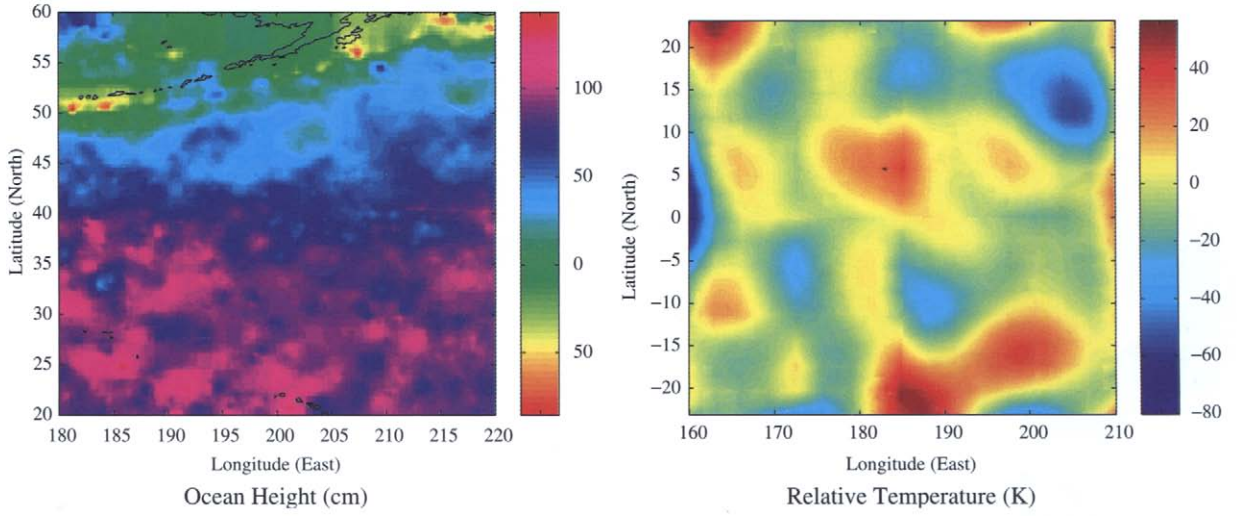


FIGURE 4.3.13 Multiscale estimation of remotely sensed fields. Left: North-Pacific altimetry based on Topex/Poseidon data. Right: Equatorial-Pacific temperature estimates based on in-situ ship data.