7

Pattern Sensitivity

I n this chapter we will consider measurements and models of human visual sensitivity to spatial and temporal patterns. We have covered topics relevant to pattern sensitivity in earlier chapters as we reviewed image formation and the receptor mosaic. Here, we will extend our analysis by reviewing a collection of behavioral studies designed to reveal how the visual system as a whole detects and discriminates spatiotemporal patterns.

The spatial-pattern vision literature is dominated by detection and discrimination experiments, not by experiments on what things look like. There are probably two reasons why these measurements make up such a large part of the literature. First, many visual technologies (e.g., televisions, printers, etc.) are capable of reproducing images that appear similar to the original, but not exactly the same. The question of which approximation to the original is visually best is important and often guides the engineering development of the device. As a result, there is considerable interest in developing a complete theory to predict when two different images will appear very similar. When we cannot reproduce the image exactly, a theory of discrimination helps the device designer; the theory identifies which of the images that the device can reproduce will appear most similar to the original. Threshold and discrimination experiments are indispensable to the design of discrimination theories.

Second, many authors believe that threshold and discrimination tasks can play a special role in analyzing the neurophysiological mech-

anisms of vision. The rationale for using threshold and discrimination to analyze the physiological mechanisms of vision is rarely stated and thus rarely debated, but the argument can be put something like this. Suppose the nervous system is built from a set of components, or mechanisms, that analyze the spatial pattern of light on the retina. Then we should identify and analyze these putative mechanisms to understand how they contribute to perception of spatial patterns. Threshold performance offers us the best chance of isolating the mechanisms because, at threshold, only the most sensitive mechanisms contribute to visibility. If threshold performance depends upon the stimulation of a single mechanism or small number of mechanisms, then threshold studies can serve as a psychologist's dissecting instrument: at threshold we can isolate different parts of the visual pathways. After understanding the component mechanisms, we can seek a unified theory of the visual system's operation.

I am not sure whether this rationale in terms of visual mechanisms adequately justifies the startling emphasis on threshold measurements. But it is plain that by now we have learned a lot from detection and discrimination experiments (DeValois and DeValois, 1988; Graham, 1989). Many of the basic ideas used in image representation and computer vision were derived from the work on detection and discrimination of spatial patterns. This chapter is devoted to an exposition of some of those experiments and ideas. In Chapters 9, 10, and 11 we will take up the question of what things look like.

A Single-Resolution Theory

The problem of predicting human sensitivity to spatial contrast patterns has much in common with other problems we have studied: image formation, color-matching, or single-unit neurophysiological responses. We want to make a small number of measurements, say, sensitivity to a small number of spatial contrast patterns, and then use these measurements to predict sensitivity to all other spatial contrast patterns.

In 1956, Otto Schade had some ideas about how to make these predictions, and he set out to build a photoelectric analog of the visual system in order to predict visual sensitivity. His idea was to use the device to predict whether small changes in the design parameters of a display would be noticeable to a human observer. For example, in designing a new display monitor an engineer might want to alter the spatial resolution of the device. To answer whether this difference is noticeable to a human observer, Schade needed a way to predict the sensitivity of human observers to the effect of the engineering change.

Figure 7.1 shows a schematic diagram of Schade's computational eye. I include the diagram to show that Schade created a very extensive
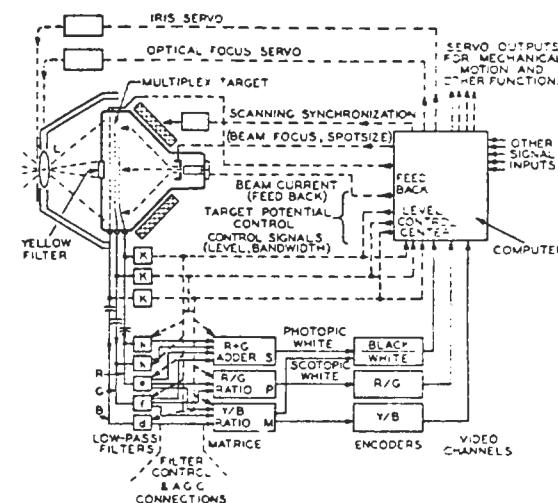
7.1 A COMPUTATIONAL MODEL OF THE HUMAN VISUAL SYSTEM. Otto Schade designed a visual simulator to predict human visu al sensitivity to patterns. His model incorporated many features of the visual pathway, and it may be the ear iest computational model of human vision. From Schade, 1 956.

model that incorporated many visual functions, such as optical image formation, transduction, integration of signals from the different cone types, adaptation, spatial integration, negative feedback, and even some thoughts about correction, interpretation, and correlation of the signals with stored information (i.e., memory). In this sense, his work was a precursor to the computational vision models set forth by Marr (1982) and his colleagues.

The Neural Image

A fundamental part of Schade's computation—and one that has been retained by the field of spatial-pattern vision—is his suggestion that we can summarize the effects of the many visual components using a single representation now called the **neural image**.[1] A real image and a couple of neural images are drawn suggestively in Figure 7.2. The idea is that there is a collection of neurons whose responses, taken as a population, capture the image information available to the observer. The responses of a collection of neurons with similar receptive fields, differing only in that the receptive fields are centered at various positions, make up a neural image.

Figure 7.2 shows how the neural-image concept permits us to visualize the neural response to an input image. At several places within

---

[1] To the best of my knowledge, John Robson first suggested this term.
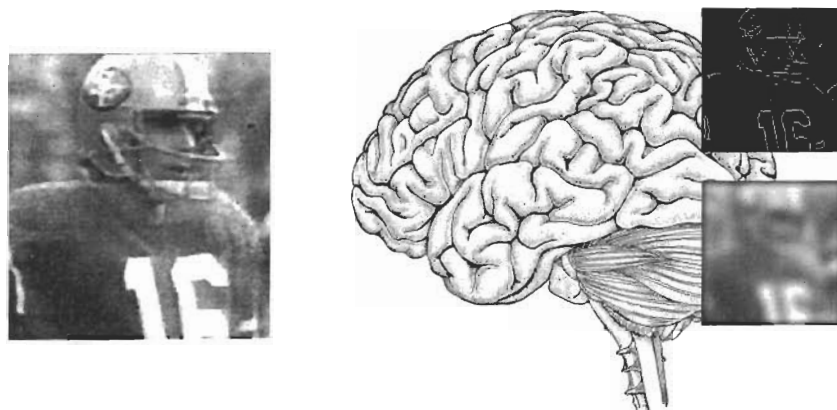
7.2 A NEURAL IMAGE represents the activity of a group of neurons. A real image, shown on the left, will be encoded by many distinct populations of neurons within the brain. The intensity at each point in a neural image, illustrated by the images on the right, represents the activity of a single neuron in the population. Each population of neurons may encode a different property of the original image. One of the neural images represents the edge locations. The other image represents an encoding of the local mean intensity, such as might be used to regulate adaptation. Other neural images, not shown, might represent aspects of color, motion, and depth.

the figure, I have represented the neural responses as an image. The intensity at each point in these neural images represents the response of a neuron single whose receptive field is centered at the corresponding image point. A bright value represents a neuron whose response is increased by the stimulus, and a dark value represents a neuron whose response is decreased.

The assumptions we make concerning the receptive-field properties of neurons comprising the neural image permit us to calculate the neural image using linear methods. For example, suppose the receptive fields of a collection of neurons are identical except for the position of the receptive-field centers; further, suppose these are uniformly spaced. In that case, we can calculate the mapping from the real image to the responses of these neurons using a shift-invariant linear mapping (i.e., convolution).

The neural images shown in Figure 7.2 illustrate the idea that different populations of neurons may represent different types of information. One neural image is shown representing the coarse aspects of the original. The values in this image were obtained by forming a local average of the original image intensities. The average local intensity represented in this image might be used for various purposes, such as regulating visual adaptation to illumination variation in the image. The

second neural image in the figure represents positions where there are likely to be edges in the original image. This image may be used as part of an object recognition system. There may be many other neural images in the cortex as well that represent different aspects of the image including color, motion, and depth. Of course, the information represented in a cortical neural image depends on transformations of the signal that take place all along the visual pathway, including lens defocus, sampling by the photoreceptor mosaic, wavelength encoding by the cones, and so forth.

Schade's Single-Resolution Theory

Schade's theory of pattern sensitivity is formulated mainly for foveal vision. Schade assumed that foveal pattern sensitivity could be predicted by the information available in a single neural image. He assumed that, for this portion of the visual field, the relevant neural image could be represented by a shift-invariant transformation of the retinal image, much like the neural image shown near the bottom of the visual cortex in Figure 7.2. In this section we review the significance of this hypothesis and also some empirical tests of it.

We know that a neural image spanning the entire visual field cannot really be shift-invariant. In earlier chapters, we reviewed measurements showing that the fovea contains many more photoreceptors and retinal ganglion cells than the periphery, and also that there is much more cortical area devoted to the fovea than to the periphery. Consequently, a neural image can have a shift-invariant representation only over a relatively small portion of the visual field, say, within the fovea or a small patch of the peripheral visual field.

Still, a model of pattern discrimination in the fovea is a good place to begin. First, the theory will be much simpler because we can avoid the complexities of visual-field inhomogeneities. Second, because of our continual use of eye movements in normal viewing, the fovea is our main source of pattern information. So, we begin by reviewing theory and measurements of pattern sensitivity in the fovea. We will consider how acuity varies across the visual field later in this chapter.

There are several ways the shift-invariant neural-image hypothesis helps us predict contrast sensitivity. Perhaps the most important is an idea we have seen several times before: if the mapping from image to neural image is shift-invariant, then the mapping from image to neural image is defined by knowing the shape of a single receptive field. In a shift-invariant neural image there is only one basic receptive-field shape. Neurons that make up the neural image differ only with respect to their receptive-field positions.

The analogy between shift-invariant calculations and neural receptive fields is useful. But we should remember that we are reasoning about behavioral measurements, not real neural receptive fields. Hence,

it is useful to phrase our measurements using the slightly more abstract language of linear computation. In these calculations, the linear receptive field is equivalent to the **convolution kernel** of the shift-invariant mapping. The shift-invariance hypothesis tells us that to understand the neural image, we must estimate the convolution kernel. Its properties determine which information is represented by the neural image and which information is not.

As we have done several times earlier in this book, we will begin our analysis using one-dimensional stimuli: vertical sinusoids varying only in the x-direction. If we use only one-dimensional stimuli as inputs, then we can estimate only the one-dimensional receptive field of the transformation. We can write the shift-invariant transformation that maps the one-dimensional contrast stimulus, a_x, to the one-dimensional neural image, n_x, using the summation formula,

$$n_x = \sum_{\langle y \rangle} l_{x-y} a_y \qquad (7.1)$$

where l_x is the one-dimensional receptive field.[2] We can also express the transformation in matrix tableau:

$$\begin{pmatrix} n_0 \\ n_1 \\ \vdots \\ n_{N-1} \\ n_N \end{pmatrix} = \begin{pmatrix} l_0 & \cdots & l_{-N/2} & \cdots & l_{-N} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ l_{N-2} & \cdots & l_0 & \cdots & l_{-N/2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ l_N & \cdots & l_{N/2} & \cdots & l_0 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{N-1} \\ a_N \end{pmatrix}$$

In matrix tableau it becomes clear that the system matrix is very simple; the rows and columns are all equal to the receptive field (i.e., convolution kernel) except for a shift or a reversal. Hence, by estimating the convolution kernel, we will be able to predict the transformation from contrast image to neural image.

The overall plan for predicting an observer's pattern sensitivity is as follows. First, we will measure sensitivity to a collection of harmonic contrast patterns. These measurements will define the observer's contrast sensitivity function (see Chapters 5 and 6). Because of the special relationship between harmonic functions and shift-invariant linear systems described in the earlier chapters and Appendix A, we can use the contrast-sensitivity function to estimate the convolution kernel of the shift-invariant linear transformation from image to neural image, l_x. Finally, we will use the estimated kernel to calculate the neural image and predict the observer's sensitivity to other one-dimensional contrast patterns. This final step will provide a test of the theory.

[2] See Appendix A for a discussion of the circular convolution formula and the meaning of the bracket notation, $\langle y \rangle$.

Shortly, it will become clear that we must make a few additional assumptions before we can use the observer's contrast-sensitivity measurements to estimate the convolution kernel. But first I will review some measurements of the human spatial contrast-sensitivity function.
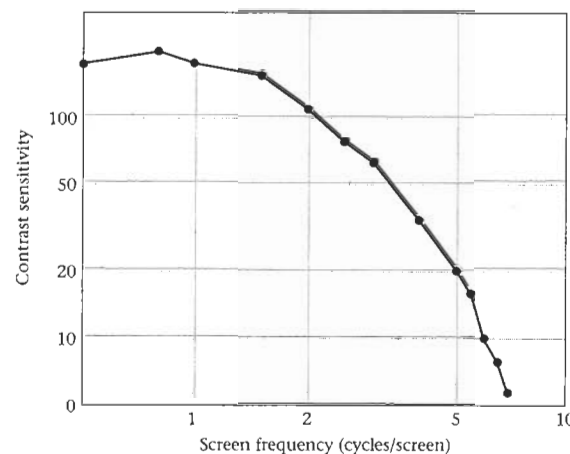
Spatial Contrast-Sensitivity Functions

Schade measured the contrast-sensitivity function by asking observers to judge the visibility of sinusoidal patterns of varying contrast. The observer's task was to decide what contrast was necessary to render the pattern just barely visible. Because of optical and neural factors, observers are not equally sensitive to all spatial-frequency patterns; the threshold contrast depends upon the pattern's spatial frequency.

To get a sense of the informal nature of Schade's (1956) experiments, it is interesting to read his description of the methods (p. 739):

> The test pattern is faded in by increasing the electrical modulation at a fixed rate and observed on the modulation meter; the observer under test gives a signal at the instant he recognizes the line test pattern, and the person conducting the test reads and remembers the corresponding modulation reading. The modulation is returned to zero, and within seconds it is increased again at the same fixed rate to make a new observation. By averaging 10 to 15 readings mentally and recording the average reading directly on graph paper, the [contrast sensitivity] function . . . can be observed in a short time and inconsistencies are discovered immediately and checked by additional observations.

The contrast-sensitivity function he measured is shown in Figure 7.3. The horizontal axis is spatial frequency as measured in terms of the display device. The vertical axis is contrast sensitivity, namely

7.3 CONTRAST-SENSITIVITY measurements of a human observer. The contrast thresholds are plotted with respect to spatial frequency on the display rather than cycles per degree of visual angle. Source: Schade, 1956.

$\log(1/c) = -\log c$, where c is the contrast of the pattern at the detection threshold. The contrast-sensitivity function has two striking features. First, there is a fall-off in sensitivity as the spatial frequency of the test pattern increases. This effect is large, but it should not surprise you since we already know many different components in the visual pathways are insensitive to high-spatial-frequency targets: the optical blurring of the lens reduces the contrast of high-spatial-frequency targets; also, retinal ganglion cells with center-surround receptive fields are less sensitive to high-spatial-frequency targets.

Second, and somewhat more surprisingly, there is no improvement of sensitivity at low spatial frequencies; there is even a small loss of contrast sensitivity at the lowest spatial frequency. The eye's optical image-formation does not reduce sensitivity at low frequencies, so the fall in contrast sensitivity at low spatial frequencies is due to neural factors. Center-surround receptive fields are one possible reason for this low-frequency fall-off.

Schade's measurements were made using a steadily presented test pattern or a drifting pattern. Robson (1966; see also Kelly, 1961) made additional measurements using flickering **contrast-reversing gratings.** Contrast-reversing patterns are harmonic spatial patterns with harmonic amplitude variation (see Chapter 5). For example, suppose the mean illumination is m. Then the *intensity* of the contrast-reversing stimulus at spatial frequency f_x, temporal frequency f_t, and contrast a is

$$[\ 1.0 + a\cos(2\pi f_t t)\cos(2\pi f_x x) \ ]m \qquad (7.2)$$

The intensity is always positive (cf. Equation 5.3). The spatiotemporal *contrast* of the pattern is

$$a\cos(2\pi f_t t)\cos(2\pi f_x x) \qquad (7.3)$$

The contrast can be both positive and negative.

As the data in Figure 7.4 show, the spatial sensitivity falls at low frequencies when we measure at a low temporal frequency (1 Hz). At high temporal frequencies, such as one might encounter during a series of rapid eye movements, there is no low-frequency sensitivity loss. As we shall see later, the contrast-sensitivity function also varies with other stimulus parameters such as the mean illumination level and the wavelength composition of the stimulus.

The Psychophysical Linespread Function

We will now return to the problem of using the contrast-sensitivity data to calculate the convolution kernel, l_x. Because this kernel defines both the rows and the columns of the shift-invariant linear transformation, it is also called the **psychophysical linespread function,** in analogy with
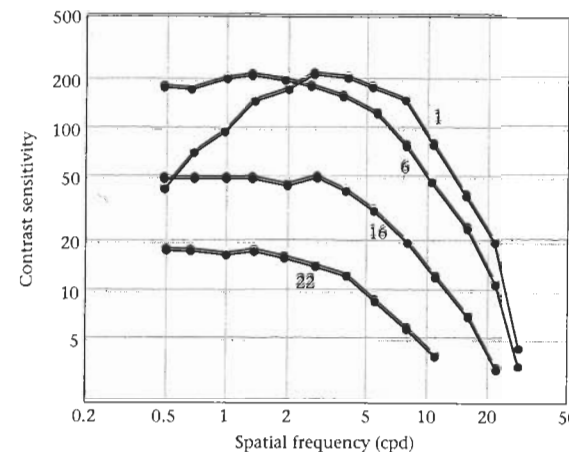
7.4 **TEMPORAL VARIATIONS CHANGE THE SHAPE OF THE HUMAN SPATIAL CONTRAST-SENSITIVITY FUNCTION.** The contrast-sensitivity functions shown here were measured with contrast-reversing targets at several different temporal frequencies. At low temporal frequencies (1 Hz) the contrast-sensitivity function is bandpass. At high temporal frequencies (22 Hz) the function is lowpass. Source: Robson, 1966.

the optical linespread function (see Chapter 2). By now you have probably noticed that each time we apply linear-systems theory, some special feature of the measurement situation requires us to devise some slightly different approach to calculating the system properties. The calculations involved in using contrast-sensitivity measurements to predict sensitivity to all contrast patterns are no exception.

We should now determine what we need to do to estimate the psychophysical linespread function. The general linear problem was illustrated in the matrix tableau representation near Equation 7.1, where the input stimulus is shown as a column vector, specifying the one-dimensional spatial contrast pattern. The matrix describes how the stimulus is transformed into the neural image. We want to make a small number of measurements in order to estimate the entries of the system matrix.

We have solved this problem before, but in this case we have a special challenge. In our previous attempts to estimate linear transformations we have been able to specify both the stimulus and the response. When we obtain the contrast-sensitivity measurements, however, we never measure the output neural image. We only measure the input threshold stimulus. Hence, we have fairly limited information available.

Because we assumed that the neural image is a shift-invariant mapping, we do know something about the neural image: when the input stimulus is a harmonic function, the output must be a harmonic function at the same frequency. But we do not know the amplitude or phase of the harmonic function in the hypothetical neural image. To estimate the psychophysical linespread function, we must make additional assumptions about the properties of a neural image that render it at

detection threshold. From these assumptions, we will then specify the phase and amplitude of the neural image at detection threshold.

Two additional assumptions are commonly made. First, we assume that the spatial phase of the neural image is the same as the spatial phase of the input spatial contrast pattern.[3] Specifically, when the input pattern is a one-dimensional cosinusoid, $\cos(2\pi f i/N)$, we assume that the neural image output pattern is a scaled copy of the input pattern, $n_i = a_f \cos(2\pi f i/N)$. The scale factor, a_f, depends on the frequency of the input signal.

Second, we must specify the amplitude of the neural image at detection threshold. The amplitude of the neural image should be related to the visibility of the pattern, and we can list a few properties that should be associated with pattern visibility. For example, whether the change introduced by the signal increases or decreases the firing rate should be irrelevant; any change from the spontaneous rate ought to be detectable. Also, detectability should depend on responses pooled across the neural image, rather than the response of a single neuron. The squared **vector-length** of the responses of the neural image is a measure that has both of these properties. The squared vector-length of the neural image, d^2, is defined by the formula

$$d^2 = \sum_{i=1}^{N} n_i^2 \qquad (7.4)$$

This formula satisfies both of our requirements since (a) the signs of the individual neural responses, n_i, are not important (because the neural image entry is squared), and (b) the formula incorporates the responses from different neurons. Other measures are possible. For example, one might assume that at detection threshold the sum of the absolute values of the neural image is equal to a constant, or one might make up a completely different rule. But, one must make some assumption, and the vector-length rule is a useful place to begin.

If we assume that at contrast detection threshold all neural images have the same vector-length, then we can specify the amplitude of the harmonic functions in the neural image. Hence, at this point we have made enough assumptions so that we can specify the complete neural image and solve for the psychophysical linespread function. Figure 7.5 shows three linespread functions estimated using these assumptions. Figure 7.5A shows a psychophysical linespread computed from Schade's measurements (note that the spatial dimension is uncalibrated). Figure 7.5B and Figure 7.5C show psychophysical linespreads, plotted in terms of degrees of visual angle, derived from Robson's measurements using 1-Hz and 6-Hz contrast-reversing functions. No single linespread

[3] We used the same assumption to infer the properties of the lens in Chapter 2.

(A)

(B)

(C)
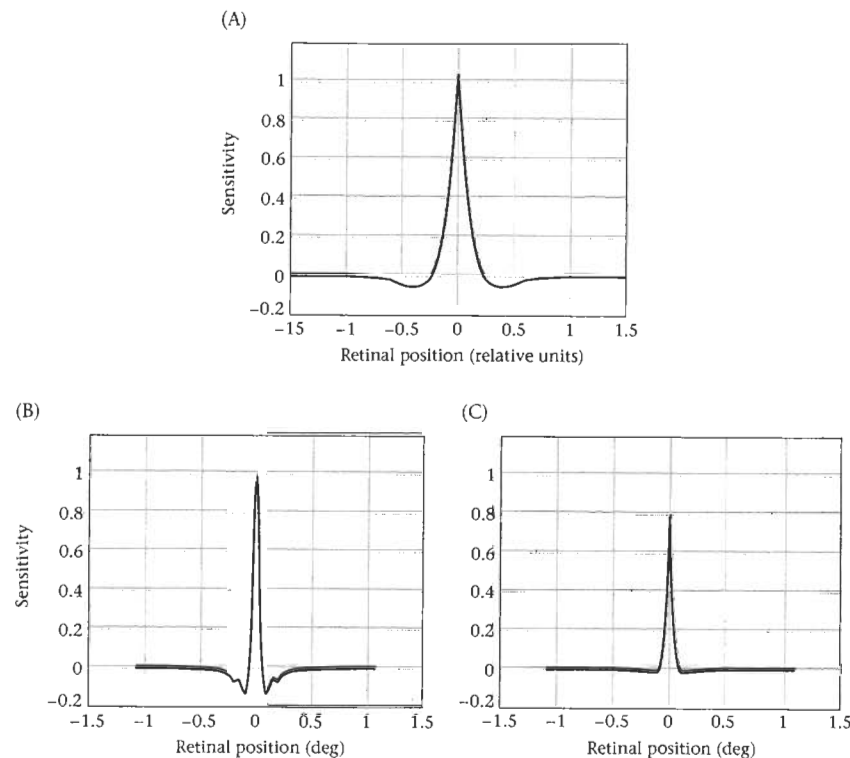
7.5 THE PSYCHOPHYSICAL LINESPREAD FUNCTION can be estimated from the contrast-sensitivity function. (A) A linespread estimated from Schade's measurements. The horizontal axis is in arbitrary units because the spatial frequency of the contrast-sensitivity function was reported in arbitrary units. Parts (B) and (C) show, respectively, linespread functions for contrast-reversing targets at 1 Hz and 6 Hz derived from Robson's measurements; the horizontal axis is in degrees of visual angle.

function applies to all stimulus conditions. We will consider how the linespread function changes with the stimulus conditions later in this chapter.

Schade (1956) suggested that the general shape of the psychophysical linespread function can be described using the difference of two Gaussian functions. This description is the same one used by Rodieck (1965) and Enroth-Cugell and Robson (1966) to model receptive fields of retinal ganglion cells. The correspondence between the psychophysical linespread function, derived from the behavioral measurement of contrast sensitivity, and the receptive-field functions of retinal ganglion cells, derived from retinal physiology, is encouraging.

Discussion of the Theory: Static Nonlinearities

To estimate the convolution kernel of Schade's hypothetical neural image, without being able to measure the neural image directly, we have been forced make several assumptions. It is wise to remember the three strong assumptions we have made:

1. The neural image is a shift-invariant linear encoding.

2. There is zero phase shift of the linear encoding.

3. The vector-length rule determines visibility.

Taken as a whole, this is a nonlinear theory of pattern sensitivity. Although the neural image is a linear representation of the input (indeed, it is even a shift-invariant representation), the vector-length rule linking the neural image to performance is nonlinear. You can verify this by noting that when a stimulus vector s_1 has length d_1 and vector s_2 has length d_2, the vector $s_1 + s_2$ need not have length $d_1 + d_2$. Thus, even when s_1 and s_2 are at one-half threshold, $s_1 + s_2$ may not be at threshold.[4]

The vector-length calculation is a static nonlinearity applied after a linear calculation (see Chapter 4). This is a relatively simple nonlinearity, so that it is straightforward to make certain general predictions about performance even though the theory is nonlinear. In the next section, we consider some of these predictions as well as experimental tests of them.

Experimental Tests

The contrast-sensitivity function by itself offers no test of Schade's theory other than reasonableness: do the inferred linespread functions seem plausible? We have seen that the linespread functions are plausible since they are quite similar to the receptive fields of visual neurons. But, because we have made so many assumptions, it is important to find general properties of the theory that we can test experimentally and in that way gain confidence in the theory's usefulness.

Harmonic functions will play a special role in testing the theory. There are two separate reasons why harmonics are important for our new test: (a) Given the assumed shift-invariance, the neural image of a harmonic is also a harmonic. We have seen this property many times before and it will be important again; (b) Harmonic functions at different frequencies are orthogonal to one another. Geometrically, orthogonality means that the vectors are oriented perpendicular to one another. Algebraically, we say two vectors a and b are orthogonal when

$0 = \sum a_x b_x$. Sinusoids and cosinusoids are orthogonal to one another, and any pair of harmonic functions at different frequencies are orthogonal to one another. We will use these two properties, combined with the vector-length rule, to test Schade's basic theory.

Suppose we create a stimulus equal to the sum of two sinusoids at frequencies, f_i, and contrasts, c_i, for $i = 1, 2$ (Figure 7.6). According to the shift-invariant theory, the neural image of these two sinusoids is the weighted sum of two sinusoids. Each sinusoid is scaled by a factor, s_i,
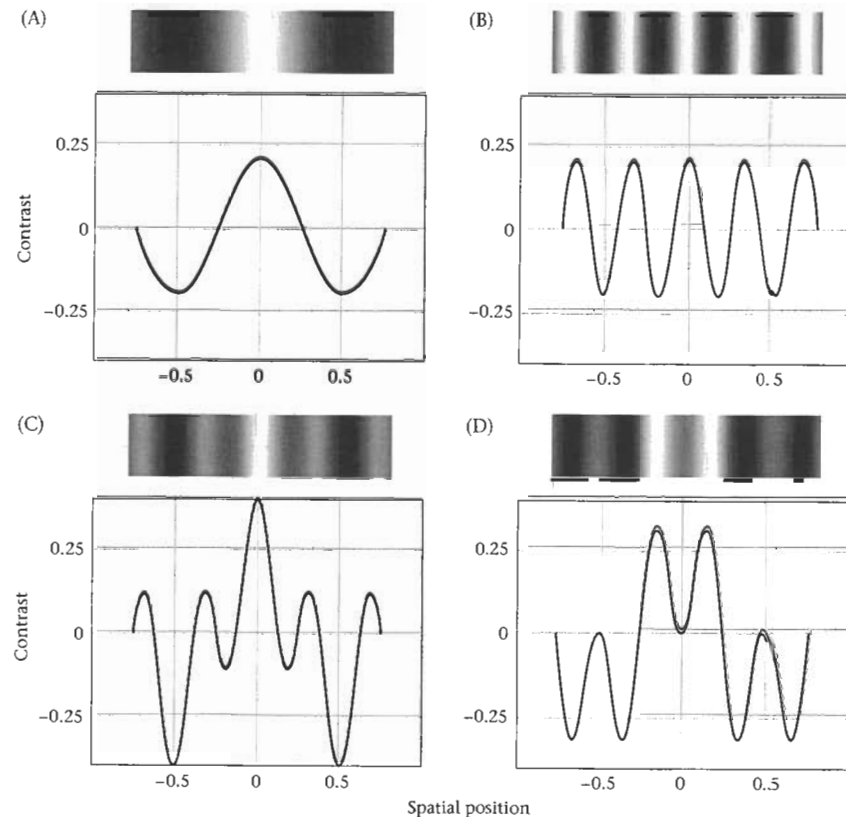
7.6 **MIXTURES OF SPATIAL CONTRAST PATTERNS.** Panels (A) and (B) show the contrast of cosinusoidal stimuli at 1 and 3 cpd. Mixtures of these two stimuli are shown in a spatial phase in which the peaks add (C) and a spatial phase in which the peaks subtract (D). The general appearance of the stimuli is indicated by the images shown above each panel.

---

[4] The only case in which the lengths will sum to $d_1 + d_2$ is when the vectors representing the neural images point in the same direction.

that defines how well the stimulus is passed by the shift-invariant system. The squared vector-length of the neural image created by the sum of the two sinusoids is

$$d^2 = \sum_{i=1}^{N} [c_1 s_1 \sin(2\pi f_1 i/N) + c_2 s_2 \sin(2\pi f_2 i/N)]^2 \qquad (7.5)$$

The squared term in the summation can be expanded into three terms

$$d^2 = [c_1 s_1 \sum_{i=1}^{N} \sin(2\pi f_1 i/N)]^2 + [c_2 s_2 \sum_{i=1}^{N} \sin(2\pi f_2 i/N)]^2$$

$$+ 2c_1 c_2 s_1 s_2 \sum_{i=1}^{N} \sin(2\pi f_1 i/N) \sin(2\pi f_2 i/N) \qquad (7.6)$$

Because sinusoids at different frequencies are orthogonal functions, the third term is zero, leaving only

$$d^2 = [c_1 s_1 \sum_{i=1}^{N} \sin(2\pi f_1 i/N)]^2 + [c_2 s_2 \sum_{i=1}^{N} \sin(2\pi f_2 i/N)]^2 \quad (7.7)$$

We can group some terms to define a new equation,

$$d^2 = (c_1 a_1)^2 + (c_2 a_2)^2 \qquad (7.8)$$

where a_i is a constant, namely, $s_i \sum_{i=j}^{N} \sin(2\pi f_i j/N)^2$.

Equation 7.8 tells us when a pair of contrasts of the two sinusoids, (c_1, c_2), should be at detection threshold. Figure 7.7 is a graphical representation of these predictions. The axes of the graph represent the contrast levels of the two sinusoidal components used in the mixture. The solutions to Equation 7.8 sweep out a curve called a **detection contour**.

7.7 THE SPATIAL TEST-MIXTURE EXPERIMENT provides a test of contrast-sensitivity models. We measure the visibility of a test-mixture whose sinusoidal components have contrasts c_1 and c_2. The set of contrast pairs such that the mixture stimulus is at detection threshold defines the detection contour. Schade's hypothesis predicts that the detection contour is an ellipse aligned with the axes of the graph, as shown by the solid curve. If the peak stimulus contrast determines contrast sensitivity to the pair, then the detection contour should fall along the contour indicated by the shaded lines.
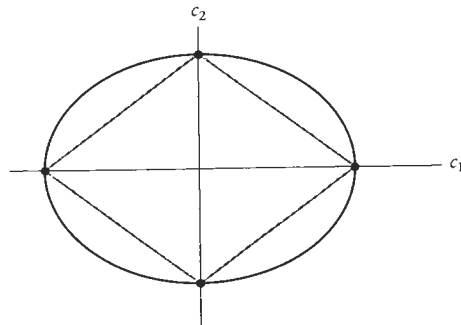
As shown in Figure 7.7, Equation 7.8 is the equation of an ellipse whose principal axes are aligned with the axes of the graph. The two unknown quantities, the scale factors a_i, are related to the lengths of the principal axes. Hence, if we scale the contrast of the sinusoidal components so that threshold contrast for each sinusoidal component is arbitrarily set to 1, the predicted detection contour will fall on a circle (Graham and Nachmias, 1971; Nielsen and Wandell, 1988).

Many alternative theories are possible. For example, had we supposed that threshold is determined by the peak contrast of the pattern, then the detection contour would fall on the diamond shape shown in Figure 7.7. The important point is that the shape of the detection contour depends on the basic theory. The prediction using Schade's theory is clear, so that we can use the prediction to test the theory.

Graham et al. (1978; see also Graham and Nachmias, 1971) measured sensitivity to mixtures of sinusoidal gratings at 1 cpd and 3 cpd. They measured thresholds using a careful psychophysical threshold estimation procedure called a **two-interval, forced-choice design**. In this procedure each trial is divided into two temporal periods, usually indicated by a tone that defines the onset of the first temporal period, a second tone that defines the onset of the second temporal period, and a final tone that indicates the end of the trial. A test stimulus is presented during one of the two temporal intervals, and the observer must watch the display and decide which interval contained the test stimulus. When the contrast of the test pattern is very low, the observer is forced to guess, and so performance is at chance. When the contrast is very high, the observer will nearly always identify the correct temporal interval. Hence, as the test-pattern contrast increases, performance varies from 0.5 to 1.0. The threshold performance level is arbitrary, but for technical reasons described in their paper, Graham et al. (1978) defined threshold to be the contrast level at which the observer was correct with probability 0.81.

The test-mixture data in Figure 7.8 do not fall precisely along the circular detection contour predicted by Schade's single-resolution theory. Specifically, thresholds measured in the 45° direction tend to fall just outside the predicted detection contour; thresholds are a little too high compared to the prediction. The theory predicts that the threshold contrasts of the individual components should be reduced by a factor of 1.414, but thresholds are reduced by only a factor of 1.2. These data are typical for these types of experimental measurements.

Is this an important difference? The point of this theory is to measure sensitivity to a small number of spatial patterns and to use these measurements to predict sensitivity to all other spatial patterns. If we see failures when we measure sensitivity to a mixture of only two test patterns, we should be concerned: The theory must be precise enough to tolerate decomposition of an arbitrary pattern into a sum of many sinusoidal patterns and then predict sensitivity to the mixtures of the multiple components. If we already see failures with two components,

7.8 SPATIAL TEST-MIXTURE THRESHOLDS measured using a 1- and 3-cpd grating. The thresholds fall outside of the detection contour predicted by the shift-invariant hypothesis and vector-length rule. Source: Graham et al., 1978.
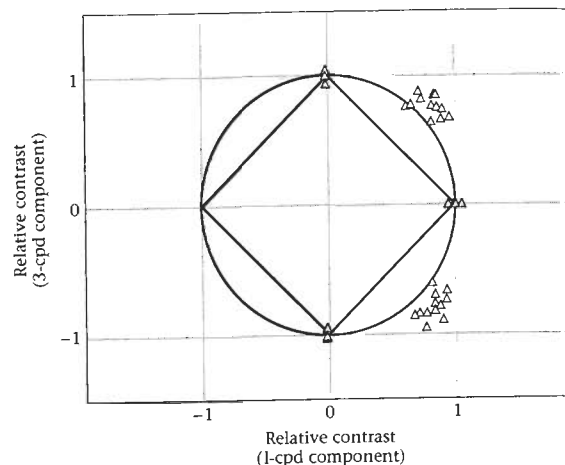
we should worry about how well the theory will do when we measure with three components.

The data in Figure 7.9 illustrate sensitivity measurements to the combination of three sinusoidal gratings. In this case, the data are plotted as three **psychometric functions**. Shown in this format, the dependence of performance on contrast is explicit. The data points connected by the dashed curves show the observer's probability of correctly detecting the individual sinusoidal grating patterns at 1.33, 4.0, or 12 cpd.

7.9 A THREE-COMPONENT SPATIAL TEST-MIXTURE EXPERIMENT. The probability of correct detection in a two-interval forced choice is shown as a function of normalized contrast. The shaded lines on the right show detection for three simple sinusoidal gratings at frequencies of 1.33, 4, and 12 cpd. The two solid lines show the probability of detecting mixtures of these three components in cosine phase and sine phase. The shaded curve on the left shows the predicted sensitivity using the shift-invariant model and the vector-length rule. Source: Graham, 1989.
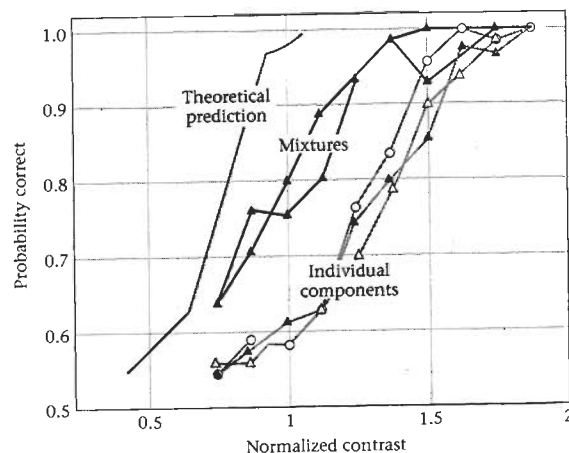
The horizontal axis measures the scaled contrast of the sinusoids in which the scale factor has been chosen to make the three curves align.

The visibility of two patterns formed by the mixtures of all three sinusoidal patterns, whose contrast ratios have been adjusted to make the three sinusoidal patterns equally visible, are shown as solid lines. One sum was formed with the peak contrasts all aligned (cosine phase) and the other with their zero-crossings aligned (sine phase).

Again, because the input signals are sinusoids or sums of sinusoids, we can predict performance based on the shift-invariant neural image and the vector-length rule. The neural image of the sum of three sinusoidal gratings will be the weighted sum of three sinusoidal neural images. The predicted threshold to the mixture of the three patterns is shown by the dot-dashed curve at the left of Figure 7.9. The vector-length rule also predicts that the probability correct will be the same in both sine and cosine phase.

The model prediction is not completely wrong; the phase relationship of the gratings does not have a significant influence on detection threshold for the mixture of three targets. But the mixture patterns are less visible than predicted by the theory: The contrasts of the three component mixtures are reduced by a factor of about 1.4 compared to their individual thresholds, while the theory predicts a contrast reduction of 1.73. The basic theory has some good features, but the quantitative predictions fail more and more as we apply the theory to increasingly complex patterns.

Intermediate Summary

We have begun formulating psychophysical theory using the simple computational ideas of shift-invariance followed by a static nonlinearity. These ideas are reminiscent of the properties of certain neurons in the visual pathway. While this theoretical formulation is a vast simplification of what we know about the nervous system, it is a reasonable place to begin. The nervous system is complex and contains many different types of computational elements. While Schade's effort to capture all of the nuances of the neural representation is inspiring, it was perhaps a bit premature. Much of the neural representation must be irrelevant to the tasks we are studying. Nonetheless, by beginning with simpler formulations, we can use psychophysical models to discover those aspects of the neural representation that are essential for predicting behavior. By comparing and contrasting the behavioral data and the neural data, we can discern the important functional elements of the neural representation for different types of visual tasks.

While the shift-invariant theory did not succeed, it has served the useful purpose of organizing some of our thinking and suggesting some experiments we should try. For the purpose of making some approximate predictions quickly, there are some good aspects of the calcula-

tion. For example, the inferred psychophysical linespread is similar to the receptive fields of some peripheral neurons. Also, for simple mixtures, the theoretical predictions are only off by a modest factor. Still, the shift-invariant theory plainly does not fit the data very well, and its performance will only deteriorate when we apply it to complex stimuli, such as natural images. We need to find new insights and experiments that might suggest how to elaborate the theory.

Multiresolution Theory

Schade's single-resolution theory of pattern sensitivity does not predict the pattern-sensitivity data accurately. But the theory is not so far wrong that we should abandon it entirely. The question we consider now is how to generalize the single-resolution theory, keeping the good parts.

Modern theories generally use an initial linear encoding consisting of a collection of shift-invariant linear transformations, not just a single one. Each shift-invariant linear transformation has its own convolution kernel and hence forms its own neural image. We will refer to the data represented by the individual shift-invariant representations as a **component-image** of the full theory.

To specify the properties of the more general theory, we need to select convolution kernels associated with each of the shift-invariant linear transformations and the static nonlinearities that follow. Also, we need to specify how the outputs of the different component images are combined to form a single detection decision.[5]

Pattern Adaptation

The motivation for building a multiresolution theory comes from a collection of empirical observations, such as the one illustrated in Figure 7.10. That figure demonstrates a phenomenon called **pattern adaptation**. To see the illusion, first notice that the bars in the patterns on top and bottom of Figure 7.10B are the same width. Next, stare at the fixation target between the patterns in Figure 7.10A for a minute or so. These patterns are called the **adapting patterns**. When you stare, allow your eye to wander across the dot between patterns, but do not let your gaze wander too far. After you have spent a minute or so examining the adapting patterns, look at the patterns in Figure 7.10B again. Particularly at first, you will notice that the bars at the top and bottom of the middle pattern will appear to have different sizes. You can try the same experiment by fixating between the adapting patterns in Figure 7.10C for a minute or so. When you examine the bars in the middle, the top

[5] For reasons I will explain in the next section, the properties of the convolution kernels of the component images are usually selected so that the expanded theory is a multiresolution representation of the image, a term I will define shortly.
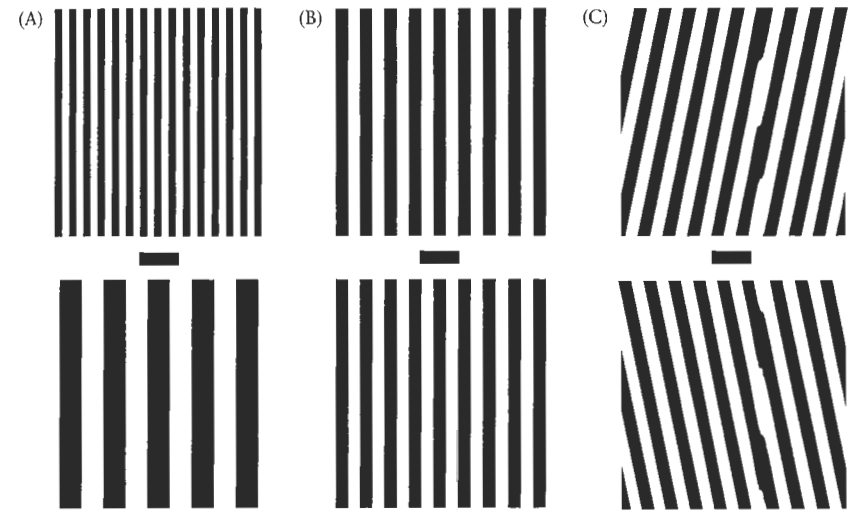
7.10 A SIZE ILLUSION AND AN ORIENTATION ILLUSION SHOWING VISUAL PATTERN ADAPTATION. The bar widths and orientations of the two square-wave patterns in the middle are the same. Stare at the fixation point between the two patterns in (A) for a minute, adapting to the two patterns in your upper and lower visual fields. When you shift your gaze to the patterns in (B) the patterns will appear to have different bar widths. Then, stare at the fixation point between the two patterns in (C) and then examine the middle pattern. When you shift your gaze to (B) the patterns will appear to have different orientations. After Blakemore and Sutton, 1969; see also DeValois, 1977a, b.

and bottom will appear to have different orientation (Gilinsky, 1968; Pantle and Sekuler, 1968; Blakemore and Campbell, 1969; Blakemore and Sutton, 1969; Blakemore, Nachmias and Sutton, 1970).

The effect of pattern adaptation can be measured by comparing the contrast-sensitivity function before and after adaptation. The curve through the filled circles in Figure 7.11A shows the contrast-sensitivity function prior to pattern adaptation. After adapting for several minutes to a sinusoidal contrast pattern, much as you adapted to the patterns in Figure 7.10A, the observer's contrast sensitivity to stimuli near the frequency of the adapting pattern is reduced, while contrast sensitivity to other spatial frequency patterns remains unchanged. The ratio of contrast sensitivity before and after adaptation is shown in Figure 7.11B. When this experiment is repeated, using adapting patterns at other spatial frequencies, contrast sensitivity falls for test patterns whose spatial frequency is similar to that of the adapting pattern (Blakemore and Campbell, 1969).

The results of the pattern-adaptation measurements suggest one way to generalize the neural image from a single-resolution theory to
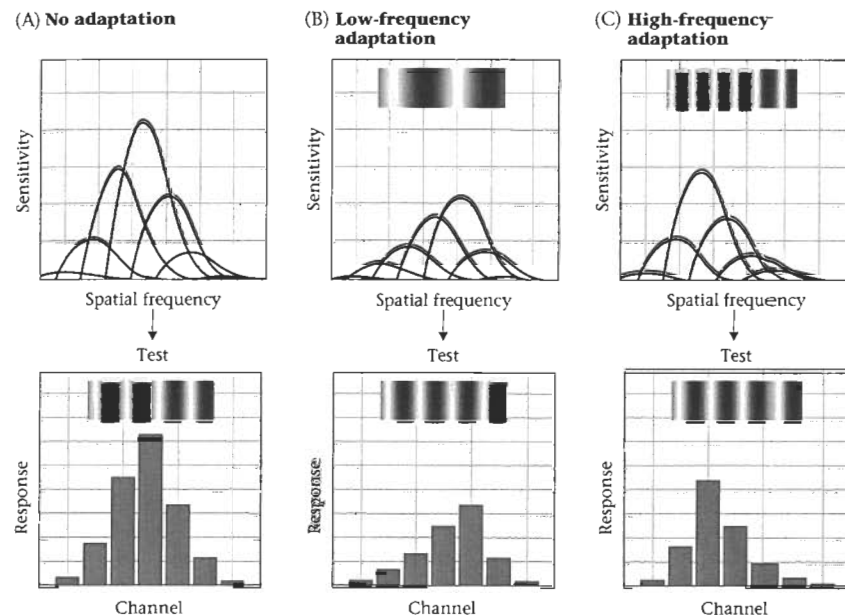
(A)

(B)
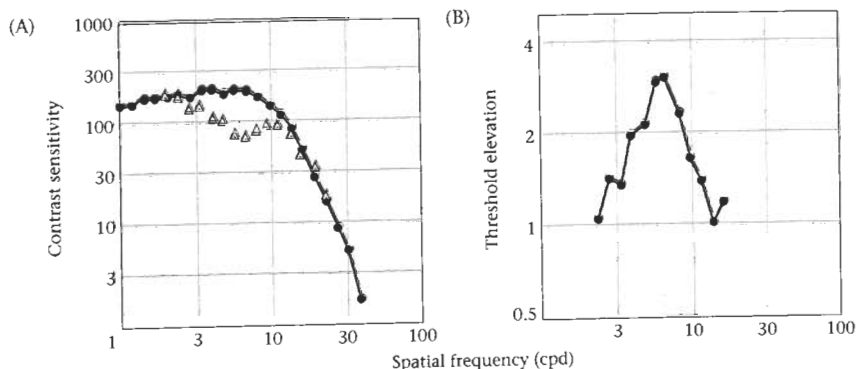
7.11 THE EFFECT OF PATTERN ADAPTATION ON THE CONTRAST-SENSITIVITY FUNCTION. (A) The curve through the filled circles shows the observer's contrast-sensitivity function before pattern adaptation. The open triangles show contrast sensitivity following adaptation to a sinusoidal pattern of 7.1 cpd. (B) Threshold elevation, that is, the ratio of contrast sensitivity before and after adaptation, is plotted as a function of spatial frequency. The threshold elevation occurs at test frequencies near the frequency of the adapting stimulus. Source: Blakemore and Campbell, 1969.

(A) **No adaptation**   (B) **Low-frequency adaptation**   (C) **High-frequency adaptation**

7.12 A MULTIRESOLUTION MODEL can explain certain aspects of pattern adaptation. (A) In normal viewing, the bar width is inferred from the relative responses of a collection of component-images, each responding best to a selected spatial frequency band. The spatial-frequency selectivity of each component image is shown above, and the amplitude of the component-image encoding of the test stimulus is shown in the bar graph below. (B) Following adaptation to a low-frequency stimulus (shown in inset), the sensitivity of the neurons comprising certain component images is reduced. Considering the responses of all the component images, the response to the test is similar to the unadapted response to a high-frequency target. (C) Following adaptation to a high-frequency pattern (shown in inset), the neural representation is consistent with the unadapted response to a low-frequency target.

a multiresolution theory: use a neural representation that consists of a collection of component-images, each sensitive to a narrow band of spatial frequencies and orientations. This separation of the visual image information can be achieved by using a variety of convolution kernels, each of which emphasizes a different spatial frequency range in the image. This calculation might be implemented in the nervous system by creating neurons with a variety of receptive-field properties, much as we have found in the variety of receptive fields of linear simple cells in the visual cortex; these cells have both orientation and spatial-frequency preferences (Chapter 6). Because the individual component images are assumed to represent different spatial frequency resolutions, we say that the neural image is a **multiresolution representation**.

Multiresolution representations provide a simple framework to explain pattern adaptation (Figure 7.12). Suppose the visual system ordinarily encodes the image using a collection of shift-invariant component images whose contrast-sensitivity curves are shown on the top of Figure 7.12A. Before adaptation, each of the component images represents the square wave at an amplitude that depends on the square-wave frequency and the channel sensitivity. The bar plot at the bottom of Figure 7.12A shows the amplitude of the component-image representations to the test pattern before adaptation.

Adaptation to a low-frequency square wave suppresses sensitivity of some of the component images, as shown in the top of Figure 7.12B. Consequently, the pattern of responses to the test frequency follow-

ing adaptation changes, as shown in the bottom of Figure 7.12B. The new pattern of responses is consistent with the responses that would be caused by the unadapted response to a finer square-wave pattern. This is the explanation of the observation that, following adaptation to a low-frequency square wave, the test pattern appears to shift to a higher spatial frequency. Figure 7.12C illustrates the component-image sensitivities following adaptation to a high-frequency square wave (top) and how the amplitude of the component image responses are altered (bottom). In this case, the pattern of responses is consistent with the unadapted encoding of a lower-frequency target.

According to the multiresolution model, pattern adaptation is much like a lesion experiment. Adaptation reduces or eliminates the contribution of one set of neurons, altering the balance of activity and

producing a change in the perceptual response. Following adaptation to a low-frequency target, the excitation in component images at higher spatial frequencies is relatively greater, giving the test bars a narrower appearance. Conversely, following adaptation to a high-frequency target, the excitation in component images representing low spatial frequencies is relatively greater, giving the test bars a wider appearance.

In summary, the empirical observations using pattern adaptation suggest that square-wave or sinusoidal adapting patterns only influence the contrast sensitivity of patterns of roughly the same spatial frequency. This observation suggests that the component images might be organized at multiple spatial resolutions.

Pattern Discrimination and Masking

There are several other experimental observations, in addition to pattern adaptation, that can be used to support multiresolution representations for human perception. Historically, one of the most important papers on this point was Campbell and Robson's (1968) study, in which they made detection and discrimination measurements using square-wave gratings and other periodic spatial patterns.

Square waves, like all periodic stimuli, can be expressed as the weighted sum of sinusoidal components using the discrete Fourier series. A square wave, $sq(x)$, that oscillates between $+1$ and -1 with a frequency of f, can be expressed in terms of sinusoidal components as

$$sq(x) = \frac{4}{\pi} \sum_{n=0}^{\infty} \frac{1}{2n+1} \sin(2\pi(2n+1)fx) \qquad (7.9)$$

A square wave at frequency f is equal to the sum of a series of sinusoids at the odd-numbered frequencies, f, $3f$, $5f$, and so forth. The amplitude of the sinusoids declines with increasing frequency; the amplitude of the $3f$ sinusoid is one-third the amplitude of the component at f, the amplitude of the $5f$ sinusoid is one-fifth, and so forth. When the overall contrast of the square wave is very low, the amplitude of the higher-order terms is extremely small and they can be ignored. At low square-wave contrasts, only one or two sinusoidal terms are necessary to generate a pattern that is very similar in appearance to the true square wave. For low-contrast values, then, the square-wave pattern can be well approximated by the pattern

$$sq(x) \approx \frac{4}{\pi} \left[\sin(2\pi fx) + \frac{1}{3} \sin(2\pi(3f)x) + \frac{1}{5} \sin(2\pi(5f)x) \right] (7.10)$$

Campbell and Robson used square waves (and other periodic patterns) to test the multiresolution hypothesis in several ways. First, they measured the smallest contrast level at which observers could detect the square-wave grating. Notice that the amplitude of the lowest-frequency
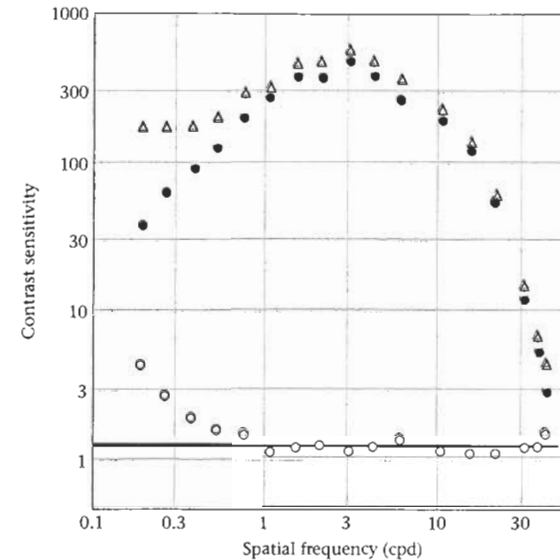
7.13 CONTRAST SENSITIVITY MEASURED USING SQUARE-WAVE GRATINGS greater than 1 cpd can be predicted from the contrast of the square-wave fundamental frequency. The open triangles and filled circles show contrast sensitivity to square waves and sine waves, respectively. The open circles show the ratio of contrast sensitivities at each spatial frequency. The solid line is drawn at a value of $4/\pi \approx 1.273$, the amplitude of a square-wave fundamental in a unit-contrast square wave. Source: Campbell and Robson, 1968.

component, which is called the **fundamental**, is $4/\pi$. Since the fundamental has the largest contrast, and since for patterns above 1 cpd sensitivity begins to decrease, Campbell and Robson argued that the neurons whose receptive-field size is well-matched to the fundamental component will signal the presence of the square wave first. If this is the most important term in defining the visibility of the square wave, then the threshold contrast of the square wave should be $4/\pi$ times the threshold contrast of a sinusoidal grating at the same frequency. The data in Figure 7.13 show contrast-sensitivity functions to both sinusoidal and square-wave targets, and the ratios of the contrast sensitivities. As predicted for patterns above 1 cpd, the ratio of contrasts at detection threshold is $1.28 \approx 4/\pi$.[6]

In addition to detection thresholds, Campbell and Robson also measured how well observers can discriminate between square waves and sinusoids. In these experiments, observers were presented with a square wave and a sine wave at frequency f. The two patterns were set in a contrast ratio of $4/\pi$, insuring that the fundamental component of the square wave and the sinusoid had equal contrast. The observers

---

[6] Campbell and Robson's square-wave detection experiment is a special case of the test-mixture experiment reviewed above. The results show that the $3f$ and higher-frequency components do not help the observer detect the square-wave grating. The more general question of how different frequency components combine is answered by test-mixture experiments, such as those performed by Graham and Nachmias (1971), in which the relative contrasts of all the components are varied freely.

adjusted the contrast of the two patterns, maintaining this fixed contrast ratio, until the square wave and sinusoid were barely discriminable. Since the contrast of the square-wave fundamental was held equal to the contrast of the sinusoid, the stimuli could only be discriminated based on the frequency components at $3f$ and higher.

Campbell and Robson found that observers discriminated between the sinusoid and the square wave when the contrast in the third harmonic reached its own threshold level. Their conclusions are based on the measurements shown in Figure 7.14. The filled circles show the contrast-sensitivity function. The open circles show the contrast of the square wave when it is just discriminable from the sinusoid. Evidently, the square-wave contrast needed to discriminate the two patterns exceeds the contrast needed to detect the square wave. However, we can explain the increased contrast by considering the contrast in the $3f$ component of the square wave. Recall that this component has one-third the contrast of the square wave. By shifting the square-wave discrimination data (open circles) to the left by a factor of 3 for spatial frequency, and downward by a factor of 3 for contrast, we compensate for these two factors. The open triangles show the open circles shifted in this way. The open triangles align with the original contrast-sensitivity measurements. From the alignment of the shifted discrimination data with the contrast-sensitivity measurements, we can conclude that the square wave can be discriminated from the sinusoid when the $3f$ component is visible at detection threshold.

7.14 DISCRIMINATION OF SINUSOIDAL AND SQUARE-WAVE GRATINGS becomes possible when the third harmonic in the square wave reaches its own independent threshold. The open circles plot the contrast-sensitivity function. The open triangles show the contrast level at which a square wave can be discriminated from its fundamental frequency. The filled squares show the square-wave discrimination data shifted by a factor of 3 in both frequency and contrast. The alignment of the shifted curve with the contrast-sensitivity function suggests that square waves are discriminated when the third harmonic reaches its own threshold level. Source: Campbell and Robson, 1968.
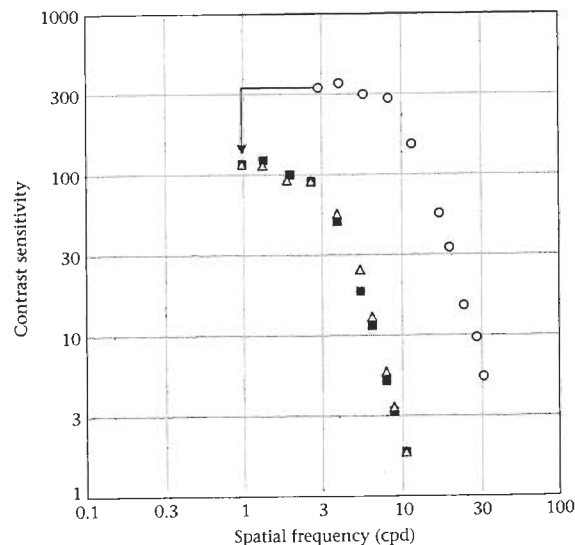
Masking and Facilitation

Campbell and Robson's discrimination results are consistent with a multiresolution representation of the pattern. It is as if the fundamental and third harmonics are encoded by different component images. Because the amplitude of the fundamental component is the same in the square wave and sinusoid, the observer cannot use that information to discriminate between them. When the contrast of the third harmonic exceeds its own independent threshold, the observer can use the information and discriminate between the two patterns.

Although multiresolution representations are consistent with this result, we should ask whether the evidence is powerful. Specifically, we should ask whether the data might be explained by simpler theories. One more general hypothesis we should consider is this: observers discriminate between two spatial patterns, S and $S + \Delta S$, whenever ΔS is at its own threshold. This is the phenomenon that Campbell and Robson report when S and ΔS are low-contrast stimuli, widely separated in spatial frequency. Can subjects always discriminate S from $S + \Delta S$ when ΔS is at its own threshold? The answer is no.

In fact, the case described by Campbell and Robson is very rare. In many cases the two patterns S and $S + \Delta S$ cannot be discriminated even though ΔS, seen alone, is plainly visible. In this case, we say the stimulus S masks the stimulus ΔS. There are also cases when S and $S + \Delta S$ can be discriminated even though ΔS, seen alone, cannot be detected. In this case, we say the stimulus S facilitates the detection of ΔS. **Masking** and **facilitation** are quite common; the absence of masking and facilitation, as in the data reported by Campbell and Robson, is fairly unusual.

The images in Figure 7.15A demonstrate the phenomenon of visual masking. The pattern shown on the left is the target contrast pattern, ΔS. This contrast pattern is added into one of the masking patterns shown in the middle column. The masking pattern on the top is similar to the target in orientation, but different by a factor of 3 in spatial frequency. If you look carefully, you will see a difference between the mask alone and the mask plus the target. Specifically, near the center of the pattern several of the bars on the left appear darkened, and several bars on the right appear lightened. The second mask is similar to the target in both orientation and spatial frequency. In this case, it is harder to see the added contrast, ΔS. The third mask is similar in spatial frequency, but different in orientation. In this case, it is easy to detect the added target.

Figure 7.15B shows measurements of masking and facilitation between patterns with similar spatial frequency and the same orientation (Legge and Foley, 1980). Observers discriminated a mask, S, from a mask plus a 2-cpd sinusoidal target, $S + \Delta S$. The vertical axis measures the threshold contrast of the target needed to make the discrimination, and the horizontal axis measures the contrast of the masking stimulus

(A) Test + Mask = Sum

7.15 MASKING AND FACILITATION. (A) These images illustrate visual masking. The test contrast pattern is shown on the left, and three different masking contrast patterns are shown in the middle column. The sum of the test and mask contrasts is shown in the right column. When the spatial frequencies of the test and mask differ by a factor of 3 (top), it is possible to see the effect of the test pattern. When the spatial frequency of the test and mask are similar (middle), it is difficult to perceive the added test. When the orientations of the test and mask are very different (bottom), it is very easy to see the added test. (B) The contrast needed to detect a 2-cpd target (ΔS, vertical axis) depends on the contrast of the masking pattern (S, horizontal axis). Each curve measures the effect of a different spatial-frequency pattern, S. When S is of low contrast and similar spatial frequency and orientation, it facilitates detection of the target; when it is of high contrast, it masks detection of the target. The curves have been displaced along the vertical axis so that each can be seen clearly. Source: Legge and Foley, 1980.

(B)
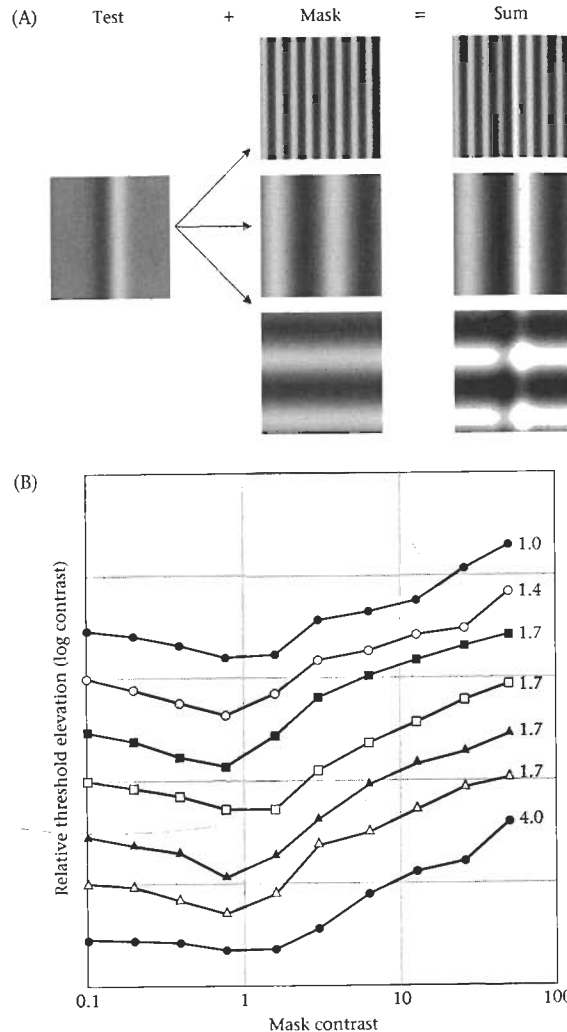
S. The different curves show results for masks of various spatial frequencies. In general, the presence of S facilitates detection at low contrasts and masks detection at high contrasts. When the spatial frequencies of ΔS and S differ by a factor of 2, the amount of facilitation is small, though there is still considerable masking. Other experimental measurements show that when the spatial frequencies of the test and mask dif-

fer by a factor of 3, the effect of masking is reduced (DeValois, 1977a, b; Wilson et al., 1983). We will discuss the effect of orientation on masking later in this chapter.

The implications of these experiments for multiresolution models can be summarized in two parts. First, Campbell and Robson's data show no facilitation or masking when S and ΔS are low-contrast and widely separated in spatial frequency. Second, the Legge and Foley data show that for many stimulus pairs more similar in spatial frequency, S influences the visibility of an increment ΔS. Taken together, these results are consistent with the idea that stimuli with widely different spatial frequencies are encoded by different component images.

The Conceptual Advantage of Multiresolution Theories

Today, many different disciplines represent images using the multiresolution format; that is, by separating the original data into a collection of component images that differ mainly in their peak spatial-frequency selectivity. The multiresolution representation has opened up a large set of research issues, and I will discuss several of these in Chapter 8. While the behavioral evidence for multiresolution is interesting, it is hardly enough to explain why multiresolution hypotheses have led to something of a revolution in vision science. Rather, I think that it is the conceptual advantages of multiresolution representations, described below, that have made them an important part of vision science.

To understand these conceptual advantages, consider two problems that arise when theorists abandon the simple shift-invariance hypothesis. First, the set of possible encoding functions, even just linear encoding functions, becomes enormous. How can one choose among all of the possible linear transformations? Second, without shift-invariance, the theorist loses considerable predictive power, since many important derivations we have made depend on shift-invariance. For example, without shift-invariance we cannot derive the same quantitative prediction concerning the test-mixture threshold of a pair of sinusoids (Equation 7.8).

Simply abandoning shift-invariance opens up the set of possible encodings too far; theorists need some method of organizing their choices among the set of possible linear encodings. The multiresolution structure helps to organize the theorist's choices, because it provides some guidelines about how to specify the properties of the collection of shift-invariant calculations. The multiresolution organization helps the theorist reason and describe the properties of the linear encoding.

Even though multiresolution theories offer a general organizational scheme for the linear encoding, they still permit considerable flexibility in designing the properties of the component images. This flexibility is used to create component images whose response properties are analogous to the receptive field properties of different populations of cortical

neurons. For example, orientation selectivity is an important property of many cortical neurons in area V1 (see Chapter 5). A simple shift-invariant theory, containing only a single convolution kernel, cannot represent orientation selectivity because if the convolution kernel is orientation selective, then the theory predicts that observer must be more sensitive to stimuli in that orientation. Since observers are not more sensitive to stimuli in one orientation than any other, oriented kernels are inappropriate for a single-resolution model. But it is possible to use orientation selective convolution kernels in multiresolution theories. As long as the collection of kernels at each resolution encodes all orientations, the model does not predict that sensitivity is better in any particular direction. For this reason, some theorists argue that there is a closer connection between the multiresolution theories and the cortical representation of pattern information.

The complexity of the calculations is an important challenge in developing multiresolution models of human pattern sensitivity. Figure 7.16 is an overview of a fairly simple multiresolution model de-
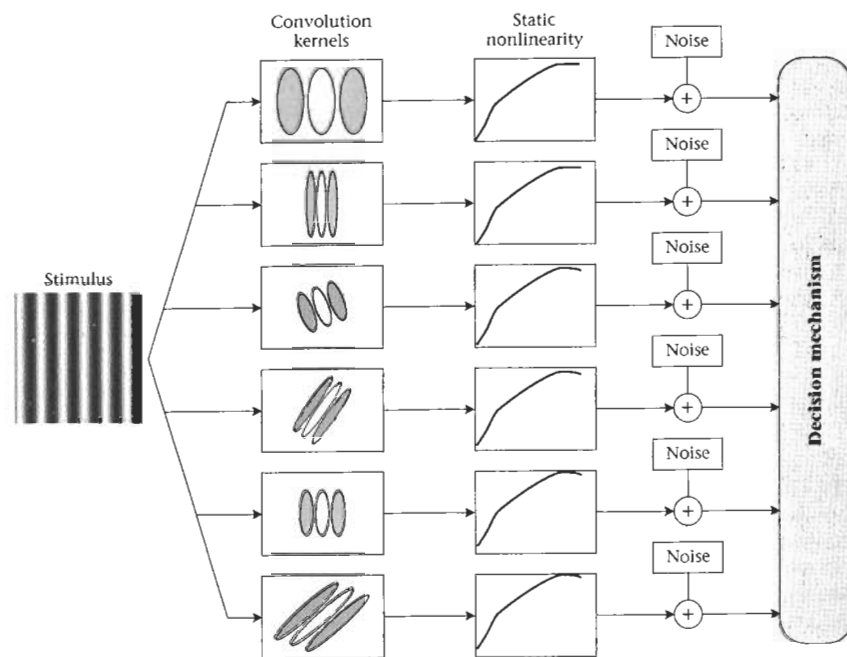
7.16   A MULTIRESOLUTION MODEL OF SPATIAL PATTERN SENSITIVITY. The stimulus is convolved with a collection of spatial filters with different peak spatial-frequency sensitivity. The filter outputs are modified by a nonlinear compression, noise is added, and the result is combined into a neural image. Source: Spillman and Werner, 1990.

scribed by Wilson and Regan (1984; see also Foley and Legge, 1981; Watson, 1983; Wilson and Gelb, 1984; Watt and Morgan, 1985). The initial image is transformed linearly into a neural image comprising a set of component images. In a recent implementation of this model, Wilson and Regan (1984) suggest that the neural image consists of 48 component images, organized by six spatial scales and eight orientations (i.e., all scales at all orientations). Each component image is followed by a static nonlinearity that is a modification of the vector-length measure. To see the development of an even more extensive model, the reader should consult the work by Watson and his colleagues (e.g., Watson, 1983; Watson and Ahumada, 1989). They have developed a substantially larger multiresolution model, using very sophisticated assumptions concerning the observer's internal noise and decision-making capabilities.

It is difficult to reason about the performance of these multiresolution models from first principles (but see Nielsen and Wandell, 1988; Bowne, 1990). Consequently, most of the predictions from these models are derived using computer simulation. Analyzing the model properties closely could easily fill up a book; in fact, Graham (1989) has completed an authoritative account of the present status of work in this area.

Challenges to Multiresolution Theory

Multiresolution theories are the main tool that theorists use to reason about pattern sensitivity. As we reviewed in the preceding section, multiresolution representations have many useful features, and they can be used to explain several important experimental results. There are, however, a number of empirical challenges to the multiresolution theories. In this section, I will describe a few of the measurements that represent a challenge to multiresolution theories of human pattern sensitivity. As you will see, many of these challenges derive from the same source as challenges to a shift-invariant theory: mixture experiments.

Pattern Adaptation to Mixtures

If we are to use pattern adaptation to justify multiresolution theories, then we should spend a little more time studying the general properties of pattern-adaptation measurements. Perhaps the first step we should take is to extend the pattern-adaptation measurements from simple sinusoids to more general patterns consisting of the mixture of two patterns.

Nachmias et al. (1973) performed a pattern-adaptation experiment using individual sinusoidal stimuli and their mixtures as adapting stimuli. The question they pose, as in all mixture experiments, is whether we can use the individual measurements to predict the behavioral performance in response to the sum of the stimuli.

The results of their measurements are shown in Figure 7.17. Each curve in the figure represents threshold elevation of sinusoidal test gratings at different spatial frequencies. The solid curve measures threshold elevation when the adapting stimulus was a 3-cpd sinusoidal grating. Confirming Blakemore and Campbell (1969), there is considerable threshold elevation at 3 cpd and less adaptation at both higher and lower spatial frequencies. The dashed curve measures threshold elevation when the adapting field was a 9-cpd grating. For historical reasons, the contrast of this grating was one-third the contrast of the grating at the fundamental. Even at this reduced contrast, the 9-cpd grating also causes a significant threshold elevation for 9-cpd test stimuli.
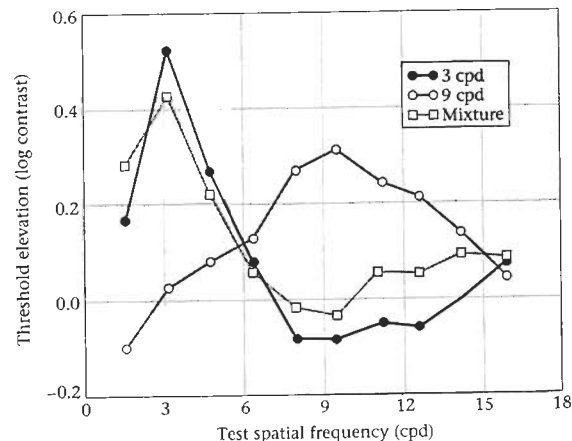
The dot-dash line shows the threshold elevation following adaptation to the mixture of the two adapting stimuli. For this observer, adaptation to the mixture shows no threshold elevation to test gratings at 9 cpd. For all of the observers in this study, the threshold elevation at 9 cpd following adaptation to the mixture is smaller than the threshold elevation following adaptation to the 3-cpd adapting stimulus. The mixture of 3- and 9-cpd stimuli is *less* potent in terms of threshold elevation than a 9-cpd stimulus alone.

This result is difficult to reconcile with the simple interpretation of adaptation and spatial frequency channels in Figure 7.12. If the adaptation to 3 cpd stimulates a different set of neurons from adaptation to 9 cpd, then why should adapting to 3-cpd and 9-cpd stimuli improve sensitivity at 3 cpd? I am unaware of any explanations of this phenomenon that also preserve the basic logical structure of the multiresolution representations.

The results of these mixture experiments should motivate us to rethink the basic mechanisms of pattern adaptation. If we plan to use this experimental method to provide support for a notion as significant as

multiresolution representation of pattern, then we should understand the adaptation phenomenon. Figure 7.18 illustrates one of the difficulties we face when we try to integrate results from detection and adaptation experiments. When we group results from detection and adaptation experiments, we assume implicitly that the visual mechanisms that limit detection are the same as those that alter visual sensitivity following pattern adaptation. But behavioral measurements provide no direct evidence that the neurons that limit sensitivity are the same as those that underlie adaptation.

The diagram in Figure 7.18 illustrates one way in which this assumption may fail. Suppose that neurons indicated as "Group T" are located early in the visual pathways and that these neurons are noisy and have low contrast gain. If these are the least reliable neurons in the pathway, then the sensitivity may be limited by their properties. In that case, we can improve the observer's detection performance by testing with stimuli that are well matched to response properties of the Group T neurons.

We have assumed that the effects of pattern adaptation are due to neural fatigue caused by strong stimulus excitation. Because the Group T neurons have relatively low gain, they will not respond very strongly

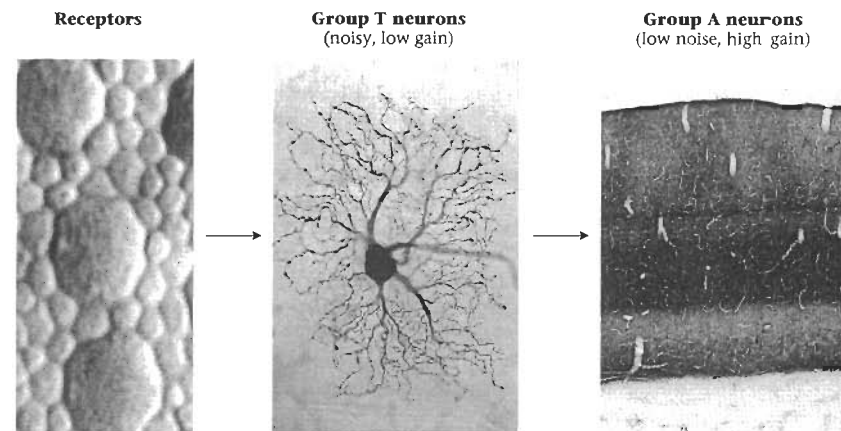7.17 PATTERN ADAPTATION MIXTURE EXPERIMENTS. These curves measures log threshold contrast elevation at various test frequencies following adaptation. The curves show the results following adaptation to a 3-cpd sinusoid (solid line), a 9-cpd sinusoid (dashed line), and their sum (dot-dash line). Threshold elevation following adaptation to the sum is smaller than threshold elevation following adaptation to the individual components. Source: Nachmias et al., 1973.

7.18 THE NEURONS THAT LIMIT DETECTION AND THOSE THAT CAUSE PATTERN ADAPTATION may not be the same. For example, one group of neurons (Group T) may be noisy and have low contrast gain. Because of their noise properties, these neurons would limit detection threshold. Neurons in a second population (Group A) may integrate the responses of the first group of neurons and have high contrast gain. Because of their high gain, this group of neurons may fatigue easily and be the neural basis of pattern adaptation. If the neural units that limit these two types of behavioral responses are different, then the spatial receptive fields of neurons that are inferred from detection and pattern-adaptation experiments may well be different. Sources: Left, courtesy of Christine Curcio; center, from Yamashita and Wässle, 1991; right, courtesy of Jennifer Lund.

to most stimuli, and consequently they may not be susceptible to pattern adaptation. Instead, it may be that another group of neurons, "Group A" neurons, are the ones most influenced by the adapting pattern. I have shown these neurons in Figure 7.18 at a later stage in the visual pathways. The spatial properties of the pattern-adaptation experiment (for example, the way test sensitivity varies with the spatial properties of the adapting pattern) may be due to the spatial receptive-field properties of the Group A neurons. Group T and Group A neurons may have quite different spatial receptive fields.[7]
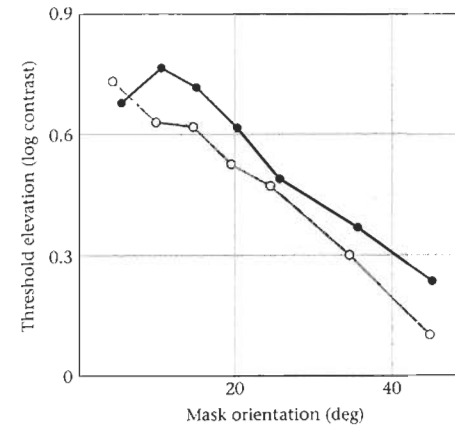
From this analysis, it should be clear that the spatial properties of the neural encoding derived from pattern adaptation may differ from the spatial properties of the neural encoding derived from detection tasks. To argue that the mechanisms limiting detection and mediating pattern adaptation are the same, we must find behavioral experimental measurements that support this point. In that case, we can piece together the results from detection and pattern adaptation to infer the organization within multiresolution models.

Masking with Mixtures

In Campbell and Robson's (1968) discrimination experiment, the observer was asked to distinguish between two stimuli S and $S + \Delta S$, where S and ΔS effectively were sinusoidal patterns. Campbell and Robson found that when S and ΔS were sinusoidal stimuli at well-separated spatial frequencies the two patterns were discriminable when ΔS was at its own threshold. In reviewing their experiments we considered how masking depends on the relative spatial frequency of the test and masking patterns (see Figure 7.15).

The data in Figure 7.19 show how masking depends on the relative orientation of the target and masker (Phillips and Wilson, 1984). In this study, the masker S and test ΔS were at the same spatial frequency. Phillips and Wilson measured the contrast needed in the test to discriminate $S + \Delta S$ from S for various orientations of the masker. The horizontal axis in Figure 7.19 measures the orientation of the masking stimulus, S. The vertical axis measures the threshold elevation of the test. As the difference in orientation between the test stimulus ΔS and masking stimulus S increases, the masking effect decreases. In this data set, when the orientation difference exceeds 40 degrees, $S + \Delta S$ can be discriminated from S when ΔS is at its own threshold level.

From these measurements, one might suspect that one-dimensional contrast patterns separated in orientation by 40 degrees are encoded by separate neurons. Experimental results like these might be used to deter-

[7] You should also consider the possibility that the basic mechanism of neural fatigue is not the main source of pattern adaptation. Recently, Barlow and Foldiak (1989) have put forward an entirely different explanation of pattern adaptation that is based on learning principles, not on neural fatigue. While the work on this topic is too preliminary to include in this volume, I think this line of research has great potential for clarifying many visual phenomena.

7.19   ORIENTATION TUNING IN THE MASKING EXPERIMENT. Threshold elevation to a 2-cpd test as a function of the orientation of a 2-cpd masking stimulus for two observers. The data include only maskers with positive orientations since masking is symmetric with respect to the orientation of the masking stimulus. Source: Phillips and Wilson, 1984.

mine the orientation-selectivity properties of convolution kernels used in multiresolution models.

Results from test-mixture experiments based on visual masking call the validity of this conclusion into question. Derrington and Henning (1989) report mixture experiments in which they measured threshold elevation using two separate masking patterns and their mixture. The individual masking patterns were 3-cpd sinusoidal gratings; one grating was oriented at +67.5 degrees and the other at −67.5 degrees relative to vertical. They measured the effect of these masking patterns on a variety of vertically oriented sinusoidal gratings.

If the two masking stimuli are represented by neurons that are different from those that represent the vertical test patterns, then the superposition of these two masking stimuli should not influence the target visibility. The data in Figure 7.20 show that this is not the case. The mixture of the two masking patterns is a potent mask even though each alone fails to have any effect.[8]

Multiresolution Theory: Current Status

The multiresolution representations are a very important theoretical tool. They help us think about the general problem of pattern sensitivity, and they provide a framework for organizing computational models of pattern sensitivity and other pattern-related tasks. There is some evidence that these representations are an important part of the human visual pathways. But there is a bewildering array of experimental methods, ranging from detection to pattern adaptation to masking, whose

[8] Similar difficulties in interpreting the effects of masking, but with respect to mixtures of sinusoidal gratings, have been studied by Nachmias and Rogowitz (1983) and Perkins and Landy (1991).

7.20   MASKING MIXTURE EXPERIMENTS. When a test and masking grating are separated in orientation by 67.5 degrees, the masker has no influence on the visibility of the test. But the combination of two masking gratings at 67.5 degrees, neither of which alone has any effect, acts as a powerful masker. The solid and shaded lines show results from two different subjects. Source: Derrington and Henning, 1989.
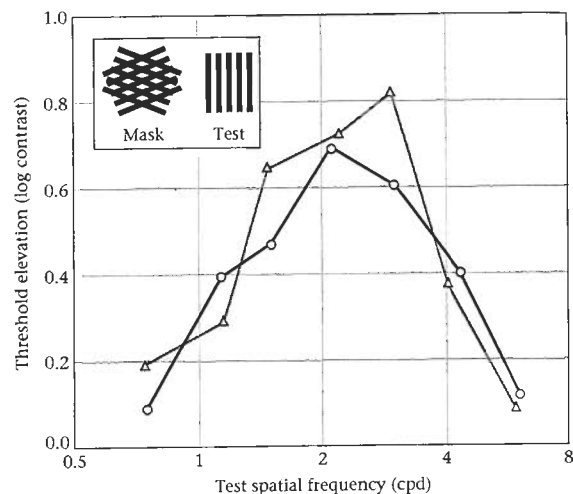
results are inconsistent with the central notions of multiresolution representations. As we have seen, mixture experiments using pattern adaptation and masking are difficult to understand if we believe that components of the image associated with spatial frequency and orientation bands are encoded by independent sets of neurons.

The conflicting pattern of experimental results show us that we have yet to achieve a complete understanding of the basic neural processes that cause adaptation and masking. Also, we do not understand how these neural processes are related to the neural processes that limit pattern sensitivity. Achieving this understanding is important because these experiments provide the key results that support multiresolution representations. Perhaps, once we understand the properties of these separate experimental methods more fully, we will understand the role of multiresolution representations and find a way to make sense of the complete set of experimental findings. Up to this point, I think you should see that we are well under way in understanding these issues, but many questions remain unanswered.

Pattern Sensitivity and Other Viewing Parameters

Next, we will review how pattern sensitivity depends on other aspects of the viewing conditions, such as the mean illumination level, the temporal parameters of the stimulus, and the wavelength properties of the pattern. In each of these cases, we will use some form of the contrast-sensitivity function as a summary of the observer's behavior.

In the remainder of this chapter, the contrast-sensitivity function plays a role that differs from the way we have used it up to now. To

this point, I have emphasized the special role of the contrast-sensitivity function in linear-systems theories. If we understand the structure of the data well enough, then the contrast-sensitivity function can be used to predict sensitivity to many other different patterns. A clear example of this is Schade's use of the contrast-sensitivity function: if visual sensitivity is limited by a shift-invariant neural image, then we can use the contrast-sensitivity function to predict sensitivity to any other pattern.

We do not have yet a complete theory that permits us to use the contrast-sensitivity function to characterize behavior generally. My purpose in continuing to treat pattern sensitivity in terms of the contrast-sensitivity function now is that it serves as a summary measure of visual pattern sensitivity. Hence, in the remainder of this chapter, we will not look at the contrast-sensitivity function as a complete description of the observer's pattern sensitivity. Rather, we will use it as a descriptive tool to help us learn something about the general pattern sensitivity of the visual system.

Part of the reason for standardizing on the contrast-sensitivity function is that the measure is used widely in both physiology and psychophysics. Hence, behavioral measurements of the contrast-sensitivity function can provide us with a measure that we can compare with the neural response at different points in the visual pathway. If a particular class of neurons, say, retinal ganglion cells, limits visual sensitivity, we should expect behavioral contrast-sensitivity curves and neural contrast-sensitivity curves to covary as we change the experimental conditions.

Light Adaptation

Figure 7.21 shows that the contrast-sensitivity function changes when it is measured at different mean background intensities. The curve in the lower left shows a contrast-sensitivity function measured at a low mean luminance level (9×10^{-4} Trolands), when rods dominate vision.[9] Under these conditions the contrast sensitivity function peaks at 1–3 cpd and the curve is lowpass rather than bandpass. The curve on the upper right shows a contrast-sensitivity function measured on a bright photopic background, one million times more intense. Under these conditions the peak of the contrast-sensitivity function is near 6–8 cpd, and the shape of the curve is bandpass. At mean background intensities higher than 1,000 Trolands, the contrast-sensitivity function remains unchanged (Westheimer, 1960; van Nes and Bouman, 1967).

The change in the shape of the contrast-sensitivity function is consistent with a few simple imaging principles. The first principle concerns the importance of achieving an adequate signal under the

[9] The Troland (Td) is a unit of measurement that represents the intensity of light in the retinal image. To measure in units of Troland it is necessary to correct for the pupil diameter. One Troland is produced on the retina when the eye is looking at a surface of 1 cd/m^2 through a pupil of area 1 mm^2.
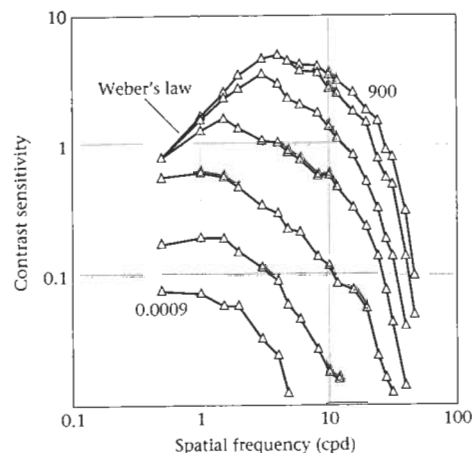
7.21 HUMAN CONTRAST SENSITIVITY VARIES WITH MEAN FIELD LUMINANCE. Each curve shows a contrast-sensitivity function at a different mean field luminance level ranging from 9×10^{-4} Trolands to 9×10^2 Trolands, increasing by a factor of ten from curve to curve. The stimulus consisted of monochromatic light at 525 nm. At the lowest level, under scotopic conditions, the contrast-sensitivity function is lowpass and peaks near 1 cpd. On intense photopic backgrounds the curve is bandpass and peaks near 8 cpd. Above these mean background levels, the contrast-sensitivity function remains constant. Source: van Nes and Bouman, 1967.

ambient viewing conditions. At very low light levels, the observer needs to integrate light across the retina in order to achieve a reliable signal. If an observer must spatially average the light signal to obtain a reliable signal, then the observer cannot also resolve high spatial frequencies. Consequently, under dim, scotopic conditions observers should have poor sensitivity to high spatial frequencies, as they do. On more intense backgrounds, when quanta are plentiful, observers can integrate information over smaller spatial regions, and spatial-frequency resolution improves.

The second principle concerns the importance of contrast, rather than absolute intensity, for visual processing. Figure 7.21 shows that contrast sensitivity of low-spatial-frequency patterns (below 1 cpd) rises with mean luminance and then becomes constant. The range in which contrast sensitivity becomes constant is called the Weber's law regime. For low-spatial-frequency patterns, Weber's law is a good description of the results. At higher spatial frequencies, contrast sensitivity continues to rise with the mean luminance. For these patterns Weber's law is not a precise description of behavioral sensitivity.

Even though Weber's law is imprecise, it does contain a kernel of truth. Consider the overall dynamic ranges we are measuring. The background intensities used in these experiments vary by a factor of one million (i.e., six orders of magnitude). Yet, the contrast sensitivity gen-

erally varies by only a factor of 20 or so (only one order of magnitude), while sensitivity to absolute light level varies by four or five orders of magnitude. The pattern of results suggests that the visual system preserves contrast sensitivity, as suggested by Weber's law, rather than absolute intensity. The visual system succeeds quite well at preserving contrast sensitivity at low spatial frequencies, and it comes close at high spatial frequencies. The significance of contrast rather than absolute intensity for vision confirms the general view we have adopted, beginning with measurements of contrast sensitivity in retinal ganglion cells and cat behavior described in Chapter 5.

Spatiotemporal Contrast Sensitivity

Figure 7.4 showed several contrast-sensitivity functions measured using contrast-reversing sinusoids. Those data illustrate how the contrast-sensitivity function varies when we measure at a few different temporal frequencies. Figure 7.22 contains a surface plot that represents how spatial contrast-sensitivity functions vary when we measure at many different temporal frequencies. One axis of the graph shows the spatial frequency of the test pattern; a second axis shows the test pattern's temporal frequency. The height of the surface represents the observer's contrast sensitivity. The surface represents the observer's **spatiotemporal**
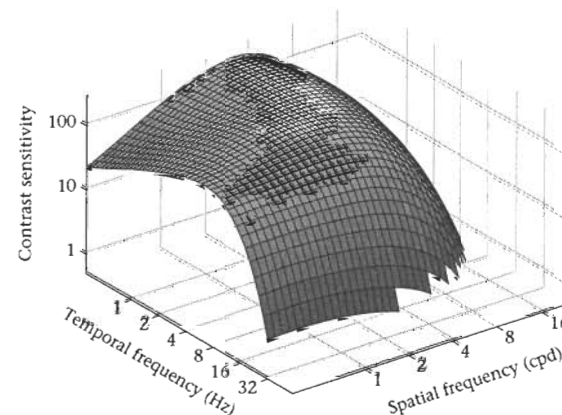
7.22 HUMAN SPATIOTEMPORAL CONTRAST-SENSITIVITY FUNCTION. The two lower axes represent the spatial and temporal frequencies of a contrast-reversing pattern; the vertical axis represents the observer's contrast sensitivity to each of the contrast-reversing patterns. The data used to estimate this surface were made on a mean background luminance of 1,000 Trolands. Curves running parallel to the spatial-frequency axis define a set of spatial contrast-sensitivity functions measured at different temporal frequencies (cf. Figure 7.4). Curves running parallel to the temporal-frequency axis represent the temporal contrast sensitivity measured at different spatial frequencies. Human spatiotemporal contrast sensitivity is not space–time separable. Source: Kelly, 1966, 1979a, b.

contrast-sensitivity function. This single surface represents a large range of spatial and temporal contrast-sensitivity functions. Paths through the surface running parallel to the spatial frequency axis represent the spatial contrast-sensitivity function; paths through the surface running parallel to the temporal frequency axis represent temporal contrast-sensitivity functions. Kelly (1979a, b) derived the analytical curve that yields the surface shape from an extensive set of psychophysical measurements.

If the spatial contrast-sensitivity functions had the same shape up to a scale factor, and similarly for the temporal contrast-sensitivity functions, we would say that human spatiotemporal contrast sensitivity is space–time separable.[10] From the shape of the contrast-sensitivity surface, it is apparent that the spatial contrast-sensitivity curves have different shapes when measured at different temporal frequencies (cf. Figure 7.4). Hence, human contrast sensitivity is not space–time separable (Kelly and Burbeck, 1984).

There are several considerations that make space–time separability an important property. First, in Chapter 5, I explained that only space–time separable systems have unique spatial and temporal sensitivity functions. When a system is not separable, it does not have a unique contrast-sensitivity function; rather it has a different function for each temporal measurement condition.

Second, space–time separability is significant because it simplifies computations and representations. For example, suppose we want to represent the spatiotemporal contrast-sensitivity function at $N_s = 60$ spatial and $N_t = 60$ temporal frequencies. If the contrast-sensitivity function is not separable, we may need to store as many as $N_s \times N_t = 3,600$ values of the sensitivity function. But, if the function is space–time separable, we need to represent only the the spatial contrast-sensitivity function and the temporal contrast-sensitivity function ($N_s \times N_t = 120$). Sensitivity to any space–time pattern can be calculated from the products of these two functions.

While the observer's behavior as a whole is not space–time separable, it is not necessary that we forgo all of the advantages of space–time separability. We may be able to describe the observer's performance as if it depends on the combination of a few space–time-separable mechanisms.[11] We first saw this approach in Chapter 5 when we studied the receptive field of retinal ganglion cells. Although their receptive fields are not space–time separable, we could model them as comprising two space–time separable components, namely, the center and surround.

Kelly (1971, 1979a, b; Kelly and Burbeck, 1984) has modeled the human spatiotemporal contrast-sensitivity function as if visual sensitivity is limited by contributions from two space–time separable components. This description of contrast sensitivity is a single-resolution description,

much like Schade's. The convolution kernel of the system is composed of a central and a surround region, much like a difference of a Gaussian, in which the two components are each space–time separable. When the two components are summed, as for retinal ganglion cells, the resulting convolution is not separable. Using suitable parameters for the Gaussians and temporal parameters, it is possible to approximate the contrast-sensitivity surface by computing the output of the convolution kernel. This single-resolution convolution kernel provides a convenient method for computing the surface, but as we have seen in other parts of this chapter the single-resolution system does not generalize well to predict sensitivity to other space–time patterns formed by the mixture of harmonic functions.

Temporal Sensitivity and Mean Luminance

The temporal contrast-sensitivity function measures sensitivity to temporal sinusoidal variations in the stimulus contrast. Figure 7.23A shows measurements of the temporal contrast-sensitivity function at a variety of mean background intensities.

First, consider how contrast sensitivity to the lowest temporal frequencies varies with background intensity. At the very lowest back-
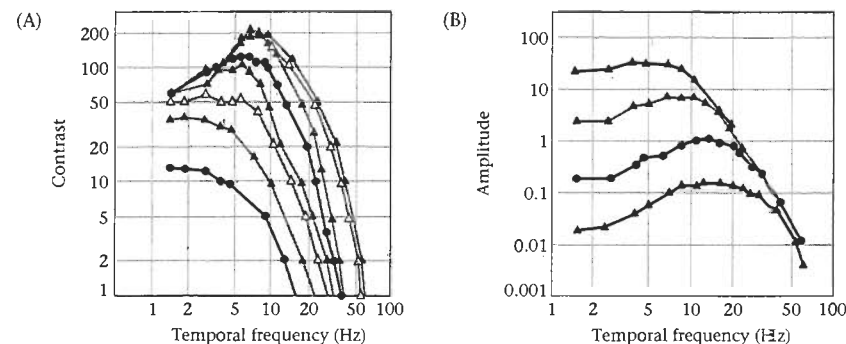
7.23 HUMAN TEMPORAL CONTRAST SENSITIVITY measured at various mean background illuminance levels. (A) Temporal contrast sensitivity. The spatial pattern was a 2-degree disk presented on a large background. Each curve measures contrast sensitivity (vertical axis) as a function of temporal frequency (horizontal axis). The curves show measurements on a variety of backgrounds. In sequence from lowest curve to highest, the mean luminance was 0.375, 1, 3.75, 10, 37.5, 100, 1,000, and 10,000 Trolands. Once the background illumination reaches roughly 5 Trolands, contrast sensitivity to low temporal frequencies remains constant, consistent with Weber's law. (B) Temporal amplitude sensitivity. The spatial pattern was a 60-degree disk. Each curve measures the threshold amplitude, not contrast, as a function of temporal frequency. The mean background levels are 0.85, 7.1, 8.5, and 850 Trolands. Notice that at high temporal frequencies the threshold amplitude appears to fall along a single curve, independent of the mean background level. This convergence is consistent with a purely linear response, involving no light adaptation, for high-temporal-frequency stimuli. Sources: A, de Lange, 1958; B, Kelly, 1961.

[10] See Chapter 5 for a discussion of space–time separability of receptive fields.

[11] Indeed, it is possible to show this is always a theoretical possibility. The result follows from an important representation in linear algebra called the *singular value decomposition*.

ground levels, contrast sensitivity increases with mean luminance. Once the mean background luminance reaches 5 Trolands, contrast sensitivity to low frequencies changes by less than a factor of 2 while the background intensity changes over a factor of 100. For low temporal frequencies, contrast sensitivity remains relatively constant across changes in the mean background intensity. This is the form of light adaptation called Weber's law.

Second, consider the contrast sensitivity at high temporal frequencies. For these tests, contrast sensitivity increases systematically at all background levels, a deviation from Weber's law. The nature of the deviation can be clarified by replotting the data as shown in Figure 7.23B where sensitivity is plotted as a function of the *amplitude* of the high-frequency flicker, not contrast (which is the amplitude divided by the mean level). When plotted as a function of amplitude, the temporal flicker sensitivity curves converge at high temporal frequencies. The convergence of the functions measured at many different mean luminance levels implies that sensitivity to high-temporal-frequency signals is predicted by the amplitude of the signal, not its contrast. This is the behavior one expects from a pure linear system, without light adaptation. In this temporal frequency range, then, Weber's law does not describe the data well at all. These data show that light adaptation does not play a significant role in determining the visibility of high-temporal-frequency flicker.

Pattern–Color Sensitivity

There is a very powerful relationship between the wavelength composition of a target and our sensitivity to pattern. In Chapter 2 we reviewed one of the most important factors that relates wavelength and pattern sensitivity: the chromatic aberration of the optics. The consequences of chromatic aberration are quite significant for the organization of the entire visual pathways. For example, based on the measurements we reviewed in earlier chapters, the chromatic aberration of the lens, coupled with the wide spacing of the S-cones, implies that a signal beginning in the S-cones can only represent signals less than 3–4 cpd (cf. Figure 2.23). This compares to the basic optical and sampling limit of nearly 50 cpd for signals initiated by a mixture of L- and M-cones. The consequences of these neural limitations and others can be measured easily in people's ability to detect, discriminate, and perceive colored patterns: people's ability to resolve short-wavelength patterns is very poor (Williams, 1986).

While it is easy to understand some of the relationship between pattern and color in terms of the optics and the cone mosaic, the limitations that relate color and pattern are best understood by thinking about the neural pathways that encode color, rather than the cones. A great deal of physiological and behavioral evidence (see Chapters 6 and 9) demonstrates that we perceive color via neural pathways that combine the signals from the three cone classes. One pathway carries the sum of the cone signals, while other pathways, called **color opponent-pathways**, carry signals representing the difference between cone signals. Signals are represented on these pathways at very different spatial resolution (Noorlander and Koenderink, 1983; Mullen, 1985; Poirson and Wandell, 1993; Sekiguchi et al., 1993b).

High-spatial-frequency signals (20–60 cpd) appear to excite only the pathway formed by summing the cone signals. We experience these patterns as light–dark modulations around the mean luminance. Spatial frequency patterns below 12 cpd can excite a pathway that encodes the difference between L and M signals. Only the lowest spatial frequencies excite the third pathway, a pathway that includes the S-cones.

These effects have been roughly understood for many years. For example, the color television broadcast system that is transmitted into many homes is organized into three color signals that correspond to a light–dark signal and two color-difference signals. Only the light–dark signal includes high-spatial-frequency information about the image; the two color channels represent only low-spatial-frequency information. This representation is very efficient for transmission, since leaving out high spatial frequencies in two of the signals permits a large compression in the bandwidth of the signal. Despite the missing spatial-frequency information, the broadcast images do not appear spatially blurred. The reason is that the high-spatial-frequency color information that is omitted in the transmission is not ordinarily perceived.

The color pathways also differ in their temporal sensitivity. Perhaps the most important observation is based on the **flicker photometry experiment**. In this experimental procedure, a pair of test lights alternate with one another. When the lights are alternated slowly the pattern appears to change between the colors of the two lights. When the lights alternate rapidly, observers fail to see the color modulation, and all differences appear as a light–dark modulation upon a steady colored background. Our temporal resolution for distinguishing blue–yellow flicker is poorest, that for red–green is in the middle, and that for light–dark is best.

The relationship between spatial resolution and temporal resolution suggests a hypothesis that we considered in Chapter 3: namely, that spatial and temporal resolution covary because they are both related through the rigid motion of objects. If the most important source of temporal variation in the image is due to motions of the eye or motions of an object, temporal-frequency and spatial-frequency resolution should covary. At a single velocity, the motion of a low-spatial-frequency image produces a slower temporal variation than motion of a high-spatial-frequency image. Hence, in those wavelength bands where only low spatial frequencies are imaged, the visual system may not require high-temporal-frequency resolution.

I have summarized the covariation of color, space, and time in Color Plate 3. The image represents how color appearance varies across

different spatiotemporal frequency ranges. We are trichromatic only in a relatively small range of low spatial and temporal frequencies represented near the origin of the figure. As the spatial or temporal frequency increases we fail to see blue–yellow variation, and vision becomes dichromatic. At the higher spatial and temporal frequencies we are monochromatic, and we see only light–dark variation.
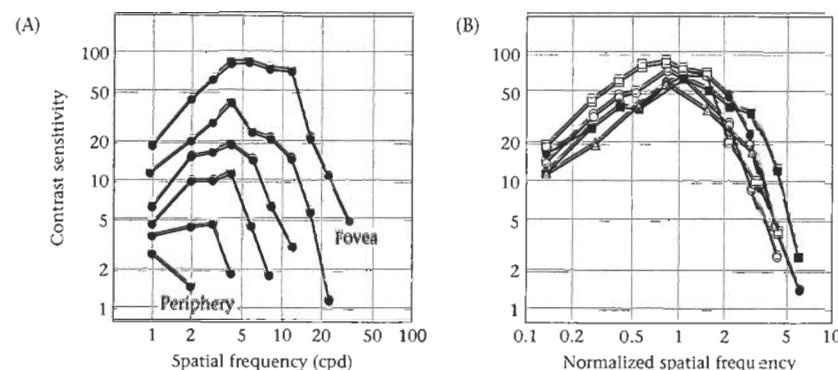
Retinal Eccentricity

The contrast-sensitivity measurements we have reviewed were all made using small patches of sinusoidal grating presented within the central few degrees of the visual field. As one measures contrast sensitivity at increasingly peripheral locations in the visual field, sensitivity decreases.[12] There are a number of neural factors that conspire to reduce both absolute sensitivity and spatial resolution. The density of the cone mosaic falls off rapidly as a function of visual eccentricity, so that there are fewer sensors available to encode the signal. The retinal ganglion cell density falls as well, as does the amount of cortical area devoted to representing the periphery. Approximately half of primary visual cortex represents only the central 10 degrees of the visual field (314 square degrees), while the remaining half of visual cortex must represent the rest of the visual field, which extends to a radius of 80 degrees (20,000 square degrees; see Chapters 5 and 6).

Figure 7.24A shows a set of contrast-sensitivity functions measured using a small grating patch at several different visual eccentricities. The top curve shows the observer's contrast sensitivity in the fovea. The observer's peak contrast sensitivity is 100 for gratings near 5–8 cpd, meaning that the observer can detect these at 1-percent contrast. In the fovea, the observer can resolve gratings as fine as 40–60 cpd. When the same stimulus is used to make measurements in the visual periphery, observers become less sensitive in all respects, so that stimuli 30 degrees in the periphery have a peak contrast sensitivity of 3 and an upper limit of 2 cpd.

We do not ordinarily notice this decrease in contrast sensitivity. When asked, most people believe that their spatial resolution is fairly uniform over a much wider extent of the image than just 2 degrees (their thumb joint at arms length). Yet, from the curves in Figure 7.24A, it is plain that our visual resolution is very poor by 7–10 degrees (a fist at arms length). Hence, our impression of seeing sharply over a large spatial extent must be due in part to our ability to integrate spatial information using eye movements.

Rovamo et al. (1978; Rovamo and Virsu, 1979; Virsu and Rovamo, 1979) suggested that the decrease in contrast sensitivity with eccentric viewing can be explained quantitatively by the reduced representation

[12] The quality of the optics does not appear to decline significantly over the first 20 degress of visual angle (Jennings and Charman, 1981).

7.24   THE CONTRAST-SENSITIVITY FUNCTION VARIES WITH RETINAL ECCENTRICITY. (A) Contrast-sensitivity functions measured using a 1 degree × 2 degree grating patch at retinal eccentricities of 0, 1.5, 4, 7.5, 14, and 30 degrees are shown. Contrast sensitivity measured using this stimulus is highest in the fovea and falls dramatically with retinal eccentricity. (B) Contrast-sensitivity functions measured with the test stimulus scaled in size and spatial frequency in order to compensate roughly for the reduced cortical area devoted to different retinal eccentricities. Source: Rovamo et al., 1978.

of the visual field in the cortex. Qualitatively, the decrease in contrast sensitivity and the coarse neural representation of the periphery do parallel one another. The rough agreement between these factors is demonstrated by the results in Figure 7.24B. These contrast-sensitivity functions, like those in Figure 7.24A, were made at different retinal eccentricities. For these measurements, however, the size and spatial frequency of the grating patch were scaled to compensate for the reduced cortical representation at that retinal eccentricity. When the size and spatial frequency of the stimulus are adjusted to compensate for the reduced cortical representation, the contrast-sensitivity functions become fairly similar.

Visual performance deteriorates with eccentricity for all known spatial-acuity tasks and spatial-localization tasks, which we will review later in this chapter; however, the performance decrease as a function of retinal eccentricity varies considerably across observers and across tasks. The reduced representation of the periphery is present in all of the neural representations beginning with the photoreceptors and continuing into the central nervous system. The variance in observers' performance coupled with the wide number of neural representations with similar decrease in the peripheral representation, make it difficult to attribute the decline in performance to any single anatomical structure. The decline of acuity with eccentric viewing is an important and widespread feature of the visual system; it may not be possible to localize its cause to a single site in the visual pathways (e.g., Westheimer, 1979; Levi et

al., 1985; Legge et al., 1987; Yap et al., 1987; Farrell and Desmarais, 1990).

Linking Hypotheses

We have now reviewed several instances in which the variation of behavioral contrast-sensitivity functions with stimulus conditions is similar to the variation of retinal ganglion cell responses. These correlations suggest that there is a causal relationship between the retinal ganglion cells' receptive fields and the behavioral measurements. But such a relationship is quite difficult to prove with the certainty that we would like to have. At this point, I think it is worth reviewing what we have learned about making such inferences.

Behavioral and neural theorizing supplement one another. Psychophysicists measure behavioral responses and then build theories about neural mechanisms. The properties of these theoretical mechanisms summarize the data and lead to new behavioral predictions. Neurophysiological measurements tell us about the neural activity directly; but we must theorize about how the neural activity influences behavior. Each field contributes part of the information about visual function.

In an influential chapter in his book, Brindley (1970) called hypotheses that connect measurements in the two fields **linking hypotheses**. He took a very conservative view concerning the type of experiments that could be used to reason about physiology from performance. His comments initiated a discussion that continues to this day (Teller, 1990; Westheimer, 1990). Brindley felt that the only truly secure argument connecting physiology and perception is this (p. 133):

> . . . whenever two stimuli cause physically indistinguishable signals to be sent from the sense organs to the brain, the sensations produced by these stimuli, as reported by the subject in words, symbols, or actions, must also be indistinguishable.

By stating his hypothesis clearly and forcefully, Brindley has drawn a great deal of attention to the problem of linking results between the separate disciplines. My purpose in writing this section is to question whether he may have succeeded too well; the emphasis on linking results from behavioral and physiological studies sometimes distracts us from making sure that the experimental logic within each discipline is complete.

We establish the most secure links between behavior and physiology when we first understand the separate measurements very well. For example, the relationship between the color-matching functions and photopigment sensitivities are strong because we have extensive quantitative studies, ranging over many measurement conditions, that tell us about each set of measurement conditions on their own. The color-matching experiment stands no matter what the photochemist observes, and the cone photopigment measurements stand no matter what the psychophysicist observes. Because each set of results stands on its own, we can feel confident that their relationship is a strong case for a connection between the two fields. If we require that the analysis within each discipline stands on its own, then when it comes time to join the two sets of observations we can have greater confidence in the link.

I mean to contrast the view stated here with an alternative approach in which the behaviorist uses the discovery of a particular neural response as the logical basis for a purely behavioral experiment; or conversely, the case in which a physiologist explains a set of recordings in terms of some potentially related behavioral measure. Such ideas may be useful in the background to help formulate specific experimental measurements. However, the logic of theories and experiments based on a web of interconnections from behavior to physiology often serve to entangle our thinking.

Given this standard, what should we think about the connection between behavioral contrast sensitivity and neural receptive fields? In this chapter we have found that there is a powerful theory underlying behavioral contrast-sensitivity functions. This theory is a good match to the logic of receptive-field organization we reviewed in earlier chapters. The psychophysical results based on the contrast-sensitivity function, however, do not fully support the basic theory. We cannot yet generalize from contrast-sensitivity functions to sensitivity to other stimuli. Hence, the association between receptive-field properties and contrast-sensitivity functions are far more tentative than the connection between the color-matching functions and the photopigment spectral sensitivities.

Having stated this limitation in our current understanding, I must add that I do not think we should be discouraged. The similarities between the properties of the contrast-sensitivity functions and neural receptive fields are too striking to ignore. By continuing to improve on the models for behavior and receptive fields separately, the links we forge and the quantitative comparisons we make could well turn out to form a complete model, linking behavioral pattern sensitivity and neural receptive fields.

Spatial Localization

In this section we will review how well human observers can localize the position of a target. Wulfing (1892) showed that human observers can make surprisingly fine discriminations between the positions of two objects. Observers can reliably distinguish spatial offsets between a pair of lines as small as one fifth the width of a single cone photoreceptor. Moreover, people can distinguish this spatial offset even when the objects are moving (Westheimer, 1979).
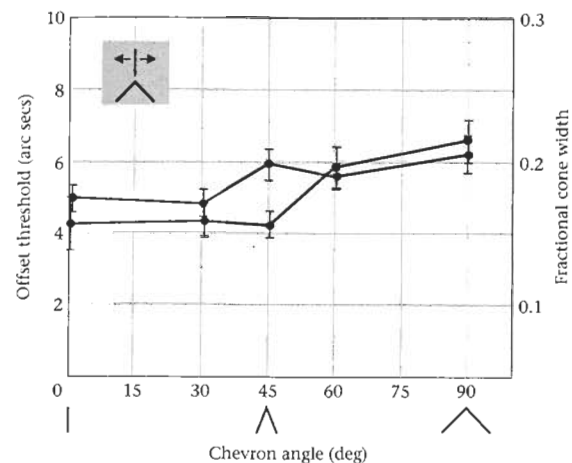
Separation   **Shift**

4 sec   4 sec

8 sec   8 sec

16 sec   16 sec
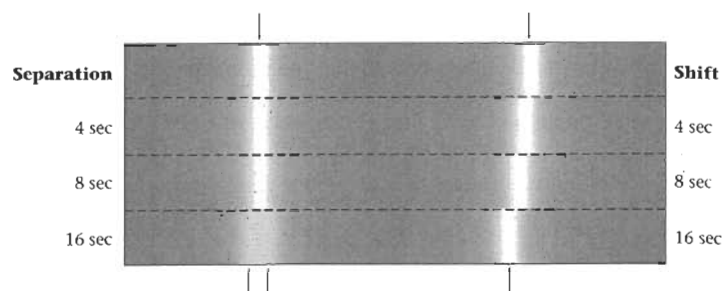
7.25   A COMPARISON OF LOCALIZATION AND SPATIAL-RESOLUTION EXPERIMENTS. In a two-line spatial-acuity experiment, the observer distinguishes between a stimulus consisting of a single line from a stimulus consisting of a pair of lines separated by a small amount. The images on the left side show the estimated retinal light distribution of a reference line and of three pairs of lines separated by increasing amounts. In a localization experiment, the observer distinguishes the position of a single line from the position of a displaced line. The images on the right side of the figure show a reference line and the estimated retinal light distribution of three offset lines.

Chevron angle (deg)

7.26   LOCALIZATION SENSITIVITY. Subjects detected whether a line was offset to the right or left of the tip of a chevron for a variety of angles of the opening of the chevron. Data from two observers are shown. Both observers could reliably report offsets as small as six seconds of arc (left vertical scale) which is one-fifth the width of a single photoreceptor (right vertical scale). Source: Westheimer and McKee, 1977.

The ability to discriminate between targets at different spatial positions is an aspect of human spatial resolution. It is important to recognize that that the ability to localize a target is a different kind of resolution from the spatial resolution we measure when we ask observers to discriminate a pattern from a uniform background.[13] The differences between the tasks are illustrated in Figure 7.25. The left side of the image in Figure 7.25 shows the estimated retinal light distributions of several stimuli a subject might be shown in a spatial resolution task. In this experiment, the subject must discriminate between the light distribution of a single line (top left) from the light distribution of a line-pair in which the two lines are separated by a small amount (bottom left). In this task, the stimuli are all centered at the same point, so there is no difference in where they are located. The right side of the image in Figure 7.25 shows the retinal light distributions of stimuli a subject might be shown in a spatial-localization task. In this experiment, the subject must discriminate the position of the retinal light distribution created by a reference line (top right) from the positions of the light distributions of a line that is offset (bottom right).

Westheimer and McKee (1977) measured observers' ability to localize a line (Figure 7.26). Subjects judged whether a line was located to the right or left of the tip of a chevron (see inset in the figure). The vertical axis of the graph measures the displacement needed to discriminate re-

[13] The terminology associated with these two types of spatial tasks can be confusing. The word *hyperacuity* refers to the fact that people localize spatial position with very high precision. Unfortunately acuity is also used to refer to the spatial-frequency sensitivity of the observer, which is a different matter. Here, I will use the term *localization* to refer to spatial resolution for position.

liably when the line is offset from the middle of the chevron. This task was repeated for chevrons with various angles; for all angles, the offset thresholds are on the order of 5 seconds of arc, roughly one-fifth the width of a single cone. Performance does not vary much as we change the stimulus. This suggests that localization performance is robust with respect to spatial manipulations of the target. This very fine localization applies to many different kinds of stimuli, including the relative positions of a pair of vertical lines, moving lines, and many other targets (Westheimer, 1979).

At first, it seems surprising to learn that we can localize targets at a finer resolution than the spacing of the cone mosaic. We know that the sampling grid determined by the cone mosaic imposes a fundamental limit on spatial pattern resolution through the phenomenon of aliasing (see Chapter 3). Why should the cone mosaic not also impose a limitation on our ability to localize position?

In fact, a coarse sampling grid does not eliminate the possibility of localizing a target precisely. The physical principles we can use to achieve fine spatial localization on a coarse sampling grid are illustrated in Figure 7.27. The main portion of the figure shows the pattern of cone absorptions we expect in response to a reference line centered over a cone and a line that is displaced to the right by 12 seconds of arc. The separation between the tick marks on the horizontal axis are set at 30 seconds, the size of an individual cone. The values were calculated using Westheimer's optical linespread function (Chapter 2).

Because the offset is very small compared to the sampling, the same cones respond to the reference line and the offset line. It follows that the identity of the cones cannot be used to estimate the locations of the
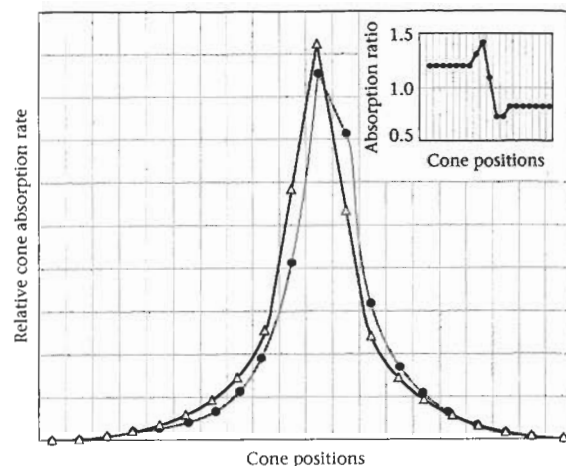
7.27 A PHYSICAL BASIS FOR LOCALIZATION IN LOCALIZATION TASKS. The points in the main graph show the estimated rate of light absorption by foveal cones in response to a fine line stimulus. The open triangles show absorptions to a reference line and the filled circles show the absorptions to a line offset by 12 seconds of arc. The grid lines on the horizontal axis are separated by the width of a single cone. The solid and shaded lines linearly connect the estimated absorption rates. The inset shows the ratio of cone absorptions from the reference line and the displaced line at each cone position. The graph shows that a 12-second shift (roughly one-third the width of a photoreceptor) changes the cone absorption rate by as much as 50 percent. This information can be used to localize the position of the line at a resolution that is substantially finer than the separation between cones.

two lines. Although the same cones respond to the two lines, the spatial pattern of cone absorptions when the lines are in these two positions is quite different. The inset to Figure 7.27 shows the ratio of the cone absorptions to the two different lines. A small spatial shift of 12 seconds of arc causes a 50-percent change in the absorption rate at an individual cone. Hence, the spatial pattern of absorption rates is a reliable signal that can be interpreted to infer the line position at a resolution finer than the spacing of the cone mosaic.

Notice that the optical blurring of the light distribution is essential if we wish to localize the line at positions finer than the sampling grid. Were there no optical blurring, the image of a line would fall within the width of a photoreceptor, and spatial displacements less than the width of a photoreceptor would not be detectable. Optical blur, which seems like a nuisance when we consider spatial resolution of contrast patterns, is a help when we consider spatial resolution in terms of localizing targets.

Our ability to localize the position of edges and lines is very robust with respect to various stimulus manipulations. If we vary the target contrast, set the display into motion, or flash the display briefly, performance remains excellent. Since performance is robust with respect

to these experimental manipulations, it is clear that the simple calculation shown in Figure 7.27 is only a demonstration of how localization is possible; the visual system must use a much more sophisticated method to calculate position. Eye movements during examination of a static display, or tracking errors during examination of a visual display, will make it impossible to compare the outputs of a single small set of cones. Rather, people must be capable of estimating the position at fine precision even though the precise identity of the cones mediating the signal varies. Although we have some basic principles to work from, how we estimate the relative position of moving targets using active eyes remains an important challenge for study.

Summary

Theories of human pattern sensitivity are organized around a few basic principles. In the earliest and simplest theories, the visibility of different types of test patterns was explained by the properties of a single shift-invariant linear system. This type of theory is simple for computation and also parallels nicely our understanding of the initial encoding of light by retinal nerve cells in certain visual streams. The convolution kernel of the shift-invariant linear system and the neural receptive field play analogous roles and provide a natural basis for comparison of behavioral and physiological data. By using common experimental measures, such as contrast-sensitivity functions, the properties of neural mechanisms and behavioral theories can be compared directly.

While certain aspects of single-resolution theories provide a reasonable description of human pattern sensitivity, they fail a number of direct empirical tests. Consequently, theorists have tried to assemble new theories in which the pattern representation is based on a collection of shift-invariant representations, not just a single one. This idea parallels the physiological notion that the visual system contains a set of visual streams. The more complex modern theories must specify a larger number of convolution kernels (receptive fields). To keep these organized, and to parallel some of the properties of cortical receptive fields, theorists generally choose convolution kernels that respond best to restricted bands of spatial frequency and to restricted stimulus orientations. These theories can predict more experimental results, but there remain many computational and experimental challenges before we will have a complete satisfactory theory of pattern sensitivity.

Because human vision constantly adapts to new viewing conditions, human pattern sensitivity cannot be described by a single pattern-sensitivity function. Pattern sensitivity covaries with the temporal properties of the test stimulus, the mean background level, and with the wavelength composition of the stimulus. Thus, a general specification of human pattern sensitivity must take all of these factors into account.

Behavioral experiments show that people are also exquisitely sensitive to the spatial location of targets. Observers can localize test stimuli

to a resolution that is considerably finer than the spacing of the cone mosaic. The ability to localize is quite robust, surviving many different stimulus manipulations. The principles of how one might localize to a very fine resolution are clear, but the methods that the visual pathways use to acquire the necessary information remain to be determined.

Exercises

1. Use a computer program, such as MATLAB or Mathematica, and experiment with vector-length calculations:

   (a) Define a discrete representation of a one-dimensional sinusoidal stimulus. Using a range of locations for x from 0 to 1 degree, and 64 sample values, create two sinusoidal patterns, s_1 and s_3.

   (b) Compute the contrast patterns $s_1 + s_3$ and $s_1 - s_3$. Plot them and compare the deviations around zero.

   (c) Compute the squared vector-length of a vector s_1 using a matrix multiplication of the form $s_1^T s_1$.

   (d) Show that a shift-invariant, vector-length calculation predicts that the visibility of the pattern $\sin(2\pi f x) + \sin(2\pi(3f)x)$ is the same as the visibility of $\sin(2\pi f x) - \sin(2\pi(3f)x)$.

   (e) Show that a shift-invariant, vector-length calculation predicts that the visibility of the pattern $\sin(2\pi f x) + \sin(2\pi(3f)x)$ is the same as the visibility of $\cos(2\pi f x) + \cos(2\pi(3f)x)$.

2. Answer these questions about shift-invariant vector-length calculations.

   (a) Consider the two stimulus contrast patterns, $\cos(2\pi f x)$ and $\cos(2\pi(3f)x)$. Suppose that the contrast threshold for the sinusoid alone is C_f and that the contrast threshold for the cosinusoid alone is C_{3f}. Compute the predicted contrast threshold for the mixture $\cos(2\pi f x) + \cos(2\pi(3f)x)$ according to a shift-invariant, vector-length theory.

   (b) What will the contrast threshold be for $\sin(2\pi f x)$?

   (c) What will the contrast threshold be for $\sin(2\pi f x) + \cos(2\pi(3f)x)$?

   (d) Suppose that we have a one-degree wide monitor region that has 32 lines. Using the Fourier series, we can describe a contrast pattern of a line as the sum of a set of 32 cosinusoids of equal amplitude and frequencies of $0, 1, \ldots, 31$ cpd. Assume that you know the visual sensitivity to each of these cosinusoids (call them s_i). Use the shift invariant, vector-length theory to predict when the contrast line will be visible.

3. Answer these questions about texture and spatial frequency.

   (a) Suppose that we are looking at a texture pattern on a wall. As we walk toward the wall, the spatial-frequency content of the image incident at our cornea changes. Describe the nature of the change in amplitude of the different spatial-frequency components.

   (b) Describe the nature of the change of the phase components of the image.

   (c) There are many instances in which we would like to predict whether one image is a good replica of another. There is a field of engineering

in which individuals have developed *image-quality metrics* to compute whether small differences will be perceptible, and if so how perceptible. Consider only the problems of black and white images and describe what image formation factors you think should be included in an image-quality metric?

   (d) How would you incorporate the human contrast-sensitivity function into an image quality metric?

   (e) Many investigators do not use the human contrast-sensitivity function to evaluate the quality of the reproduction. Instead they evaluate the error of the replication by computing the mean squared error between the original and the replica. Can you devise a thought experiment to illustrate when this procedure will surely fail?

   (f) For many images, the mean squared error between the original and the replica performs reasonably well as a predictor of the perceptual difference between the image and its replica. Can you explain what factors might make this so?

4. Answer these questions about light adaptation.

   (a) Design a set of stimuli and experimental procedure for measuring the contrast-sensitivity function by signals initiated in the L-cone mosaic. What will be the main difficulties in creating the stimuli? What about for the S-cone mosaic?

   (b) The experiments we reviewed in this chapter were based mainly on backgrounds and targets that appear light–dark. Suppose that we adjust the color of the background light such that the L-cones are adapted to a much higher level of illumination than the S-cones. What do you think will happen to the contrast-sensitivity function measured with a light–dark modulation? What do you think will happen to the contrast-sensitivity function measured with a colored target?

   (c) Why should contrast sensitivity, rather than absolute light sensitivity, be a key visual variable preserved by adaptation?

   (d) As the mean background intensity increases, people can perceive increasingly rapid flicker. What imaging principles does this illustrate?

5. Answer these questions about wavelength and pattern.

   (a) Chromatic aberration reduces spatial resolution in the short-wavelength portion of the spectrum. Does this mean that a blue–yellow neural pathway must have poor spatial resolution?

   (b) Define pattern–color separability.

   (c) Suppose that pattern–color separability is a desirable feature of a opponent-color pathway. How would you create a blue–yellow neural signal that is pattern–color separable?

   (d) Suppose that pattern–color separability is not important and that only spatial resolution is valuable. Describe how you would create a blue–yellow neural signal.

6. Answer these questions about the physical basis for localization.

   (a) Is there any reason that localization would improve if the human cones were spaced more finely?

   (b) How well should we be able to localize using only the signals from the S-cone mosaic?

   (c) Is there any reason that localization would deteriorate if the human cones were spaced more coarsely? What effect would broadening (narrowing) the linespread function have on localization?