Pergamon

PII: S0042-6989(97)00154-5

Vision Res., Vol. 38, No. 2, pp. 231–257, 1998

Printed in Great Britain
0042-6989/98 $19.00 + 0.00

Spatial Summation in Simple (Fourier) and Complex (Non-Fourier) Texture Channels

NORMA GRAHAM,*‡ ANNE SUTTER†

Received 6 November 1996; in revised form 8 May 1997

Complex (non-Fourier, second-order) channels have been proposed to explain aspects of texture-based region segregation and related perceptual tasks. Complex channels contain two stages of linear filtering with an intermediate pointwise nonlinearity. The intermediate nonlinearity is crucial. Without it, a complex channel is equivalent to a single linear filter (a simple channel). Here we asked whether the intermediate nonlinearity is piecewise-linear (an ordinary rectifier), or compressive, or expansive. We measured the perceptual segregation between element-arrangement textures where the contrast and area of the individual elements were systematically varied. For solid-square elements, the tradeoff between contrast and area was approximately linear, consistent with simple linear channels. For Gabor-patch elements, however, the tradeoff was highly nonlinear, consistent with complex channels in which the intermediate nonlinearity is expansive (with an exponent somewhat higher than 2). Also, substantial individual differences in certain details were explainable by differential intrusion from "off-frequency" complex channels. Lastly, the results reported here (in conjunction with those of other studies) suggest that the strongly compressive intensive nonlinearity previously known to act in texture segregation cannot be attributed to a compressive nonlinearity acting locally and relatively early (before the spatial-frequency and orientation-selective channels) but could result from inhibition among the channels (as in a normalization network). © 1998 Elsevier Science Ltd

Texture Nonlinear Rectifier Non-Fourier Spatial frequency

INTRODUCTION

Channels selectively sensitive to spatial frequency and orientation, with some relatively simple nonlinearities, explain many of the phenomena of texture segregation and related visual tasks. Two different types of nonlinearity—one intensive and one spatial—seem to be necessary (e.g. Graham, 1991, 1994; Graham, Beck, & Sutter, 1992; Malik & Perona, 1990; Sperling, 1989; Wilson, 1993). Yet much remains unknown about these simple nonlinearities. The primary aim of the current experiments was to study further the spatial nonlinearity in texture segregation and related tasks, but the results also have implications for the intensive nonlinearity in these tasks.

The intensive nonlinearity: relatively early and local, or inhibition among channels (normalization)?

Two kinds of processes have been suggested as an explanation for the effects requiring an intensive nonlinearity (e.g. Bowen & Wilson, 1994; Graham, 1991; Graham *et al.*, 1992; Malik & Perona, 1990; Sperling, 1989; Wilson, 1993). One candidate is a relatively early and local process occurring before the channels (although it must occur at some stage after the relative sensitivities to different spatial frequencies and orientations at different mean luminances have been set. Graham & Sutter, 1996). The other is mutual inhibition or normalization occurring among the channels outputs (modeled as a gain control or normalization network, e.g. Heeger, 1991). Of course, both kinds of processes are known to occur in the visual system, so the question here is to what extent either (or both) affect perception in various circumstances. For texture segregation tasks, we had not yet been able to tell these two candidates apart (Graham & Sutter, 1996).

The spatial nonlinearity—complex (second-order, non-Fourier) channels

Complex channels like that in Fig. 1 have been proposed to explain a number of phenomena of spatial perception that cannot be explained by simple linear channels. They have been called variously complex channels, non-Fourier or second-order mechanisms, collector units, or collator units. We called them "complex channels" (Sutter, Beck, & Graham, 1989) having been inspired by John Robson's suggestion that complex

*Department of Psychology, Columbia University, New York, NY 10027, U.S.A.
†Department of Psychology, Loyola University, 6525 N. Sheridan Road, Chicago, IL 60626, U.S.A.
‡To whom all correspondence should be addressed [*Fax*: +1-212-854-3609; *Email*: nvg@psych.columbia.edu].

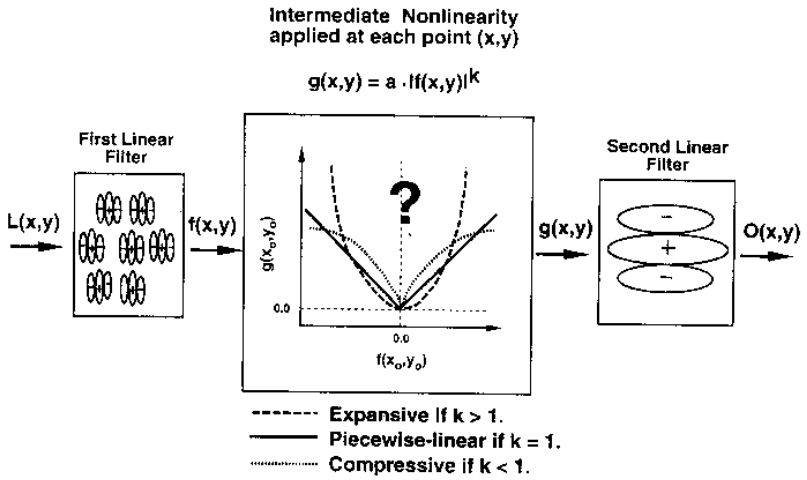

FIGURE 1. Diagram of complex channel consisting of two stages of linear filters separated by an intermediate pointwise nonlinearity. *L(x,y)* is the stimulus pattern represented as luminance at each point *(x,y)*; *f(x,y)* is the output of the first filter; *g(x,y)* is the output after the intermediate nonlinearity; and *O(x,y)* is the output of the second filter and therefore of the channel at point *(x,y)*. Although only a few receptive fields are shown for the first and second filters, they represent a whole collection of receptive fields of the same size and shape spread densely across the visual field. The channel output *O(x,y)* goes, along with the outputs from simple and other complex channels, to a comparison and decision stage shown in Fig. 4.

cortical cells could perform such perceptual tasks (e.g. Robson, 1980), but the possible analogy to physiology should be made with care. Indeed, others used the analogy with complex cells in other ways. (See p. 732 top, Graham, Beck, & Sutter, 1992 and comments in Part I of Appendix I here.) The complex channel in Fig. 1 consists of two stages of linear filters—where the first is sensitive to a higher spatial frequency than the second—with an intermediate pointwise nonlinearity. To put it another way, complex channels consist of big receptive fields at the second stage that "paste together" the (nonlinearly transformed) outputs of little receptive fields at the first stage. A number of complex channels sensitive to different spatial frequencies and orientations are presumed to exist at each point in the visual field. They are usually assumed to exist alongside simple linear channels.

The complex channels investigated in this study are those that explain why certain textures and patterns can be segregated or discriminated in spite of their very similar Fourier amplitude spectra (e.g. Graham, 1991; Graham *et al.*, 1992; Kingdom & Keeble, 1995, 1996; Landy, 1996; Landy & Ternes, 1995; Lin & Wilson, 1996; Solomon & Sperling, 1995; Sperling, 1989; Sperling, Chubb, Solomon, & Lu, 1994; Sutter, Beck, & Graham, 1989; Sutter & Graham, 1995; Sutter, Sperling, & Chubb, 1995; Victor & Conte, 1989, 1991, 1996). These complex channels seem quite similar to mechanisms proposed for other texture discriminations (Bergen & Landy, 1991; Fogel & Sagi, 1989; Grossberg & Mingolla, 1985; Landy & Bergen, 1991; Malik & Perona, 1990; Rubenstein & Sagi, 1993, 1996; Wolfson & Landy, 1995) but they are not necessarily identical (see Appendix I, Part I).

Mechanisms like complex channels have been proposed for many other tasks involving spatial patterns. These proposed mechanisms contain, as do the complex texture channels, two layers of filtering separated by a nonlinearity, where the first filtering stage is sensitive to higher spatial frequencies than the second. These other tasks include: pattern masking with compound gratings, especially the effect of beat frequencies (Badcock & Derrington, 1989; Henning, Hertz, & Broadbent, 1975); illusory contours formed by offset gratings or Kaniza triangles (Shapley & Gordon, 1985; Wilson & Richards, 1992; Wilson, 1993); the lateral enhancement of perceived contrast (Sagi & Hochstein, 1985); detecting a path defined by the alignment of elements along a smooth path that is imbedded in a background of unaligned elements (Moulden, 1994; also see Beck, Rosenfeld, & Ivry, 1989 and Field, Hayes, & Hess, 1993 although their proposed mechanisms are more active and interactive than the complex channels here); detection of a target of one or more Gabors in a field of different Gabors (Rubenstein & Sagi, 1993; Sagi, 1990); computing shape from texture (Sakai & Finkel, 1995); stereoacuity (e.g. Hess & Wilcox, 1994; Wilcox & Hess, 1996); several varieties of spatial localization tasks like separation discrimination (Hess & Badcock, 1995) and alignment accuracy (Hess & Hayes, 1994; Levi & Waugh, 1996; also see other references given in these articles); detecting the centroid of a cluster of separated dots of opposite contrast polarity (Morgan & Glennerster, 1991); and some geometrical illusions (Morgan & Hotopf, 1989; Morgan, Hole, & Glennerster, 1990). It is too early to know whether these mechanisms proposed for other tasks are the same as the complex texture channels just described.

The complex texture channels of Fig. 1 are also structurally analogous to the non-Fourier mechanisms proposed for various phenomena of motion perception (e.g. Boulton & Baker, 1993; Chubb & Sperling, 1988, 1989; Chubb, McGowan, Sperling, & Werkhoven, 1994; Derrington & Badcock, 1992; Derrington & Henning, 1994; Fleet & Langley, 1994; Lu & Sperling, 1995; Pantle, 1992; Sperling, 1989; Sperling, Chubb, Solomon, & Lu, 1994; Turano & Pantle, 1989; Werkhoven, Sperling, & Chubb, 1993; Wilson, Ferrara, & Yo, 1992; Wilson, 1994; Wilson & Kim, 1994). The channels for texture and motion might, in principle, even be parts of the same population of channels although serving different parts of the spatial-frequency/temporal-frequency range. There are, however, already known to be a number of differences between the channels serving texture and motion (e.g. Graham, 1994; Hammett & Smith, 1994; McOwan & Johnston, 1996; Solomon & Sperling, 1995; Sperling, Chubb, Solomon, & Lu, 1994) and we would not expect the results of the studies here to necessarily generalize to motion.

Finally, another set of proposed mechanisms—which will be called "higher-order" here for the sake of having a short name—seem structurally similar to the complex channels of Fig. 1 but differ in an interesting way. These higher-order mechanisms still consist of two stages of filtering, but the two stages are rather different from those in a complex channel. The second stage of a complex channel collects outputs from first-stage receptive fields, all of much the same shape but *differing in position*, whereas the second stage of a *higher-order mechanism* collects outputs from first-stage receptive fields all at much the same position but *differing in shape*. Higher-order mechanisms of several types have been suggested. Mechanisms where the pooling is across receptive fields of different symmetries (phases) at the same location have been proposed for encoding lines and edges and explaining some geometrical illusions (e.g. Morrone & Burr, 1988; Burr & Morrone, 1994). Various kinds of pooling across spatial frequencies or across orientations (which may be flexible and depend on task, and which is often coupled with a pooling across a limited area of space as well) have been proposed to explain the perceived appearance of and discrimination among several kinds of suprathreshold patterns including textures (e.g. Georgeson, 1992, 1994; Klein, Stromeyer, & Ganz, 1974; Meese, 1995; Meese & Georgeson, 1996; Polat & Sagi, 1993; Rubenstein & Sagi, 1996; Sakai & Finkel, 1995; Thomas, Olzak, & Shimozaki, 1993; Thomas & Olzak, 1996), as well as the performance of specific tasks like encoding of binocular disparity (e.g. Fleet, Wagner, & Heeger, 1996). It is still quite unclear whether any of these higher-order mechanisms are more than superficially similar to each other or to the complex (second-order, non-Fourier) channels invoked for texture and motion (see Appendix I, Part I).

All the proposals mentioned above-complex channels for textures and for other spatial tasks, non-Fourier motion mechanisms, higher-order mechanisms—have in common that the second stage collects responses from multiple receptive fields at the first stage. For this scheme to be useful when the two stages are linear filters, as they usually are, there must always be some intermediate nonlinearity involved in the collection. If there were not, the two linear-filtering stages would collapse to a single linear-filtering stage (and thus to a simple linear channel). Thus, the intermediate nonlinearity is a crucial part of these schemes.

This study

In this study we try to characterize further the intermediate nonlinearity for at least one case of complex channels—those in texture-based region segregation. While a certain amount is known about other properties of these complex channels, relatively little is known about the intermediate nonlinearity. (What is known will be presented in the Discussion and compared with the results from this study.) Here we study the intermediate nonlinearity using element-arrangement textures like those in Fig. 2. These patterns contain three regions, each composed of the same two element types but distinguished by the arrangement of the elements—striped arrangements on the sides and a checkerboard arrangement in the middle. The observer is asked to rate how immediately and effortlessly the different regions in the pattern are perceived to segregate.

To characterize the intermediate nonlinearity, we study spatial summation by varying the area and contrast of Gabor-patch elements in patterns like that in Fig. 2(a, b). (The patches used in the experiment had twice the number of cycles shown in this figure.) The first question is whether any complex channel of the form shown in Fig. 1 is consistent with the observers' ratings. The answer to this general question turned out to be "yes". Therefore, we could ask a more specific question: is the intermediate pointwise nonlinearity in the hypothesized complex channels best thought of as piecewise linear (as the solid line in Fig. 1 shows) or as some other function. In particular, we ask which value of k in the following function best predicts our data:

$$g(x,y) = a \cdot |f(x,y)|^k \qquad (1)$$

where $g(x,y)$ is the output at point (x,y) of the intermediate nonlinearity in the complex channel, $f(x,y)$ is the input at point (x,y), and a is an arbitrary constant. If $k = 1$, then g is a piecewise-linear function (solid line in Fig. 1) as in conventional full-wave rectification. If $k > 1$, then g is an expansive function (dashed line in Fig. 1). If $k < 1$, g is compressive (dotted line in Fig. 1). We might have considered a family of functions larger than that in equation (1) but this family proved sufficient for our purposes.

Examples of compressive, linear, and expansive functions abound in the visual system. The intensive nonlinearity in texture segregation mentioned above is compressive and, indeed, compressive at quite low contrasts for constant-difference series of element-arrangement patterns (Graham, 1991; Graham *et al.*,

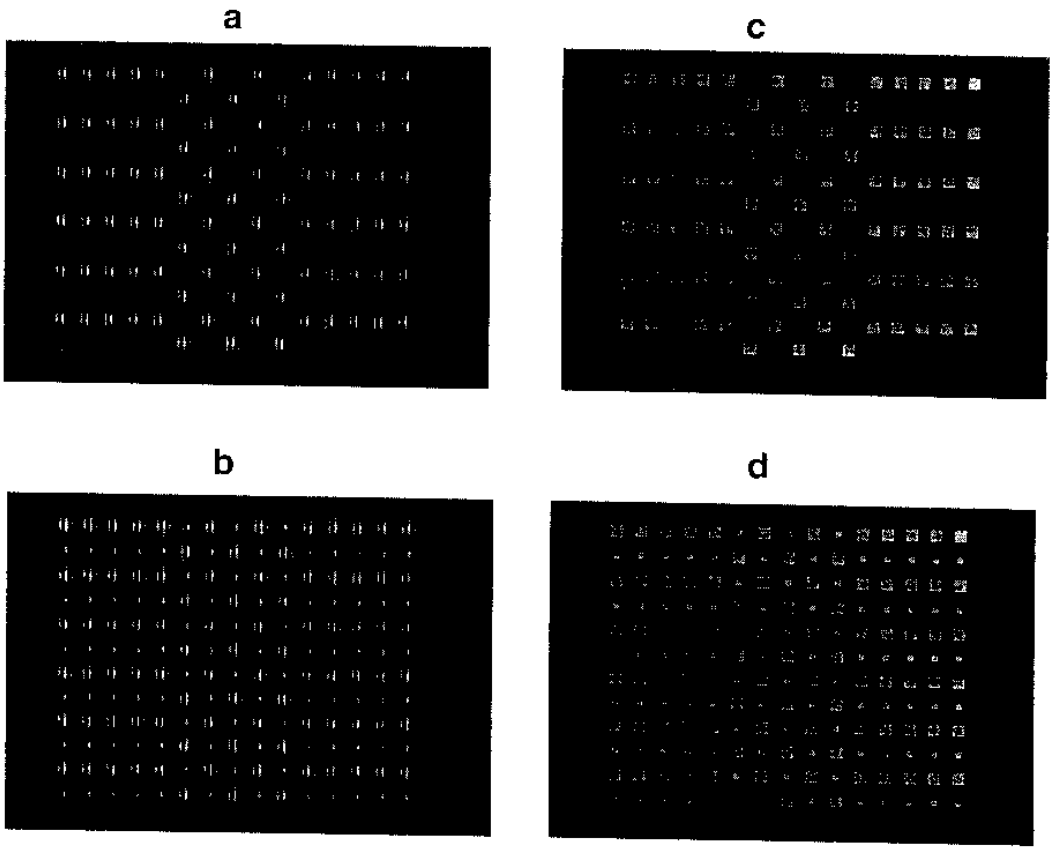

FIGURE 2. Examples of the element-arrangement texture stimuli with grating-patch elements in (a) and (b) and square elements in (c) and (d). Each contains three regions composed of the same two element types but distinguished by the arrangement of the elements—striped on the sides and checkered in the middle. The examples in (b) and (d) have equal contrast in the large and small patches. The examples in (a) and (c) have zero contrast in the small elements (and are sometimes called one-element patterns). The patches we used in the study had twice the number of cycles shown here. For half the observers the orientation of the grating patches was perpendicular to the orientation of the stripes (as shown here). For the other half the patches were parallel to the stripes.

1992; Graham & Sutter, 1996); the possible relationship of this intensive nonlinearity to the complex channels' intermediate nonlinearity will be discussed below. The intermediate nonlinearity in the complex channels has often been assumed to be either piecewise-linear or expansive (see Discussion for references), although with little evidence for either. The results of the study reported here will favor an expansive function with a power somewhat higher than 2, perhaps 3 or 4.

A conceptually related question is whether the intermediate nonlinearity should be of the full-wave type [even-symmetric around zero as in equation (1) above] or of the half-wave type. The results here cannot distinguish among these alternatives, but available evidence (briefly described in the Discussion) suggests both forms may act in texture perception.

We have phrased these questions in terms of a model of a complex channel that has an intermediate nonlinearity between two stages of linear filtering (as in Fig. 1). One could, in principle, add a compressive or expansive nonlinearity to one of the two stages of filtering instead. (Several ways to do so are described in Part III of Appendix I. For the results here, they are formally identical to the model of Fig. 1.) For simplicity's and concreteness's sake, although one might talk more generally of the "nonlinearity associated with a complex channel", we will continue in the main text to talk about the "intermediate nonlinearity" in a complex channel like that in Fig. 1.

In addition to investigating spatial summation in complex channels, the study in this paper investigates spatial summation in simple (Fourier) channels by studying the tradeoff between the area and contrast of solid-square elements [Fig. 2(c, d)]. This is a replication and extension of one experiment in Sutter *et al.* (1989). We find that the summation for solid-square elements is consistent with the assumption that simple linear channels are primarily responsible for the segregation of these patterns.

Substantial individual differences have been found in texture segregation and related tasks (e.g. Graham, Sutter, & Venkatesan, 1993; Cannon & Fullenkamp,

1993; Graham & Sutter, 1996). Thus, a third aim of the study here is to collect enough data from each observer that individual observers' results could be described and any individual differences understood. In the event, there were substantial individual differences in two aspects of the results. These differences among observers may be the result of differential intrusion from channels other than those tuned to the spatial characteristics of the patterns at issue.

Finally, the findings of this study, when considered together with those of other studies (Graham, 1991; Graham *et al.*, 1992; Graham & Sutter, 1996; and Graham and Sutter, in preparation), have implications for the intensive as well as the spatial nonlinearity in texture segregation. In particular, they suggest that the intensive nonlinearity is probably *not* a relatively local process occurring before the spatial-frequency and orientation-selective simple and complex channels. It may, however, be inhibition among these channels (as in a contrast-gain or normalization network).

THE MODEL'S PREDICTIONS

This section attempts to provide insight into why using element-arrangement patterns like those of Fig. 2 can reveal properties of spatial summation in simple and complex channels, respectively. In particular, it concentrates on the predictions of the channels "tuned" to the spatial characteristics of the pattern. (Note that we use the words "tuned channels" to refer to those channels for which the best stimulus is the pattern under discussion, while calling all other channels "untuned channels" although these other channels may well be tuned in the sense of being selectively sensitive. We also sometimes use the words "on-frequency" and "off-frequency channels" instead, with the caveat that those words are somewhat misleading in the case of complex channels, where a single frequency is not at issue.) Square-element patterns and the tuned simple channels are discussed first, followed by grating-element patterns and the tuned complex channels. The intrusion of channels other than the tuned channels will be discussed after the initial presentation of experimental results.

Predictions—square-element patterns and simple channels

The use of square-element patterns to explore summation in simple channels is illustrated in Fig. 3. The output of a tuned simple channel is shown in response to three stimuli (three columns). Figure 3 (top row) shows explicitly only a small portion of the striped region of each stimulus and shows the receptive field of the

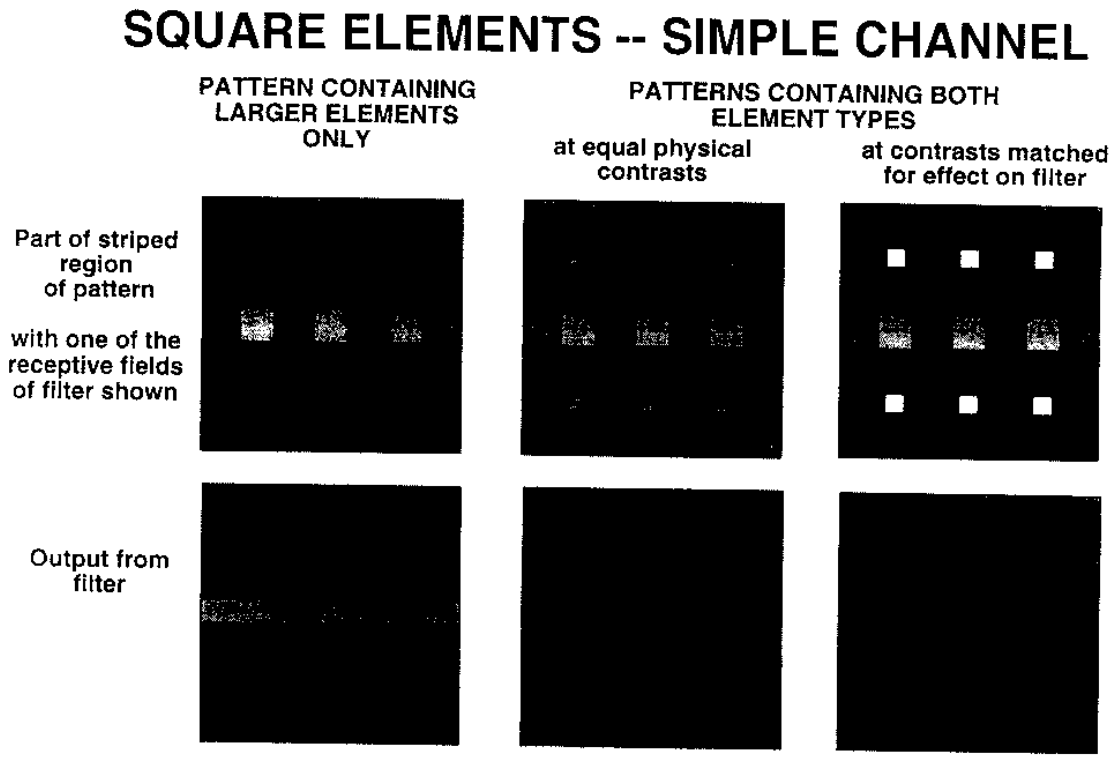

FIGURE 3. Diagram illustrating use of square-element patterns to investigate spatial summation in simple channels. The top row shows a section of the striped region of the pattern with a receptive field from the tuned simple channel superimposed. The bottom row shows the output of the channel. Three stimuli are represented in the three columns (as labeled).

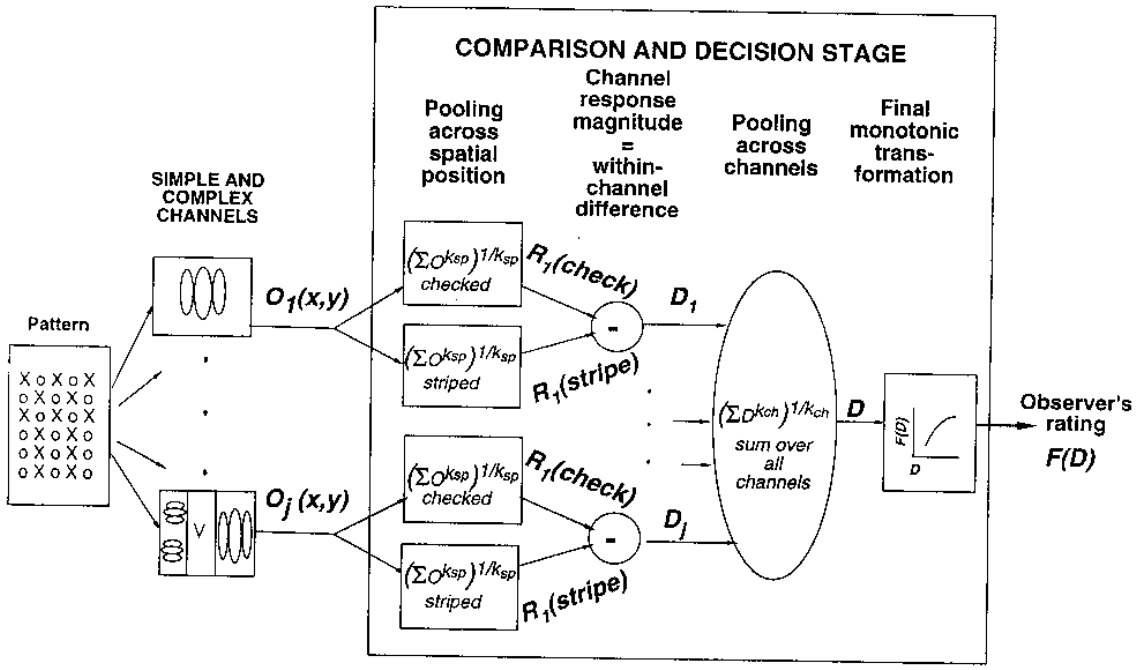

FIGURE 4. Diagram of the comparison and decision stage we use for the task of texture segregation. These are simplified rules representing all higher levels of visual processing. Symbols are defined and equations for the computations are given in Part II of Appendix I.

corresponding tuned simple channel. The second row shows the output of the tuned simple channel to these three stimuli. Mid-gray represents a zero response, brighter areas represent positive responses, and darker areas represent negative responses. Notice that this tuned channel produces modulated output to both the large-element-only stimulus (left) and to the stimulus containing both elements at equal physical contrasts (middle) but not to the stimulus in which the elements have the same contrast-area product and thus are matched for effectiveness (right column).

In the checkerboard region, the output of this particular tuned channel is unmodulated for all three stimuli shown, indeed for any contrast ratio of square elements. (If shown here, it would simply be a gray square for all three stimuli. Since the channel's receptive field is horizontally oriented, both types of elements in the checkerboard region fall within the excitatory center, and also both fall within the inhibitory surround. Hence, the net response is close to zero.) Thus, using the output from this tuned channel, the observer could perceptually segregate the different regions in the stimuli in the left and middle columns of Fig. 3, but could *not* segregate the stimulus in the righthand column.

Note that the simple channel in this diagram is not the only simple channel tuned to this pattern: analogous reasoning would apply, for example, to an obliquely-oriented filter tuned to the checkerboard region; but that filter would produce modulated responses in the checkerboard region of the appropriate stimuli (the same stimuli to which the filter in Fig. 3 produces modulated responses in the striped region), while never producing modulated responses in the striped region.

From channel outputs to the observer's response—the comparison-and-decision stage. To use the outputs of channels (or any other intermediate entities one is investigating) to quantitatively predict the response of the observer, one always needs assumptions linking the channels' outputs to the observer's response. We have generally considered a family of rules which includes the rules used by many other investigators. These rules compute various measures of the degree to which there are gross differences in overall activity between the channel outputs in the checkerboard region and the channel outputs in the striped regions. Figure 4 shows these rules diagrammatically, calling them a "comparison and decision stage". For models containing only simple channels this is approximately saying that two textures can be segregated if, and only if, they differ sufficiently in the amount of some spatial-frequency/orientation they contain (a generalization of the original Julesz, 1975 conjecture, not unlike the corresponding stage in so-called "energy" or "sum-of-quadrature" models, see, e.g. Bergen & Landy, 1991 or Clark, Bovik, & Geisler, 1987). Notice that we are ignoring the very interesting questions of whether the boundary between regions is found or the regions are growing or some combination of both (e.g. Lee, 1995), as well as the interesting asymmetries in performance (e.g. Beck, 1973; Gurnsey & Browse, 1987; Rubenstein & Sagi, 1996). Further description of our comparison and decision stage, and discussion of its implications and relationship to others'

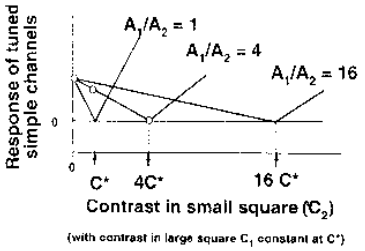

FIGURE 5. The output of the tuned simple channels in response to square-element patterns. The contrast of the larger elements C_1 is held constant at C^*. The contrast in the smaller elements C_2 is shown on the horizontal axis. The three curves are for three different area ratios, where A_1 is the area of the larger elements (always the same in this study) and A_2 is the area of the smaller elements (which varied). The three open circles represent the stimuli used in the three columns of Fig. 3.

models can be found in Parts I and II of Appendix I here and in Graham *et al.*, 1992 (with further discussion in Sutter at al., 1989, Graham, 1991 and the Appendix in Graham *et al.*, 1993).

For the first part of this paper, we will make the simplifying assumption that the observer's response is entirely determined by the tuned channel or channels (more generally a group of channels all acting proportionally to the tuned channels). Then the comparison and decision stage reduces to the following:

1. Consider, without loss of generality, a single one of the group of tuned channels; we will refer to this as *the* tuned channel.
2. The magnitude of the tuned channel's response is the amplitude of the modulated output in one region (e.g. the striped region) minus that in the other (the checkerboard region). This quantity is labeled *the within-channel difference* in Fig. 4
3. The observer's rating of perceived segregation is a monotonic function (F in Fig. 4) of the magnitude of the tuned channel's response.

Note that, although we are currently considering the case of tuned simple channels, this same derivation will apply for the case of tuned complex channels below. (The effects of untuned channels and of the intensive nonlinearity will be returned to later.)

Predicted position and magnitude of dips in response-vs-contrast curves for square-element patterns. Predictions from the three cases in Fig. 3 appear as the three open circles in Fig. 5. Figure 5 shows the magnitude of the response from the tuned simple channels (on the vertical axis) as a function of contrast in the smaller elements C_2 (horizontal axis) when the contrast in the large elements was held constant at $C_1 = C^*$ (where C^* is not zero). Several different ratios of element areas are represented in the different curves (A_1 and A_2 are the areas of the large and small elements, respectively). The most important feature to note is that the predicted response of the tuned simple channels falls to zero when the area-times-contrast products of the two elements are equal, or equivalently, when

$$\frac{C_2}{C_1} = \frac{A_1}{A_2} \tag{2}$$

For example, when the large square has four times the area of the small square, then the minimum occurs when the small square has four times the contrast of the large.

Equation (2) is a consequence of the linearity of the simple channels. It applies to the patterns used in this study, where the squares are exactly as wide as the inter-square spaces (where the duty-cycle is one-half). In our previous study (Sutter *et al.*, 1989), where we computed predictions from a full simple-channel model involving a whole array of channels, the predicted dip occurred at a slightly different value because the squares were somewhat wider than the inter-square spaces.

Predictions—grating-element patterns and complex-channels

Figure 6 illustrates the logic of the experiments using patterns composed of grating-patch elements. Portions of the striped regions from three such stimuli are shown in the three columns of Fig. 6. All three stimuli are composed of large and small patches differing in area by a factor of 4. The contrast in the small grating patches is zero (left column), or physically equal to that of the large elements (middle column), or is the contrast which matches the large and small grating patches for effectiveness on the second-stage filter of the tuned complex channel (right column). Note that, if the complex channel's intermediate nonlinearity is either conventional half-wave or conventional full-wave rectification (both piecewise-linear functions), then this match for effectiveness will occur when the area-times-contrast products of the two patches are equal. However, if the pointwise intermediate nonlinearity is a power function with power unequal to 1, this matching contrast could be quite different (as will be described below in connection with Fig. 7).

The stimuli in Fig. 6 cannot be segregated by simple channels, since the Fourier amplitude spectra in the two regions are very similar (see Graham *et al.*, 1992 and Graham, Sutter, & Venkatesan, 1993). These stimuli can, however, be segregated by complex channels. The complex channel whose outputs are illustrated in Fig. 6 is the one "tuned" to the striped region of this set of stimuli. The top row shows explicitly only a small portion of the striped region of each stimulus, with a receptive field from the first filter of the tuned complex channel superimposed. This first filter is tuned to the grating elements (and thus to a relatively high spatial frequency and, in this example, a horizontal orientation).

The second row of Fig. 6 illustrates the output of the complex channel's first filter. The third row shows the output after the pointwise intermediate nonlinearity. It also shows a receptive field from the second stage superimposed. The complex channel's second filter is tuned to the fundamental frequency of the striped region (hence a relatively low spatial frequency and, in this example, a horizontal orientation). And finally, the fourth

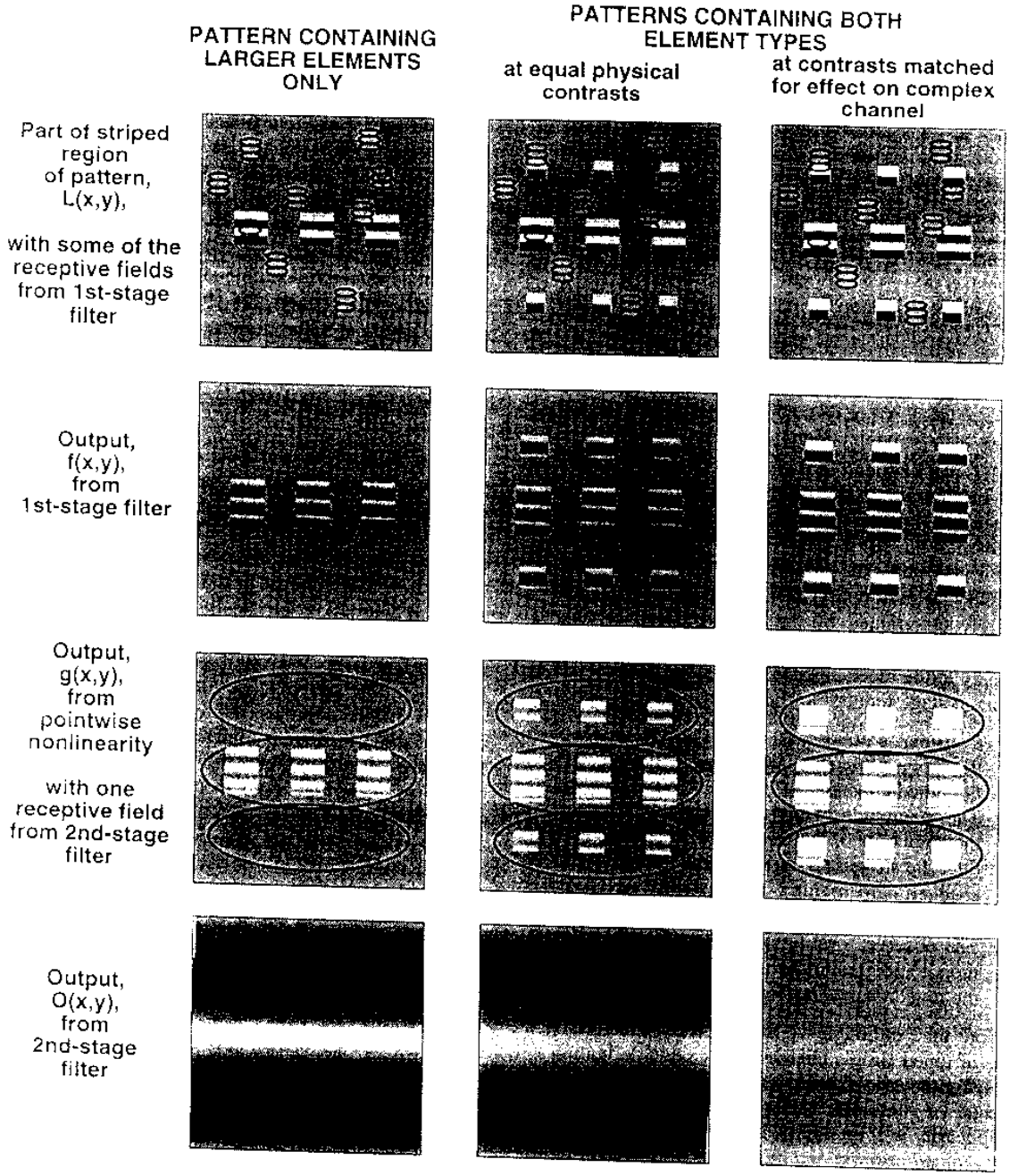

FIGURE 8. Diagram illustrating use of grating-patch-element patterns to investigate spatial summation in complex channels. The top row shows a section of the striped region of the pattern with a receptive field from the first filter of the tuned complex channel superimposed. The second row shows the output of the first filter $f(x,y)$. The third row shows the output $g(x,y)$ after the intermediate nonlinearity with a receptive field of the second filter superimposed. The bottom row shows the output of the channel. Three stimuli are represented in the three columns (as labeled).

row illustrates the overall output from the complex channel, that is, the output from the second filter. The modulation in this output is most pronounced in the left column (large patches only). It is reduced in the middle column by the presence of the small elements. It is maximally reduced in the third column when the contrasts of the small and large patches are matched for effectiveness. If the contrast in the small patches were increased still further (not shown), the modulation would again increase but in the opposite phase.

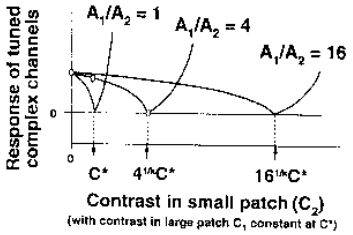

FIGURE 7. The output of the tuned complex channels in response to grating-element patterns. The contrast of the larger elements C_1 is held constant at C^*. The contrast of the smaller elements C_2 is shown on the horizontal axis. The three curves are for three different area ratios, where A_1 is the area of the larger elements (always the same in this study) and A_2 is the area of the smaller elements (which varied). The three open circles represent the stimuli used in the three columns of Fig. 6. The curvature shown here is appropriate for an expansive function ($k > 1$). For $k = 1$, the lines will be straight (exactly like those in Fig. 5 in fact) and for $k < 1$ (compressive functions) the curvature will be in the opposite direction. The location of the dips also depends on k. It is given by equation (3) and is indicated by the labels on the horizontal axis here.

In the checkerboard region the output of this particular tuned channel would be unmodulated for all three stimuli shown, and would be so for any contrast ratio of grating elements. (If shown here, the output would simply be a gray square. Remember that both types of elements in the checkerboard region fall within the excitatory center of the horizontal receptive field, and also both fall within the inhibitory surround. The net response, therefore, is close to zero.) Thus, using the output from this tuned channel, the observer could perceptually segregate the different regions in the stimuli in the left and middle columns of Fig. 6, but could *not* segregate the stimulus in the righthand column.

To get from this channel output to the observer's response requires the comparison-and-decision stage we have already discussed in connection with the square-element experiments (see Fig. 4).

Predicted position of and magnitude of dips for grating-element experiments. Figure 7 shows the magnitude of the tuned complex channels' response (the amplitude of modulation in its output) as a function of small-patch contrast (following the same conventions as Fig. 5). Predictions from the three cases in Fig. 6 appear as the three circles in Fig. 7. Just as in Fig. 5, the curves in Fig. 7 dip to a minimum of zero. In Fig. 7, however, the minimum no longer necessarily occurs when equation (2) holds, and the functions are not necessarily straight lines.

Rather, the match occurs whenever

$$\frac{C_2}{C_1} = \left|\frac{A_1}{A_2}\right|^{(1/k)} \tag{3}$$

Equivalently, since C_1 is held constant at C^*, whenever

$$C_2 = C^* \cdot \left|\frac{A_1}{A_2}\right|^{(1/k)} \tag{4}$$

The parameter k in equation (3) and equation (4) equals the exponent k in the power function at the intermediate stage of the complex channel [equation (1)].

When $k > 1$ (expansive intermediate nonlinearity), the position of the dip in Fig. 7 moves *less* dramatically to the right than it does for $k = 1$. This happens for the following reason: with an expansive nonlinearity, a somewhat larger contrast in the smaller than the larger patch translates into a much larger response at each point in the smaller than the larger patch. For example, to compensate for an area-ratio of 4 when $k = 2$, the small-patch contrast need only be $4^{1/2} = 2$-times the large-patch contrast.

If, on the other hand, the intermediate nonlinearity is compressive ($k < 1$), the position of the dip will move *more* dramatically to the right than in the linear case. To compensate for an area-ratio of 16, the small patch will need to have more than 16-times the contrast of the large patch because the response at each point in the small patch is compressed more than in the large patch. For example, for an area-ratio of 4, when $k = 1/2$, the small-patch contrast must be $4^2 = 16$ times the large-patch contrast.

The curvature as well as the location of the dip depends on k. The curvature shown in Fig. 7 is that appropriate for an expansive function ($k > 1$). Curvature in the opposite direction is predicted by compressive functions.

Some further justification of the prediction in equation (3) is given in Part III of Appendix 1.

EXPERIMENTAL METHODS AND PROCEDURES

It will be useful to let the words "pattern" and "stimulus" take on specific meanings. The word "pattern" will mean particular spatial characteristics without specifying contrast. For example, we might speak of the "pattern" in which both elements are squares of a particular size without implying anything about the contrasts. The word "stimulus" will mean a particular pattern with contrasts specified. For example, we might speak of the "stimulus" in which both elements are squares of a particular size and the contrasts of the two element types have been set at 5 and 10%, respectively. (The background luminance, temporal characteristics, and viewing distance are constant in the study reported here, with one exception explicitly noted later, and thus do not need to enter this discussion.)

There were eight different patterns seen by each observer in the main experiment. They were the 2×4 combinations of two kinds of elements (square and grating patches) with four different ratios of element areas (where the larger elements were 1, 1.78, 4, and 16-times the area of the smaller ones).

For each of the 8 patterns, 114 different contrast combinations were used. These combinations are diagrammed in Fig. 8, where each symbol represents a contrast combination that was used. Thus, there were $8 \times 114 = 912$ different stimuli in the experiment. In general, each observer saw each stimulus 4-times (distributed across eight sessions as described below) for a total of $4 \times 912 = 3648$ trials.

There were seven observers.

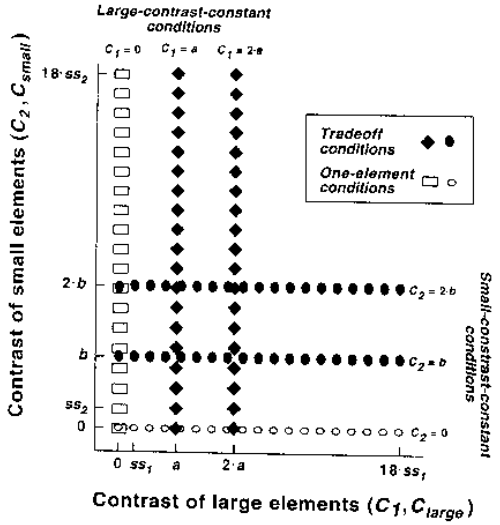

FIGURE 8. Each symbol represents one of the 114 contrast combinations at which any given pattern was presented. The one-element conditions are shown by open symbols and the tradeoff conditions by closed symbols. Contrasts were equally spaced, and the values of the contrast increments for the various conditions (the parameters a, b, ss_1, and ss_2) are given in Table 1 for each of the eight patterns (4 area ratios × 2 element types).

Details of the stimuli

Spatial characteristics of the patterns. The numbers, spacing, and arrangements of the elements can be see in Fig. 2.

For four of the seven subjects (cv, jh, ws, and nh), the orientation of the grating-patch elements was as shown in Fig. 2 with the bars of the gratings parallel to the stripes, or, in other words, the local and global orientations were consistent. For the other three subjects (ch, lz, sj), the orientation of the grating-patch elements was perpendicular to that shown in Fig. 2.

The center-to-center spacing between neighboring elements was 32 pixels (0.67 deg at the viewing distance of 0.91 m).

The repetition period (within a given region) was two rows and two columns of elements and was 1.33 × 1.33 deg (64 × 64 pixels). Thus the fundamental frequency (the reciprocal of the repetition period) of either the checkerboard or the striped region was 0.75 c/deg both horizontally and vertically.

Both element types were solid squares, or both were circular Gabor patches. The two element types in a given pattern were generally of different widths (and areas). The width of a Gabor-patch element was taken to be the full width at half peak of the circular gaussian envelope. The width of the larger element type in a pattern was always 16 pixels (0.33 deg at the viewing distance of 0.91 m) and that of the smaller type was either 16, 12, 8 or 4 pixels (0.33, 0.25, 0.17, or 0.08 deg). Thus patterns of four different area ratios were used: $A_1/A_2 = 1, 1.78, 4,$ and 16, where A_1 is the area of the larger element and A_2 the area of the smaller. (For ease of expression, the area ratio of 1.78 will sometimes be referred to as 2.)

The Gabor patch elements were actually truncated to be a square with width and height of 32 pixels so as not to overlap with the neighboring elements. The harmonic oscillation in the Gabor function was in sine phase with respect to the window so that the space-average luminance across each Gabor element was the same as the background luminance. The orientations of all the patches were either vertical or horizontal (as described above). The spatial frequency of the harmonic oscillation was always 12 c/deg (a period of 4 pixels). In order to insure their visibility in Fig. 2, the period in the grating patches used in that figure was 8 pixels; thus each patch in Fig. 2 contains half as many cycles as the stimuli in the experiments.

Contrast of the stimuli. Figure 8 shows the full set of contrast combinations used for any particular pattern. Each axis of Fig. 8 gives the contrast of one element type:

$$C_i = \Delta L_i / L_{bkd}$$

where L_i is the luminance of the elements of type i, L_{bkd} is the luminance of the background and $\Delta L_i = L_i - L_{bkd}$. Here C_1 will refer to the contrast of the large elements and will sometimes be called C_{large}; it is plotted on the horizontal axis. Similarly, C_2 will refer to the contrast of the small elements and will sometimes be called C_{small}; it is plotted on the vertical axis.

The full set of contrast combinations contains six subsets, each containing 19 levels of contrast in one of the element types while the contrast in the other was fixed. There were a total of $6 \times 19 = 114$ contrast combinations.

In the three *large-contrast–constant conditions*, the contrast in the large element is held constant, while the contrast in the small element is varied. Analogously there are three *small-contrast–constant* conditions.

Two of the six subsets will occasionally be referred to as *one-element conditions* since only one type of element is visible in these conditions (open symbols in Fig. 8). The other four subsets (closed symbols in Fig. 8) will be referred to as *tradeoff conditions*.

Contrasts are given in integral multiples of several quantities—a, b, ss_2 and ss_1—and the values of these increments for different patterns and different observers are given in Table 1. The grating-element contrasts were set higher for observers ch and sj than for the other observers because, in preceding experiments, their sensitivity to grating elements had been lower.

Viewing conditions and background luminance. The luminous screen was approximately 16 cm high and 21 cm wide, which, at the viewing distance of 0.91 m, was 10 × 13 deg of visual angle. The background luminance of the screen was constant throughout each experiment (during stimuli, fixation points, and inter-stimulus intervals) at 18 ft-L.

The observer viewed the screen binocularly while sitting in a chair with unrestrained head and natural pupils. The chair was set so that the distance between the eye and the screen was 0.91 m initially (presumably changed a few centimeters by unintended shifts in head

position). There was a small lamp 6 feet behind the observer which— along with the CRT screen itself— provided some ambient illumination in the room.

Equipment and calibration. The patterns were generated and the experiments run by a Macintosh IIci on a standard Apple monochrome monitor using Pascal programs built upon programs kindly supplied by Hugh Wilson. Based on calibrations with a uniform field, there were 150 linearized gray levels available. The background luminance of all our stimuli was set at the midpoint and hence the smallest contrast step was nominally 1/75 or 1.3%. The actual contrast of the smallest squares and of the Gabor patches (based on the nominal contrast step of 1/75) may have been slightly lower than those calculated from these calibrations (and reported in Table 1) due to the spatial characteristics of the monitor, but the actual contrast values are irrelevant for the conclusions here.

TABLE 1. Element widths and contrast-increment parameter values (columns) for the eight patterns (rows: 4 area ratios × 2 element types)

Area ratios	Element widths (in deg visual angle)		Contrast-increment parameters (in proportion contrast)			
A_1/A_2	W_1	W_2	a	ss_2	b	ss_1
Grating-element patterns						
1	0.33	0.33	0.16 {0.24}	0.053	0.16 {0.24}	0.053
2	0.33	0.25	0.16 {0.24}	0.053	0.21 {0.27}	0.053
4	0.33	0.17	0.16 {0.24}	0.053	0.27 {0.32}	0.053
16	0.33	0.08	0.16 {0.24}	0.053	0.43 {0.48}	0.053
Square-element patterns						
1	0.33	0.33	0.04	0.013	0.04	0.013
2	0.33	0.25	0.04	0.027	0.08	0.013
4	0.33	0.17	0.04	0.040	0.12	0.013
16	0.33	0.08	0.04	0.053	0.24	0.013

See Fig. 8 for diagram of the stimuli used for each pattern. Numbers inside brackets are the values for subjects ch and sj in the cases where their values differed from other subjects'.

Details of the experimental procedures

Structure of the sessions. Practical considerations prevented all the 912 different stimuli (2 kinds of elements × 8 area-ratios × 114 contrast combinations) from being presented in a single session. Instead, a single session contained trials of only half of the stimuli. Each of these 456 stimuli were presented once in the session randomly intermixed. Each observer participated in eight sessions, during the course of which he/she saw each stimulus four times.

For observers cv, jh, nh, and ws, both grating-element and square-element stimuli appeared in the same session but only half the contrast combinations for each were used. The small-contrast–constant conditions were presented in some sessions and the large-contrast–constant conditions were in others. For ch, lz, and sj, the square-element stimuli were presented in some sessions, and grating-patch-element stimuli were presented in others. No difference due to this difference in the structure of the sessions was found, and thus it will be mentioned no further.

Structure of a trial and the response scale. Each trial started when the observer pressed the top inch of a response device (an "Unmouse"). A small fixation pattern then appeared for 1 sec in the middle of the screen. Immediately after the fixation pattern's offset, the stimulus was presented for 1 sec with an abrupt onset and offset.

After stimulus offset, a 1-sec delay occurred and then a beep signaled that the observer could make a response by pressing the appropriate position within a rectangle (about 10 cm wide × 2.5 cm high) on the response device. We used the 1-sec delay between the stimulus offset and the observer's response for reasons briefly discussed in Graham, Sutter, & Venkatesan (1993) and Sutter & Graham (1995). After the observer's response, there was a double-beep.

Although the responses were actually recorded on a finely divided scale (from 0 to 100, as the observers knew), five equally spaced numerals were written on the face of the response device to guide their responses. A sheet of paper was available whenever they wished to look at it stating that the meaning of these numerals were:

0 No segregation between the regions
1 Barely perceptible segregation between regions
2 Perceptibly segregated regions
3 Moderately segregated regions
4 Highly segregated regions.

Observers and instructions. There were seven observers, all of whom were students at Columbia University. All the observers were naïve as to the purpose of these experiments at the time they ran them, but all of them had participated in related texture-segregation experiments prior to running in these. Several of these experiments have been published: The observers ws and cv appear in Graham & Sutter (1996); cv also appears in Graham *et al.* (1993); observers ch (called cs there) and jh appear in Sutter & Graham (1995).

Before participating in their first segregation-rating experiments, the observers all received 15–30 min of instructions including a series of practice patterns. They were told to maintain fixation at the center of the screen (even after the fixation mark had disappeared) and to indicate by their response the degree to which the regions immediately and effortlessly segregated. They were explicitly instructed NOT to focus on the individual squares or any other form of local information and NOT to indicate the result of scrutinizing the patterns for differences. They were asked to ignore factors such as the overall size and the degree of brightness of the pattern. They were also instructed to maintain a focus of attention that was global (while, however, continuing to fixate the center of the screen).

All observers either had 20/20 vision without correction or wore eyeglasses or contact lenses of current prescription.

Calculation of summary numbers from the experimental results

For each observer, for each of the stimuli, the average

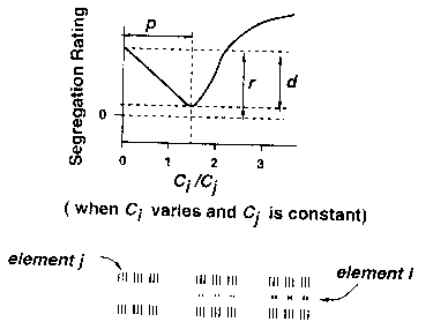

FIGURE 9. Schematic drawing showing the dip in typical results from tradeoff conditions. The vertical axis shows segregation rating and the horizontal axis contrast ratio c_i/c_j, where c_j is fixed at some non-zero value. The magnitude of the dip is d/r where r is the segregation rating when only element j is present, and d is the maximal decrease in rating when element i is added at various contrasts. The position p of the dip is the contrast ratio at which the lowest segregation rating occurs. In this figure $d/r = 0.8$ and $p = 1.5$.

segregation rating was computed over the four repetitions of the stimulus and then the quantities described in the following paragraphs were calculated using these average ratings.

The locations and magnitudes of the minima (dips) in the tradeoff conditions. When results from a typical tradeoff condition are plotted, the curves show dips as in the prototype of Fig. 9. (Examples of experimental results plotted this way are shown in Fig. 10.) To summarize each of these curves, we use two quantities: (1) the magnitude of the dip (as a proportion of the excursion of the curve, d/r in Fig. 9), and (2) the position of the dip on the horizontal axis (the contrast ratio at which the lowest point of the curve occurred, p in Fig. 9).

To compute the magnitude of the dip d/r we took r to be the average segregation rating at a contrast ratio of zero, and d was determined by the very lowest point in the averaged data curve.

We measured the position of the dip, p, as the midpoint of the range of contrasts leading to very low segregation ratings (ratings for which d/r was within 10% of the maximal d/r). This way of measuring the dip position was generally the same as estimating the position from the single lowest point in the data curve. It only differed in the cases where the data produced large flat-bottomed dips, and then the single point that happened to be numerically lowest was sometimes very misleading (at the end of the contrast range, for example). If the estimated relative magnitude of the dip (d/r) for a

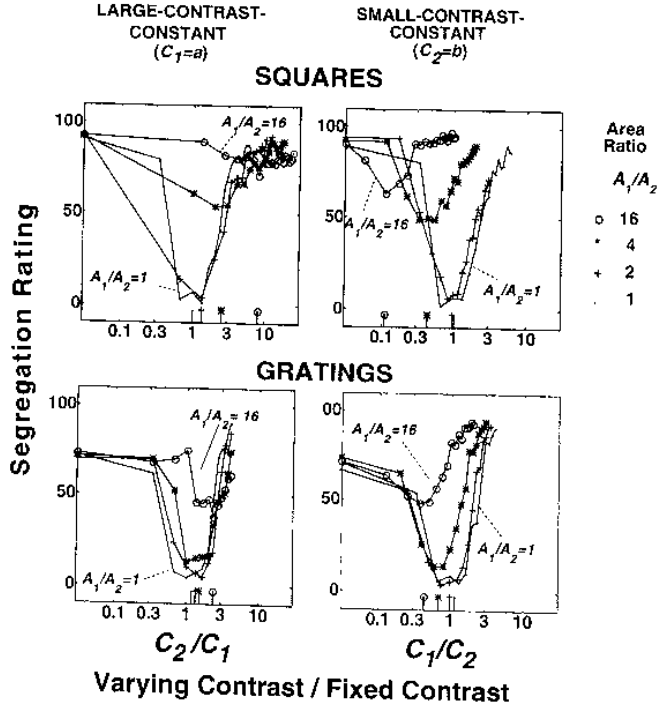

FIGURE 10. Segregation-versus-contrast curves for the average observer (that is, the curves show the results averaged across the seven observers) from two tradeoff conditions: the large element's contrast is constant at a in the left column and the small element's contrast is constant at b in the right column. Results for square-element patterns (top row) and grating-element patterns (bottom row) are shown. Each curve is for a different ratio of element areas, where A_1 is the area of the larger elements (always the same in this study) and A_2 is the area of the smaller elements (which varied). The horizontal axis shows the ratio of the contrasts in the two element types; the numerator is the contrast of whichever element had varying contrast in that condition. The small vertical lines on the horizontal axis show the calculated positions of the dips in each curve.

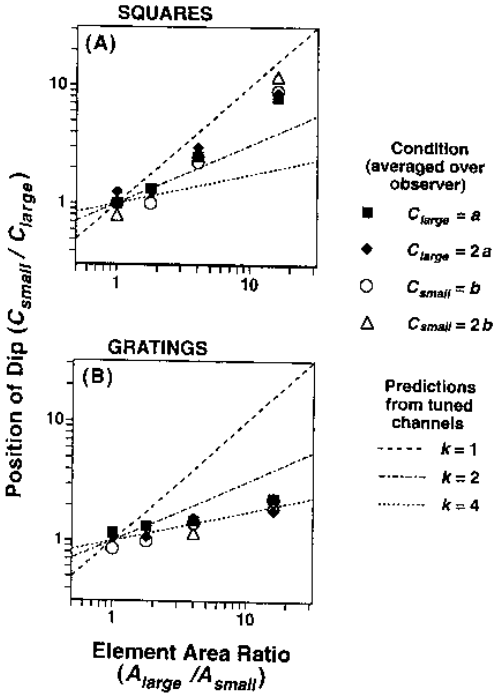

FIGURE 11. Position of the dips plotted as a function of area ratio in the segregation rating versus contrast curves for the average observer (e.g. Figure 10). Different symbols represent the four tradeoff conditions with the results for square-element and grating-element patterns in (A) and (B), respectively. The lines show the predicted positions of the dips for various values of the power k, assuming that the tuned channels determine the observer's response. For the squares the data are near the predictions for $k = 1$. For the gratings, the data points fall near the predictions for $k = 3$ (not shown) or 4 (dotted line), indicating an expansive intermediate nonlinearity in the complex channels.

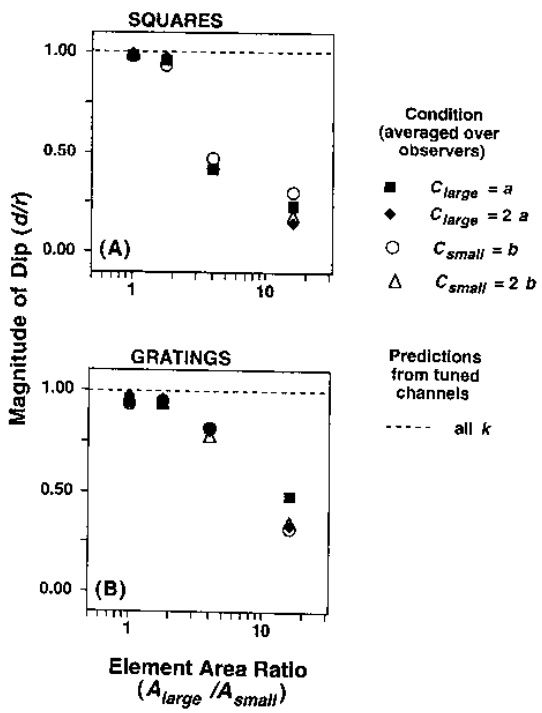

FIGURE 12. Magnitude of the dips plotted as a function of area ratio from the segregation rating vs contrast curves (e.g. Figure 10) for the average observer. Different symbols represent the four tradeoff conditions with the results for square-element and grating-element patterns in (A) and (B), respectively. The line shows that the predicted magnitude of the dips is 1.0, no matter what the value of k, if the tuned channels determine the observer's response. The observed values depart from 1.0 at large area ratios, more dramatically for square-element than for grating-element patterns.

particular observer and session was less than 0.1, the position of the dip was not entered into any of the averages, as it seemed more likely to contribute variability than information.

For each observer, for each of the eight patterns (4 area ratios × 2 element types), for each of the four tradeoff conditions, both the dip's position and magnitude were estimated. Figures 11 and 13 will show the positions, and Figs 12, 14, and 16 the magnitudes.

Segregation thresholds in the one-element conditions. For each observer, for each of the eight different kinds of *elements* used in these experiments (grating vs square, four different areas), a "segregation threshold" was calculated. The segregation threshold for a particular kind of element (e.g. the smallest squares) is based on the results from the *one-element condition* (open symbols, Fig. 8) containing the set of stimuli composed only of that kind of element (but of varying contrast). The segregation threshold is the contrast (found by interpolation) of that one-element stimulus to which the observer would give a mid-scale segregation rating (a rating of 50). The ratio of the thresholds for the smallest and largest element will be presented below in Fig. 16 (where C_S^* and C_L^* are the values of the segregation threshold C_0, for the smallest and largest elements, respectively).

RESULTS FROM TRADEOFF CONDITIONS

Figure 10 shows results from the "average observer" (i.e. the results averaged across all seven observers) for both grating-element and square-element patterns (bottom and top panels, respectively) for all four area ratios (four different curves in each panel) from two of the tradeoff conditions (for C_{large} = a in the left panels and $C_{small} = b$ in the right panels). The vertical axis is average segregation rating (averaged across observers as well as across stimulus repetitions), and the horizontal axis is the ratio of contrast in the varying element divided by the contrast in the fixed element.

The calculated position of the dips in these curves is indicated by the small vertical lines in the bottom of each panel. (See Methods for details of calculation.) Notice that in the large-contrast-constant conditions (left panels), the dip moves more dramatically to the right (that is, moves to greater values of C_{small}/C_{large}) for square

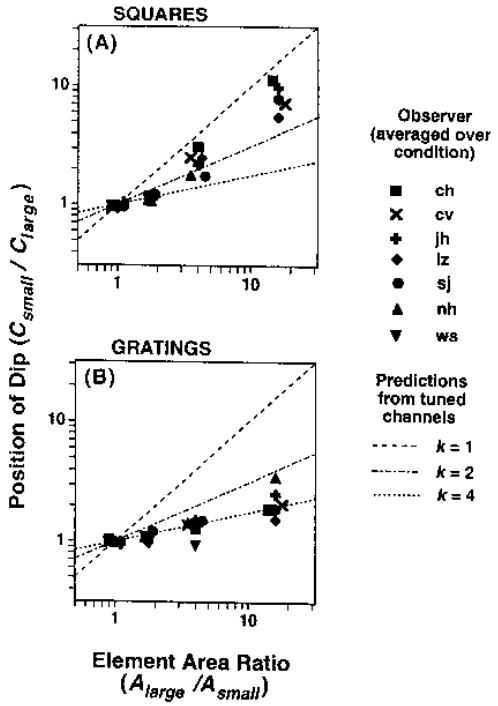

FIGURE 13. Positions of the dips for individual observers (averaged over the four tradeoff conditions) in the same format as Fig. 11.

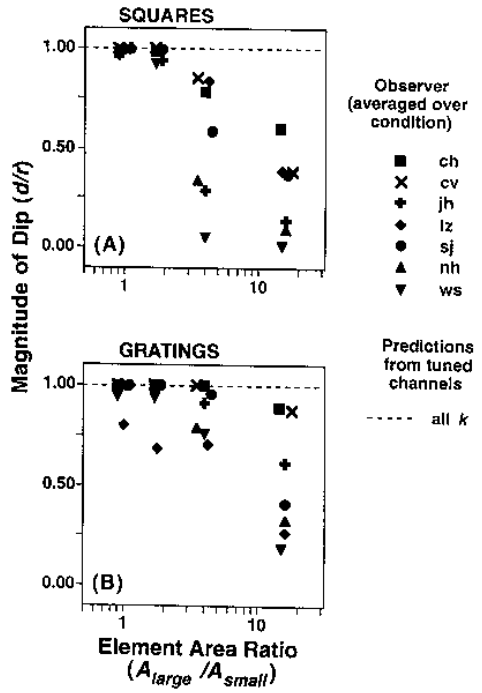

FIGURE 14. Magnitudes of the dips for individual observers (averaged over the four tradeoff conditions) in the same format as Fig. 12.

elements (top row) than for grating elements (bottom row). Analogously, in the small-contrast–constant conditions (right panels), the dip moves more dramatically to the left (which is again a movement to greater values of C_{small}/C_{large}) for the square than for the grating elements.

Figure 11 (as squares and circles) shows the positions of the dips from the curves shown in Fig. 10 and also (as diamonds and triangles) from the two tradeoff conditions using higher constant contrasts. The position of the dip is given on the vertical axis (expressed as C_{small}/C_{large}). The area-ratio is given on the horizontal axis (expressed as A_{large}/A_{small}).

Notice in Fig. 11 that the results from the four tradeoff conditions superimpose quite well. Also shown are lines representing equation (3) for $k = 1$ (top line), 2, and 4. The results from the square-element patterns (top panel) hover between the $k = 1$ and $k = 2$ line. For an area ratio of 16, for example, the dip occurs at a contrast ratio of about 10 (rather than 16 as on the $k = 1$ line or 4 as on the $k = 2$ line).

The results from the grating-element patterns (bottom panel) are very close to the $k = 4$ line. For example, for an area ratio of 16, the contrast ratio at the dip is only 2. To put this another way, in order to compensate for a 16-fold difference in area, the contrast of the small grating patch only needs to be twice that of the large.

Figure 12 shows the magnitudes of the dips (d/r) for the average observer (e.g. Figure 10) in the different conditions in much the same format as Fig. 11 showed the positions. Notice the magnitudes of the dips for the average observer become substantially smaller as the area ratio becomes larger.

Figures 13 and 14 show the positions and magnitudes of the dips for individual observers (averaged over condition) in the same formats as Figs 11 and 12. The results for individual observers resemble those for the average subject in both the positions and magnitudes of the dips, although there is more scatter. Much of the increased scatter in the position of the dips and some of the scatter in the magnitudes of the dips can be attributed to the smaller amount of data entering each point. However, the rest of the scatter, which is considerable in the case of magnitudes, is the result of systematic individual differences. This is easy to verify by looking at the curves (not shown here to save space) like those in Fig. 10 but for individual observers in individual sessions—the differences in dip magnitudes are clear in individual sessions. We will return to these individual differences below.

A further detail—there is no evidence for any interaction between observer and condition. To put it another way—suppose one looks at graphs like those of Figs 11 and 12 but with the average observer replaced by each individual. There will necessarily be considerable scatter among the points, since relatively small amounts of data are now entering into each point. However, to the extent allowed by this considerable intrinsic variability,

an individual's results for all four tradeoff conditions appear to be the same.

Predictions from tuned channels

Let us compare these results in Figs 11, 12, 13, and 14 to the predictions from the tuned channels. (Remember that the "tuned channels" are the channels for which the pattern under discussion is the best stimulus. All other channels are called "untuned channels.")

The measured dip positions in Figs 11 and 13 are at least approximately consistent with the predictions from the tuned channels: the dip positions for square elements resemble the predicted linear behavior for simple channels (Fig. 5) although the $k = 1$ line is systematically missed in Figs 11 and 13. And the dip positions for grating elements in Figs 11 and 13 are as predicted by complex channels having an intermediate power–law nonlinearity with a power somewhat higher than 2, perhaps 3 or 4 (Fig. 7).

However, the measured dip magnitudes in Figs 12 and 14 are not even approximately consistent with the predictions from the tuned channels. The predicted magnitudes (expressed as d/r) are always 1.0, since the predicted curves dip to 0 in both Figs 5 and 7. At large area ratios, this predicted magnitude is far larger than the measured magnitudes. Further, as mentioned above, the magnitudes of the dip vary quite substantially from observer to observer.

Thus, the tuned channels by themselves cannot explain all features of the data—in particular, they cannot account for the dip magnitudes. One ordinarily assumes, however, that all responsive channels affect the observer's response to some extent. So the question becomes: on the basis of our existing knowledge, are there plausible channels (other than the tuned channels) that respond enough to these patterns to influence the observer's response? And, if so, will the intrusion of these other channels explain the discrepancies between the tuned channels' predictions and the individual differences in the responses?

Untuned (off-frequency) simple channels cannot explain the discrepancies

In our earlier paper (Sutter *et al.*, 1989) we conducted an extensive computational study of the predictions for square-element patterns from models containing simple channels tuned to different spatial frequency and orientations across the full visible range. We investigated a number of different candidate rules for pooling across spatial position and also for pooling the responses from different simple channels (briefly described here in Part II of Appendix I). As it turned out, the predictions were always dominated by the simple channels tuned to the pattern in question; in particular, the dip magnitudes and positions were those predicted from the tuned channels alone (even though the model allowed all other simple channels to influence the observer's response).

We have never conducted as extensive a computational study of simple channels' responses to grating-element patterns because near-zero responses are expected. The computations we have done, however, show the expected result: simple channels respond little if at all to element-arrangement patterns made of grating elements or other luminance-balanced elements (elements in which the average luminance equals that of the background). Some relevant computations are described in Graham, Beck, & Sutter (1992) and Graham, Sutter, & Venkatesan (1993).

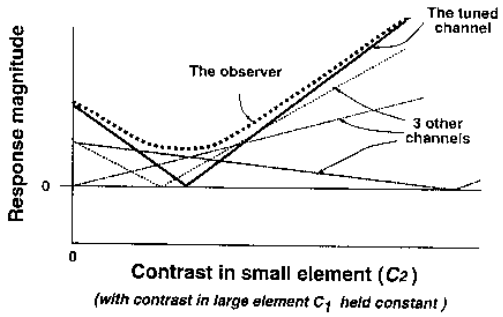

FIGURE 15. The observer's response (as a function of contrast ratio) is shown as the approximate envelope of the response of several channels: one tuned channel and three "other" channels that are somewhat sensitive to the patterns. This sketch is appropriate for both square-element patterns and grating-patch-element patterns.

In short, the discrepancies between the predictions from the tuned channels and the observer's segregation ratings as measured here (for both square or grating elements) cannot be explained by the action of other (untuned) simple channels.

Complex channels can explain the discrepancies for square-element patterns

The results here for square-element patterns replicate some of those in our earlier study (Sutter *et al.*, 1989). There we suggested that complex channels might explain the discrepancy between the simple model predictions and the experimental results, in particular, that they might explain the fact that the observed dip magnitudes were less than 1.0. The appropriate complex channels are similar to that in Fig. 6: they have first filters that are sensitive to relatively high spatial frequencies (in this case to the edges of the square elements rather than to the dominant frequency of a grating patch) and second filters that are sensitive to the period of the striped or checkered pattern. We still think these complex channels the likely explanation of discrepancies. To illustrate the intrusion of these complex channels, Fig. 15 shows response magnitude versus contrast-ratio curves (like the earlier ones in Fig. 5 and 7) not only for the tuned channel (thick solid line) but also for three other channels (thick dashed and dotted curves). There are a number of complex channels tuned to different first-stage frequencies, and, in general, they will have different dip positions (although a bias in these positions will be discussed below). Figure 15 shows the observer's response (thick dashed line) as the approximate envelope of individual channels' responses.

and this qualitative assumption will be sufficient for our concerns here. (Any of the wide family of precise pooling and comparison rules in Fig. 4 leads to this assumption.) As indicated in Fig. 15, these other complex channels may substantially attenuate the dip in the predictions for the observer by "filling in" the dip in the tuned channel's curve.

Consider, in particular, the complex channels that are sensitive to such high spatial frequencies that they respond only to the square-elements' edges. Their responses will be approximately proportional to the total amount of edge and hence to the elements' widths rather than to their areas. Hence their response-vs-contrast curves will dip to a minimum when the contrast ratio compensates for the width ratio (rather than for the area ratio as predicted by the tuned simple channels). Notice that the difference between the width ratio and the area ratio is greater at greater area ratios. Hence, the "filling in" by these complex channels would be expected to be greater for greater area ratios. And indeed the measured dip magnitudes for square-element patterns (Figs 12 and 14, top panels) depart more dramatically from the predicted value of 1.0 as area ratio increases. A prediction can also be made about the position of the dip in the observer's curves. Since the ratio of large-to-small widths is always less than the ratio of large-to-small areas, the observers' dip's position will tend to be systematically displaced away from the area ratio toward the width ratio. This direction of bias could explain why the measured dip positions for the square elements in the observer's results (Figs 11 and 13, top panels) are not exactly on the $k = 1$ line but pulled toward the $k = 2$ line.

Varying the high spatial-frequency content of the stimuli. If intrusions of complex channels sensitive to the high-spatial frequency content of the square elements are the explanation of the dip magnitudes being less than 1.0, then diminishing (or enhancing) the effective high-spatial-frequency content of the stimuli might increase (or decrease) the magnitudes of the dip. How effective the high spatial-frequency content of a pattern is (relative to its low spatial-frequency content) can be changed by varying the scale of the pattern (e.g. by varying viewing distance) which moves the whole spatial-frequency content of the pattern to different parts of the spatial-frequency dimension. In our earlier study (Sutter *et al.*, 1989), we reported the average of a group of observers' segregation ratings as we varied the spatial scale of square-element patterns over a factor of 8 (Sutter *et al.*, 1989, Fig. 15). Here we ran one individual subject (ws) at different viewing distances (halving and doubling the distance used in the main study). We obtained the expected result both times: namely, at smaller scales (larger viewing distances, less effective high spatial-frequency content), the dip magnitudes are larger, indicating less intrusion from other channels.

We also varied high spatial-frequency content by comparing square elements (with abrupt spatial edges) to "blob" elements (of the same effective area but with gradual edges and, therefore, with less high spatial-frequency content). We did so with two observers (ws, and a new observer, so) and again obtained the expected result: namely, larger dip magnitudes (less intrusion from other channels) with blob elements than with square elements.

Untuned (off-frequency) complex channels can explain the discrepancies for grating-element patterns

The tuned channels for grating-element patterns are complex ones (like that in Fig. 6) where the first filter is sensitive to the dominant spatial frequency and orientation of the little grating patches. However, grating patches contain a range of spatial frequencies and orientations in addition to the dominant one and, further, smaller grating patches contain a broader range of these other spatial frequencies and orientations than do larger patches. These extra spatial frequencies are both higher and lower than the dominant spatial frequency of the grating patches.

Hence, as the contrast ratio of the smaller to larger grating-patch elements in our patterns varies, the relative responses of a number of complex channels (in addition to the tuned channel) will vary. The minimum in the curve for each of these other channels will occur at a different contrast ratio. (The sketches in Fig. 15 are again appropriate here.) Thus, these channels can attenuate the dip predicted by the tuned channel for the observer's response. This attenuation would be expected to be greater for greater area ratios (since the range of frequencies/orientations in the small patch gets broader as the patch gets smaller) which is in agreement with the experimental results showing smaller dip magnitudes for larger area ratios (Figs 12 and 14, bottom panels).

Note one further aspect of these results. There is more attenuation of dip magnitude for square elements than for grating elements (Figs 12 and 14). In fact, for grating elements, the only substantial attenuation of dip magnitude is at the largest area ratio, and there are two observers (ch and cv) who do not even show this effect. Why should there be this difference between square and grating elements? It is again what is expected on the notion of other-channel intrusion. For the square elements, as mentioned above, the predicted dip position from the other channels is at the width ratio, which becomes quite different from the area ratio (the dip position of the tuned channels). For the grating elements, however, the predicted dip positions are those of the complex channels responsive to the "other frequencies and orientations" in the small grating patch. These other frequencies are not very different from the dominant frequency affecting the tuned channel; hence the curves of these other channels should not dip at very different positions from that of the tuned channel and could not, therefore, attenuate the dip magnitude in the observer's curve very much.

The question arises as to whether these other untuned complex channels may affect the observer's dip position as well as dip magnitude for grating-element patterns (as they do for square-element patterns). As mentioned

above, the dip positions of these other channels are not, on the whole, very different from those of the tuned channel in the case of grating elements. Thus, any bias will be modest. Some calculations do suggest, however, that when the dip magnitude is affected (at the largest area ratios), there might be a small effect on dip position (although smaller than the effect for square elements). This possible small bias would (as in the case of square elements) be toward lower values of C_{small}/C_{large} and hence would tend to make k seem larger than the actual value from the intermediate nonlinearity of the tuned channels.

In light of this possible bias in dip position caused by the intrusion of other complex channels, it seems conservative to conclude that the power of the intermediate nonlinearity in the tuned complex channels is "somewhat less than 4" rather than "near 4" as suggested by initial inspection of the results. On the other hand, note that for two observers (ch and cv) there is hardly any attenuation of dip magnitude at all, and for no observer is the attenuation large at the second-highest area ratio (bottom panel, Figs 12 and 14); yet the dip positions for those points agree with all the other points (bottom panel, Figs 11 and 13) and suggest a value certainly greater than 2 and approximately equal to 4. Thus, we will tentatively conclude from the tradeoff condition results that the power of the intermediate nonlinearity is probably near 3 or 4.

RESULTS FROM THE ONE-ELEMENT CONDITIONS AND INDIVIDUAL DIFFERENCES

If intrusion from channels other than the tuned ones explains why the dip magnitudes (d/r) for the average observer are less than 1.0, then differences among observers (which are primarily in dip magnitudes) might well result from individual variation in other-channel intrusion: those observers with *less* other-channel intrusion would show *greater* dip magnitudes.

There is a logically independent subset of data from this study that can be used to test this supposition. This subset of data is that from the one-element conditions, which we have not yet looked at. (These are the conditions represented by open symbols in the matrix of stimulus conditions of Fig. 8.) If the intrusion of other channels is minimal in the tradeoff conditions (as indicated by large dip magnitudes) for a particular observer, then other-channel intrusion might also be expected to be minimal in the one-element conditions for that observer. How does other-channel intrusion show up in the one-element conditions?

If other-channel intrusion is completely absent in the one-element conditions (if the tuned channels by themselves mediate segregation in these conditions), then a clean prediction about spatial summation in that condition can be made in terms of the segregation threshold (the contrast necessary to produce a criterion segregation amount for the one-element patterns—see Methods for further definition). In words, the small-element threshold is predicted to be greater than the large-element threshold by the amount that compensates for their difference in area. In symbols:

$$\frac{C_S^*}{C_L^*} = \left|\frac{A_L^*}{A_S^*}\right|^{(1/k)} \tag{5}$$

where the subscripts S* and L* stand for the one-element stimuli containing small or large elements respectively; the quantities A_S^* and A_L^* are the areas of those stimuli; C_S^* and C_L^* are the segregation thresholds; and the parameter k equals 1 for square elements and equals the power of the intermediate nonlinearity for grating-patch elements.

To gain insight into the prediction of equation (5), the reader might go through this exercise; imagine varying the size of the elements in the left column of Fig. 3 or Fig. 6 (the column representing one-element stimuli) and note the resulting behavior of the tuned channel.

This equation (5) is identical to equation (3) with the

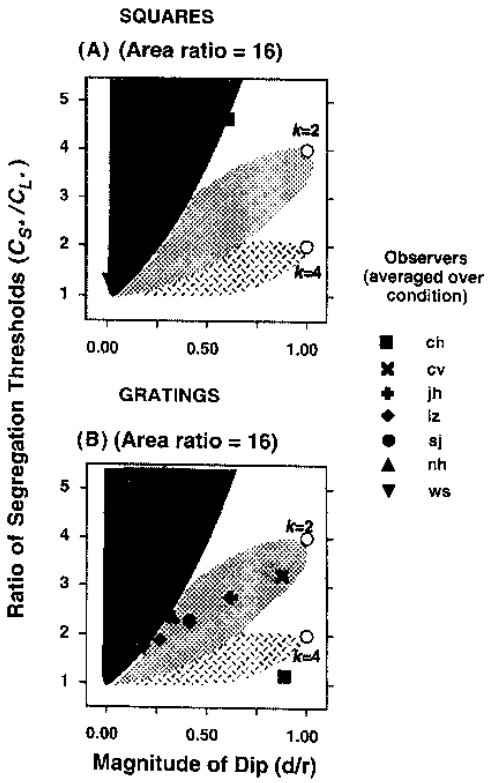

FIGURE 16. Individual differences in the tradeoff conditions and the one-element conditions. The horizontal axis shows dip magnitude in the tradeoff conditions for an area ratio of 16, and the vertical axis shows the the ratio of segregation thresholds in the one-element conditions for elements having an area ratio of 16. The results for square-element and grating-element patterns are in (A) and (B), respectively. The open circles show the predictions, for various values of k, if the tuned channels alone determined texture segregation for all stimuli. The point for $k = 1$ is at (1,16) and outside the visible part of the graph although its direction is indicated by an arrow. As channels other than the tuned ones intrude, both the dip magnitude in the tradeoff conditions and the ratio of segregation thresholds in the one-element conditions diminish, although, as suggested by the shaded areas, the relationship need not be linear.

subscripts 2 and 1 in equation (3), representing the two element types in the same pattern, replaced by S* and L* in equation (5), representing different types of one-element patterns.

Now consider what happens to the prediction for spatial summation in equation (5) if other channels intrude. In general, these other intruding channels are expected to be more sensitive to the small element relative to the large element than are the tuned channels, because these other channels are responsive to the edges (of square elements) or to the extra frequencies (in the smaller grating patches). This extra sensitivity to the smaller elements should *reduce* the amount of spatial summation seen as you go from the smaller to the larger elements.

Figure 16 shows the behavior of individual observers in both the one-element condition (spatial summation as measured by the ratio of segregation thresholds C_S*/C_L*, on the vertical axis) and the tradeoff conditions (dip magnitudes, on horizontal axis). Results are shown for square-element patterns (top panel) and grating-element patterns (bottom panel), where the ratio of the areas of the two elements was always 16. (Other area ratios lead to less-diagnostic results so are not shown here.) Each solid symbol is for an individual observer. The open circles are predictions, when only the tuned channels are active, for several values of k. The point for $k = 1$ is outside the visible area of the graph; it has coordinates (1,16) indicated by the arrow.

As other channels intrude, an observer's performance is predicted to move both down (toward less spatial summation in the one-element conditions) and to the left (toward smaller dip magnitudes in the tradeoff conditions) from the tuned-channel-only predictions in Fig. 16. The shaded areas between the origin and the open circles are drawn to indicate this qualitative expectation since, in the case of these predictions, there is no reason to expect exact proportionality between the effects in the one-element and tradeoff conditions.

As can be seen in Fig. 16, both for grating and square elements, observers having smaller dip magnitudes in the tradeoff conditions also tend to have less spatial summation in the one-element conditions. There is one outlying point—ch and grating elements (the solid square in the bottom panel). This outlying point cannot be entirely the result of intrinsic variability in the rating responses (although some of it probably is), because individual sessions for this observer show the same pattern of results. Some of the discrepancy might be attributable to ch's complex channels being characterized by a higher value of k than the other observers (although, if this were the complete explanation, one might expect that higher power to also have shown up in the position of the dips in Fig. 13). The explanation of this anomalous point is unclear.

Overall, however, the results of Fig. 16 provide support to two conclusions: (i) other-channel intrusion is the source of the earlier discrepancy between the tuned-channel predictions and the dip magnitude data in the tradeoff conditions; (ii) the variation among individual observers in this study—which occurs both for dip magnitudes (although not for dip positions) and for spatial summation in segregation thresholds—seems explainable on the basis of other-channel intrusion.

There is another consistency in these results that can be seen by noting the order of the symbols from lower left to upper right in each panel of Fig. 16; namely, observers showing the greatest other-channel intrusion (symbols closest to the lower left) with square elements tend to do so also with grating elements. This consistency may be due to the fact that, for both square-element and grating-element patterns, the intruding channels are complex channels sensitive to relatively high spatial frequencies. Thus, although more subtle difference among observers' complex channels or higher processes cannot be ruled out (certainly not for the one discrepant observer), the individual differences here may largely occur because some observers have more (or more sensitive) complex channels of the type that have first filters sensitive to relatively high spatial frequencies.

Implications for value of k

This pattern of variation among individual observers can also say something about the actual values of k for square and grating elements. Any conclusion should be viewed with caution since we have not been able to make a more definite prediction than the qualitative prediction crudely indicated by shaded areas in Fig. 16. How big those shaded areas should be, for example, given reasonable parameters in a fully elaborated multiple-channel model, is not at all clear to us and the positions of the points in Fig. 16 are quite scattered. (Note that they are a great deal more scattered than those in Figs 11 and 13 for dip positions.) With this caution in mind, the individual observer results in Fig. 16 suggest a value of k near 1 (or perhaps a little higher) for the square elements and a value of k somewhere between 2 and 3 for the grating elements (for six of seven observers—the last observer may have a higher k). Given the sources of uncertainty in interpreting Fig. 16, these values seem satisfactorily similar to those suggested by the earlier analysis based on dip position in the trade-off conditions (Figs 11 and 13). In fact, the suggested value of k for square elements is essentially identical ($k = 1$ or perhaps a little higher). The suggested value for grating elements is a little lower here than in the earlier analysis (which suggested k near 4 but perhaps 3 or 3.5). Putting the two estimates together (considering the fact that the estimate from Fig. 16 is less stable) suggests a conservative conclusion that the intermediate nonlinearity in complex channels is characterized by a value of k somewhere in the range 2.5–4 but probably toward the higher end of it. Or, for simplicity, we will say that k for complex channels equals 3 or 4.

Some caveats about value of k

We have been speaking as if all complex channels had the same power of intermediate nonlinearity and as if the

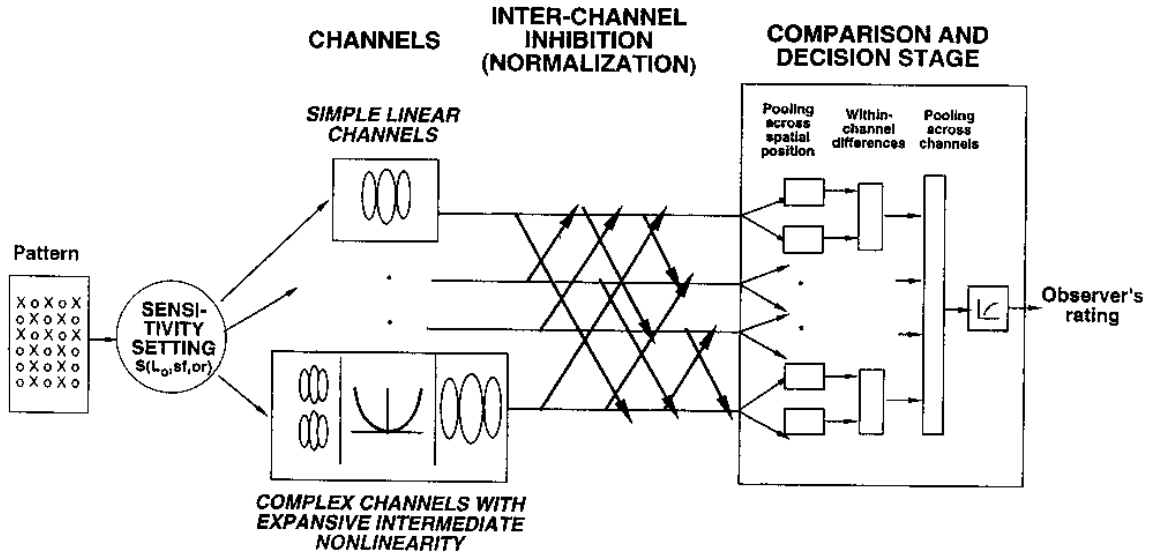

FIGURE 17. Model of texture segregation consistent with the results of this study. The nonlinearity in the complex channels is expansive and can be described by a power function with an exponent of about 3 (more conservatively, in the range from 2 to 4). The highly compressive intensive nonlinearity we identified in previous work (Graham *et al.*, 1992; Graham & Sutter, 1996) may result from inhibition among channels (shown on the diagram as a normalization network), but the study reported here suggests that it cannot result from a local nonlinearity occurring before the channels. Thus, the early sensitivity-setting stage shown on the diagram does not introduce any compression for contrasts less than 100%. It does set a sensitivity factor that depends on mean luminance, spatial frequency, and orientation.

same power of nonlinearity held over the full contrast range. This is the simplest model consistent with our results perhaps, but not the only one.

For one thing, we have only studied a limited range of spatial and temporal frequency, and different intermediate nonlinearities might well hold in different ranges. (For that matter, simple channels might act differently within different ranges as well.)

But even within one range, our results here could not distinguish between the existence of a number of different kinds of complex channels having different exponents (with the conglomerate effect of a power of 3 or 4) and the existence of only one kind (with a power of 3 or 4).

Also, our conclusions about the values of k are only valid at the contrasts of the dips (since it is on the data at the dips that these conclusions are based). It is conceivable that for higher or lower contrasts different values would hold. However, we used a fairly broad range of contrasts (Table 1) and so, for simplicity in this paper, will continue to talk as if the same value held over the full contrast range (about 1 for the simple channels and 3 or 4 for the complex channels).

DISCUSSION

About the intensive nonlinearity

To explain aspects of texture perception one needs, in addition to the "spatial nonlinearity" embodied in the complex channels, an "intensive nonlinearity" that is compressive at quite modest levels of contrast. As mentioned in the Introduction, two quite different relatively early visual processes have been suggested as candidates for the intensive nonlinearity. In more detail, as applied to texture segregation, these two candidates are:

(i) A relatively early, relatively local compressive nonlinearity that comes before the simple and the complex channels but after a stage at which sensitivity to different background luminances, spatial frequencies, and orientations is set (see Fig. 10 of Graham & Sutter, 1996). It is approximately equally compressive for luminances below and above the background. Its physiological substrate might lie in light adaptation processes at the retinal level that readjust the operating range of the system to be around the background luminance.

(ii) Mutual inhibition among the channels (e.g. intracortical inhibition), which may be modeled as a contrast-gain control or normalization network (see Fig. 2 of Graham & Sutter, 1996 and Fig. 17 here). Note that this also comes after the stage at which sensitivity to different background luminances and spatial frequencies and orientations is set (since it comes after the channels themselves) and is approximately equally compressive for luminances below and above the background. Its physiological substrate may be the inhibition among cells in cortical area V1 or V2 or even higher.

Up to this point, we have ignored the intensive non-

linearity in interpreting the results of this study. Let us now consider what effect, if any, these two candidates for the intensive nonlinearity should have on the conclusions here and, vice versa, what effect the results here may have on our understanding of the intensive nonlinearity.

Relatively early, local nonlinearity. With element-arrangement textures, the compressive effects of the intensive nonlinearity are particularly obvious in studies of constant-difference series of patterns (Graham, 1991; Graham, Beck, & Sutter, 1992; Graham & Sutter, 1996). The difference between the two element types' contrast is constant for all patterns in such a series. As will be explained more fully elsewhere (in preparation), if a relatively early (before the channels), local (pointwise) nonlinearity existed, its effect in area-contrast tradeoff experiments like those here and in constant-difference-series experiments would be the same. It would appear in both as compression, or else it would appear in both types of experiments as expansion. However, the area-contrast tradeoff experiments reported here showed linearity or even slight expansiveness for squares and expansiveness for gratings, while the constant-different series experiments show compression for both squares (studies referenced above) and gratings (Graham and Sutter, in preparation). Thus a relatively early and local nonlinearity by itself is probably not the underlying cause of both phenomena. To be certain we still need to investigate the same contrast ranges for the same observers. If the results hold up in the same conditions, as seems likely, then any attempt to include an early, local nonlinearity in the model leads to a contradiction with one set of experiments or the other unless very elaborate further modifications are also made to the model. In short, the results of the area-contrast tradeoff experiments and constant-difference series experiments strongly suggest that the compressive intensive nonlinearity is NOT a relatively early and local process.

Inhibition among channels (in a normalization network). The case of inter-channel inhibition (the other candidate for the intensive nonlinearity—see Fig. 17) is very different. It seems likely that inhibition among channels modeled as a normalization network will not substantially affect the predicted results of area-contrast tradeoff experiments like those here. This seems likely because the tradeoff between the two elements' areas and contrasts occurs in the complex channels *before* the inter-channel inhibition (the normalization) is applied. Thus, the existence of inter-channel inhibition would not change our conclusion about the intermediate nonlinearity in the complex channels nor about the linearity of the simple channels. Consequently, we think it likely that inhibition among channels both explains the results of constant-difference-series experiments and is consistent with the results here.

About a model of both the spatial and intensive nonlinearities

To summarize our conclusions about models: the following model is consistent with all our results from area-contrast-tradeoff and constant-difference series experiments and is shown in the diagram in Fig. 17:

(Property 1) There are both simple and complex channels.

(Property 2) The intermediate nonlinearity in the complex channel is expansive with a power of 3 or 4.

(Property 3) There is inhibition among all the channels (e.g. a normalization network).

Shown also in Fig. 17 are two boxes explicitly representing the simplifications of all other stages of visual processing used in making predictions of the observer's behavior from the three properties of major interest. The sensitivity-setting box represents all that comes between the stimulus and the channels (the retina and the lateral geniculate nucleus at least). It does NOT introduce any compression for contrasts less than 100%. It does set a sensitivity factor that depends on mean luminance and that may also depend on spatial frequency and orientation. The comparison-and-decision box was discussed earlier (Fig. 4) and represents all that comes between the channels and the response of the observer (at least all higher cortical levels of processing).

We have noted some relatively minor variations of this model that would also work, e.g. different complex channels might have somewhat different power functions at their intermediate stage rather than all having the same power. Also, several alternative forms of nonlinear-pooling inside complex channels could substitute for the intermediate nonlinearity (see Part III of Appendix I).

To say that the model of Fig. 17 is consistent with our results does not state the implications of our results as strongly as possible. We have also argued that several modifications of the diagram in Fig. 17 produce models which are *not* consistent with our results here. To summarize these points, here is a list of potential modifications in Fig. 17 that would NOT be consistent with our results (when everything else remains as shown) and thus can be rejected:

(Point 1) Neither the simple nor the complex channels can be omitted.

(Point 2) The intermediate pointwise function in the complex channel cannot be compressive or piecewise-linear.

(Point 3) There cannot be a relatively early and local compressive nonlinearity (occurring before the simple and complex channels) added to this diagram (unless it only becomes compressive for contrasts outside the range used here).

(Point 4) There must be inhibition among channels (e.g. a normalization network) or some other form of compressive intensive nonlinearity acting after the channels, and it must be active at low contrasts.

A final comment: As is always true, however, to have decided among several possible explanations of phenom-

ena is not to have decided among all possible explanations. There are always alternative explanations (or unconsidered factors or confounding factors, depending on how you look at them). Here we mention two of some interest.

(Alternative 1) Sparse first layer rather than a traditional linear filter. We have been assuming that there is such a dense set of neurons as part of the first filter that its output is effectively continuous (as in the second row of Fig. 6). Perhaps, however, the second filter of a particular complex channel only collects outputs from a small number of first-stage receptive fields, perhaps only a few in a line parallel to the main orientation of the first-stage receptive fields. Thus, when increasing the diameter of circularly symmetric Gabor patches (as in our experiments), the stimulation to the second stage would not be increasing in proportion to the patches' area but only in proportion to its linear dimension (the length of bar). To test this particular kind of sparse sampling one could run experiments varying only one dimension of the Gabor patches.

In another variety of sparse sampling, the first-stage fields might be scattered about incoherently (out of phase with one another).

(Alternative 2) Dynamics interactions. Complex channels (second-order mechanisms, collector units, etc.) were presented here as rather hard-wired or, at least, nothing was said about their not being hard-wired. In the current literature, the main alternative to complex channels and their relatives is an appeal to in-place dynamic interactions occurring between spatially separated neurons (e.g. the "impletion" process of Caelli, 1985; the "association field" of Field, Hayes, & Hess, 1993; the active reentrant connections of Sporns, Tonini, & Edelman, 1991; the in-place spatial interactions of Polat & Sagi, 1993, 1994). Such dynamic interactions have been studied for visual neurons (e.g. Gilbert, 1994), but their status as an explanation for experiments like these is unclear. In many cases, at least, they will act just like hard-wired entities and so cannot be distinguished for the purposes of explaining these experiments. However, they certainly represent an interesting possibility.

Relation to previous studies of intermediate nonlinearity in complex channels

Previous to this current study we (and others) had often assumed that the intermediate nonlinearity in complex (non-Fourier, second-order) texture channels was piecewise linear (sometimes only by casually drawing piecewise-linear functions on the diagram) although this assumption has usually had little effect on the predictions of interest (Graham, 1991; Graham *et al.*, 1992, 1993, 1996; Sperling *et al.*, 1994; Wilson, 1993).

On a number of other occasions, expansive functions had been assumed instead. Lin & Wilson (1996) used an expansive function (a power of 2) in their model of non-Fourier channels for pattern discrimination, although again this assumption apparently had little effect. Solomon & Sperling (1994) also used an expansive function (a power of 2), which was necessary in their study to ensure that certain stimuli were detected only by half-wave mechanisms of the appropriate polarity (see also Solomon, Sperling, & Chubb, 1993). An expansive function (either a fourth power or a linear function with a threshold) was also used by Victor & Conte (1989, 1991, 1996) in their texture model, and they do provide some evidence for the necessity of the expansiveness (e.g. 1991, p. 1484). Their studies are primarily of visual evoked potential in response to alternations among certain classes of texture (occasionally compared to formal psychophysics), so its generalizability to perceived texture segregation is not indubitable but suggestive. Another recent report is at least consistent with the suggestion of an expansive intermediate nonlinearity: Landy (1996) reported that discrimination between textures differing in second-order contrast shows a dipper-shaped contrast discrimination function like that found in ordinary (first-order) contrast discrimination with sinusoidal gratings, and he suggested that the same explanations (subthreshold summation, an expansive nonlinearity, or reduction of uncertainty) might apply to second-order as are often applied to first-order dippers. Given the nature of the textures used here, the expansive nonlinearity at the intermediate stage in complex channels would predict the dipper.

In short, the results reported here, which demonstrate that the intermediate nonlinearity in the complex channels is expansive and well described by a power function with an exponent of 3 or 4 (within the contrast range studied, of course), are consistent with the little information that has been available before about the intermediate nonlinearity in complex channels like those of Fig. 1 (i.e., those in which the second stage collects from receptive fields of like shape but different position). Note that we are explicitly NOT discussing what were called "higher-order" mechanisms in the Introduction. The intermediate nonlinearity there may be quite different.

Half- versus full-wave intermediate nonlinearities. A related question is whether the intermediate pointwise nonlinearity is of the half-wave or full-wave type. The power function in equation (1) and those sketched in the intermediate nonlinearity box of Fig. 1 are of the full-wave type. A power function of the positive half-wave type is zero for all inputs below zero:

$$\begin{aligned} g(x,y) &= a \cdot |f(x,y)|^k \text{ for } f(x,y) \geq 0 \\ &= 0 \text{ otherwise} \end{aligned} \tag{6}$$

In a power function of the negative half-wave type the non-zero outputs g occur for negative values of input f. Our results here are consistent with either full- or half-wave functions (Part III of Appendix I).

Previous evidence suggests that most observers have both half-wave and full-wave mechanisms active in some kinds of texture perception (Malik & Perona, 1990; Sperling, Chubb, Solomon, & Lu, 1994; Solomon & Sperling, 1994) but that the half-wave ones do not support as many perceptual effects—e.g. second-order

Mach bands or lateral contrast induction—as do the full-wave (Sperling *et al.*, 1994; Lu & Sperling, 1996).

Interestingly, the results for non-Fourier channels in motion are somewhat different: 2/3 of the observers may not have any half-wave mechanisms at all but rely entirely on full-wave mechanisms (Sperling, Chubb, Solomon, & Lu, 1994; Solomon & Sperling, 1994). It would be interesting to know what the expansiveness or compressiveness of the intermediate nonlinearity in motion is.

SUMMARY AND CONCLUSIONS

The tradeoff between contrast and area for square elements and for Gabor-patch elements was studied in element arrangement patterns like those of Fig. 2. The tradeoff was approximately linear for square elements—in other words, the minimal segregation occurred when the product of area and contrast was equal for the two element types. This tradeoff was highly nonlinear for grating elements, however; the minimal segregation occurred when the product of area and the fourth-power of the contrast was equal for the two element types.

Overall, the results are consistent with the model shown in Fig. 17. For square elements, perceived segregation is primarily the result of a simple linear (Fourier, first-order) channel with peak sensitivity at the fundamental frequency/orientation of this pattern. For grating-patch elements, perceived segregation is primarily the result of the complex channel tuned to this pattern (in which the first filter is tuned to the grating spatial frequency, and the second filter is tuned to the fundamental frequency determined by the arrangement of elements).

The newest conclusion about the complex channels is this: The nonlinearity associated with these channels (shown in Fig. 1 as an intermediate pointwise nonlinearity between the two linear filter stages) is expansive. It can be described as a power function with an exponent substantially greater than 1.0, probably between 3 and 4.

Individual observers were studied in detail. One important aspect of the results—the ratio of contrasts that causes maximal interference (the position of the dip in the curves in Figs 11 and 13)—is quite stable across individual subjects. It is this ratio which reveals the compressiveness or expansiveness of the spatial summation. However, two other aspects vary substantially among observers. One is the magnitude of the maximal interference (the depth of the dip). The other is in the segregation of one-element patterns (patterns in which one of the two element types has contrast zero)—in particular in the degree to which the perceived segregation in one-element patterns increases with element area. These differences are correlated across observers and across element type (squares/gratings). Both these individual differences may be understood as the result of differences in the extent to which channels other than the tuned ones contribute to the perceived segregation of these patterns. These other channels would be primarily complex channels sensitive to relatively high spatial frequencies (high relative to the fundamental frequency of the textures). The model in Fig. 17 would apply to each observer, therefore, but the population of complex channels would be somewhat different from observer to observer.

Another implication of the results in this paper (when considered with those of other studies) is for the previously identified intensive nonlinearity active in texture segregation, a nonlinearity which is compressive at quite low contrasts (e.g., Wilson, 1993; Graham *et al.*, 1992; Graham & Sutter, 1996). Two candidates to explain this intensive nonlinearity have been considered in the past: a relatively early local nonlinearity (although it must occur after sensitivity to background luminance, spatial frequency and orientation have been set—Graham & Sutter, 1996), and inhibition among the channels (modeled as a normalization or contrast-gain network). The results of the study here, while consistent with inhibition among the channels (as in Fig. 17), are quite difficult and perhaps impossible to reconcile with a relatively early local nonlinearity if the results of appropriate constant-difference series experiments are considered as well.

REFERENCES

Badcock, D. R. & Derrington, A. M. (1989). Detecting the displacement of spatial beats: no role for distortion products. *Vision Research, 29*, 793–797.

Beck, J. (1973). Similarity grouping of curves. *Perceptual and Motor Skills, 36*, 1331–1341.

Beck, J., Rosenfeld, A. & Ivry, R. (1989). Line segregation. *Spatial Vision, 4*, 75–101.

Bergen, J. R., & Landy, M. S. (1991) Computational modeling of visual texture segregation. In M. S. Landy and J. A. Movshon (Eds.), *Computational models of visual processing* (pp. 253–271). Cambridge, MA: MIT Press.

Boulton, J. C. & Baker, C. L. (1993). Different parameters control motion perception above and below a critical density. *Vision Research, 33*, 1803–1811.

Bowen, R. W. & Wilson, H. R. (1994). A two-process analysis of pattern masking. *Vision Research, 34*, 645–657.

Burr, D. C., & Morrone, M. C. (1994) Features in visual images. In *Higher-order processing in the visual system* (pp. 129–146). Chichester (Ciba Foundation Symposium 184): Wiley.

Caelli, T. (1985). Three processing characteristics of visual texture segmentation. *Spatial Vision, 1*, 19–20.

Cannon, M. W. & Fullenkamp, S. C. (1993). Spatial interactions in apparent contrast: individual differences in enhancement and suppression effects. *Vision Research, 33*, 1685–1695.

Chubb, C., McGowan, J., Sperling, G., & Werkhoven, P. (1994) Non-Fourier motion analysis. In *Higher-order processing in the visual system (Ciba Foundation Symposium No. 184)* (pp. 193–210). Chichester: Wiley.

Chubb, C. & Sperling, G. (1988). Drift-balanced random stimuli: a general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America A, 5*, 1986–2007.

Chubb, C. & Sperling, G. (1989). Two motion perception mechanisms revealed by distance driven reversal of apparent motion. *Proceedings of the National Academy of Sciences USA, 86*, 2985–2989.

Clark, M., Bovik, A. C., & Geisler, W. S. (1987). Texture segmentation using a class of narrowband filters. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*.

Derrington, A. M. & Badcock, D. R. (1992). Two-stage analysis of the motion of 2-dimensional patterns. What is the first stage? *Vision Research, 22*, 691–698.

Derrington, A. M., & Henning, G. B. (1994). Implications of motion detection for early non-linearities. In *Higher-order processing in the visual system* (pp. 309–329). Wiley: Chichester (Ciba Foundation Symposium 184).

Field, D. J., Hayes, A. & Hess, R. F. (1993). Contour integration by the human visual system: evidence for a local "association field". *Vision Research, 33*, 173–193.

Fleet, D. J. & Langley, K. (1994). Computational analysis of non-Fourier motion. *Vision Research, 34*, 3057–3079.

Fleet, D. J., Wagner, H. & Heeger, D. J. (1996). Neural encoding of binocular disparity: Energy models, position shifts, and phase shifts. *Vision Research, 3612*, 1839–1857.

Fogel, I. & Sagi, D. (1989). Gabor filters as texture discriminators. *Biological Cybernetics, 61*, 103–113.

Georgeson, M. A. (1992). Human vision combines oriented filters to compute edges. *Proceedings of the Royal Society of London Series B Biological Sciences, 249*, 235–245.

Georgeson, M. A. (1994). From filters to features: location, orientation, contrast and blur. In *Higher-order processing in the visual system* (pp. 147–169). Chichester (Ciba Foundation Symposium 184): Wiley.

Gilbert, C. D. (1994). Circuitry, architecture and functional dynamics of visual cortex. In *Higher-order processing in the visual system* (pp. 35–62). Chichester (Ciba Foundation Symposium 184): Wiley.

Graham, N. (1991). Complex channels, early local nonlinearities, and normalization in perceived texture segregation. In M. S. Landy & J. A. Movshon (Eds), *Computational models of visual processing* (pp. 273–290). Cambridge, MA: MIT Press.

Graham, N. (1994). Nonlinearities in texture segregation. In *Higher-order processing in the visual system (Ciba Foundation Symposium No. 184)* (pp. 309–329). Chichester: Wiley.

Graham, N., Beck, J. & Sutter, A. (1992). Nonlinear processes in spatial-frequency channel models of perceived texture segregation. *Vision Research, 32*, 719–743.

Graham, N. & Sutter, A. (1996). Effect of spatial scale and background luminance on the spatial and intensive nonlinearities in texture segregation. *Vision Research, 36*, 1371–1390.

Graham, N., Sutter, A. & Venkatesan, C. (1993). Spatial-frequency- and orientation-selectivity of simple and complex channels in region segregation. *Vision Research, 33*, 1893–1911.

Grossberg, S. & Mingolla, E. (1985). Neural dynamics of perceptual grouping: textures, boundaries, and emergent features. *Perception and Psychophysics, 38*, 141–171.

Gurnsey, R. & Browse, R. A. (1987). Micropattern properties and presentation conditions influencing visual texture discrimination. *Perception and Psychophysics, 41*, 239–252.

Hammett, S. T. & Smith, A. T. (1994). Temporal beats in the human visual system. *Vision Research, 34*, 2833–2840.

Heeger, D. J. (1991). Computation model of cat striate physiology. In M. S. Landy & J. A. Movshon (Eds), *Computational models of visual processing*. Cambridge, MA: MIT Press.

Henning, G. B., Hertz, B. G. & Broadbent, D. E. (1975). Some experiments bearing on the hypothesis that the visual system analyzes spatial patterns in independent bands of spatial frequency. *Vision Research, 15*, 887–899.

Hess, R. F. & Badcock, D. R. (1995). Metric for separation discrimination by the human visual system. *Journal of the Optical Society of America, 12*, 3–16.

Hess, R. F. & Hayes, A. (1994). The coding of spatial position by the human visual system: effects of spatial scale and retinal eccentricity. *Vision Research, 34*, 625–643.

Hess, R. F. & Wilcox, L. M. (1994). Linear and non-linear filtering in stereopsis. *Vision Research, 34*, 2431–2438.

Julesz, B. (1975). Experiments in the visual perception of texture. *Scientific American, 232*, 34–43.

Kingdom, F. A. A. & Keeble, D. R. T. (1995). Sensitivity to orientation modulation in micropattern-based textures. *Vision Research, 35*, 79–91.

Kingdom, F. A. A. & Keeble, D. R. T. (1996). A linear systems approach to the detection of both abrupt and smooth spatial variations in orientation-defined textures. *Vision Research, 36*, 409–420.

Klein, S. A., Stromeyer, C. F. & Ganz, L. (1974). The simultaneous spatial frequency shift: a dissociation between the detection and perception of gratings. *Vision Research, 14*, 1421–1432.

Landy, M. S. (1996). Second-order contrast discrimination. *Investigative Ophthalmology and Visual Science, 37(3)*, 5243.

Landy, M. S. & Bergen, J. R. (1991). Texture segregation and orientation gradient. *Vision Research, 31*, 679–691.

Landy, M. S., & Ternes, C. M. (1995). 2nd-order spatial contrast sensitivity. Presented at the annual meetings of the Optical Society of America.

Lee, T. S. (1995). A Bayesian framework for understanding texture segmentation in the primary visual cortex. *Vision Research, 35*, 2643–2657.

Levi, D. M. & Waugh, S. J. (1996). Position acuity with opposite-contrast polarity features: evidence for a nonlinear collector mechanism for position acuity?. *Vision Research, 36*, 573–588.

Lin, L. M. & Wilson, H. R. (1996). Fourier and non-Fourier pattern discrimination compared. *Vision Research, 36*, 1907–1918.

Lu, Z. L. & Sperling, G. (1995). The functional architecture of human visual motion perception. *Vision Research, 35*, 2697–2722.

Lu, Z. L. & Sperling, G. (1996). Second-order illusions: Mach bands Chevreul, and Craik-O'Brien-Cornsweet. *Vision Research, 36*, 559–572.

Malik, J. & Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A, 7*, 923–932.

McOwan, P. W. & Johnston, A. (1996). A second-order pattern reveals separate strategies for encoding orientation in two-dimensional space and space–time. *Vision Research, 36*, 425–430.

Meese, T. S. (1995). Phase-reversal discrimination in one and two dimensions: performance is limited by spatial repetition, not by spatial frequency content. *Vision Research, 35*, 2157–2167.

Meese, T. S. & Georgeson, M. A. (1996). The tilt aftereffect in plaids and gratings: channel codes, local signs and "patchwise transforms". *Vision Research, 36*, 1421–1437.

Morgan, M. J. & Glennerster (1991). Efficiency of locating centers of dot-clusters by human observers. *Vision Research, 31*, 2075–2083.

Morgan, M. J., Hole, G. J. & Glennerster (1990). Biases and sensitivities in geometrical illusions. *Vision Research, 30*, 1793–1810.

Morgan, M. J. & Hotopf, W. H. N. (1989). Perceived diagonals in grids and lattices. *Vision Research, 29*, 1005–1015.

Morrone, M. C. & Burr, D. C. (1988). Feature detection in human vision: a phase-dependent energy model. *Proceedings of the Royal Society of London Series B Biological Sciences, 235*, 145–221.

Moulden, B. M. (1994). Collator units: second-stage orientational filters. In *Higher-order processing in the visual system* (pp. 170–192). Wiley: Chichester (Ciba Foundation Symposium 184).

Pantle, A. (1992). Immobility of some second-order stimuli in human peripheral vision. *Journal of the Optical Society of America A, 9*, 863–867.

Polat, U. & Sagi, D. (1993). Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments. *Vision Research, 34*, 73–78.

Polat, U. & Sagi, D. (1994). The architecture of perceptual spatial interactions. *Vision Research, 34*, 73–78.

Robson, J. G. (1980). Neural images: the physiological basis of spatial vision. In C. S. Harris (Ed.), *Visual coding and adaptability* (pp. 177–214). Hillsdale, NJ: Erlbaum.

Rubenstein, B. S. & Sagi, D. (1993). Effects of foreground scale in texture discrimination task: Performance is size, shape, and content specific. *Spatial Vision, 7*, 293–310.

Rubenstein, B. S. & Sagi, D. (1996). Preattentive texture segregation: the role of line terminations, size, and filter wavelength. *Perception and Psychophysics, 58*, 489–509.

Sagi, D. (1990). Detection of an orientation singularity in Gabor textures: effect of signal density and spatial frequency. *Vision Research, 30*, 1288–1377.

Sagi, D. & Hochstein, S. (1985). Lateral inhibition between spatially adjacent spatial frequency channels? *Perception and Psychophysics, 37*, 315–322.

Sakai, K. & Finkel, L. H. (1995). Characterization of the spatial-frequency spectrum in the perception of shape from texture. *Journal of the Optical Society of America A, 12*, 1208–1224.

Shapley, R. M. & Gordon, J. G. (1985). Nonlinearity in the perception of form. *Perception and Psychophysics, 37*, 84–88.

Solomon, J. A. & Sperling, G. (1994). Full-wave and half-wave rectification in second-order motion perception. *Vision Research, 34*, 2239–2257.

Solomon, J. A. & Sperling, G. (1995). 1st- and 2nd-order motion and texture resolution in central and peripheral vision. *Vision Research, 15*, 59–64.

Solomon, J. A., Sperling, G. & Chubb, C. (1993). The lateral inhibition of perceived contrast is indifferent to on-center/off-center segregation, but specific to orientation. *Vision Research, 33*, 2671–2683.

Sperling, G. (1989). Three stages and two systems of visual processing. *Spatial Vision, 4*, 183–207.

Sperling, G., Chubb, C., Solomon, J. A., and Lu, Z.L. (1994) Full-wave and half-wave rectification in motion and texture. In *Higher-order processing in the visual system* (pp. 287–308). Chichester (Ciba Foundation Symposium 184): Wiley.

Sporns, O., Tonini, G. & Edelman, G. M. (1991). Modeling perceptual grouping and figure–ground segregation by means of active reentrant connections. *Proceedings of National Academy of Sciences USA, 88*, 129–133.

Sutter, A., Beck, J. & Graham, N. (1989). Contrast and spatial variables in texture segregation: Testing a simple spatial-frequency channels model. *Perception and Psychophysics, 46*, 312–332.

Sutter, A. & Graham, N. (1995). Investigating simple and complex mechanisms in texture segregation using the speed-accuracy trade-off method. *Vision Research, 35*, 2825–2843.

Sutter, A., Sperling, G. & Chubb, C. (1995). Measuring the spatial-frequency selectivity of second-order texture mechanisms. *Vision Research, 35*, 915–924.

Thomas, J. P., Olzak, L. A. & Shimozaki, S. S. (1993). The role of Fourier components in discrimination between two types of plaid patterns. *Vision Research, 33*, 1573–1579.

Thomas, J. P. & Olzak, L. A. (1996). Uncertainty experiments support the roles of second-order mechanisms in spatial frequency and orientation discriminations. *Journal of the Optical Society of America A, 13*, 689–696.

Turano, K. & Pantle, A. (1989). On the mechanism that encodes the movement of contrast variations: velocity discrimination. *Vision Research, 29*, 207–221.

Victor, J. D. & Conte, M. (1989). Cortical interactions in texture processing: scale and dynamics. *Visual Neuroscience, 2*, 297–313.

Victor, J. D. & Conte, M. (1991). Spatial organization of non-linear interactions in form perception. *Vision Research, 31*, 1457–1488.

Victor, J. D. & Conte, M. M. (1996). The role of high-order phase correlations in texture processing. *Vision Research, 36*, 1615–1631.

Werkhoven, P., Sperling, G. & Chubb, C. (1993). The dimensionality of texture-defined motion: a single channel theory. *Vision Research, 33*, 463–485.

Wilcox, L. M. & Hess, R. F. (1996). Is the site of non-linear filtering in stereopsis before of after binocular combination? *Vision Research, 36*, 391–400.

Wilson, H. R. (1993). Nonlinear processes in visual pattern discrimination. *Proceedings of the National Academy of Sciences USA, 90*, 9785–9790.

Wilson, H. R. (1994). The role of second-order motion signals in coherence and transparency. In Higher-order processing in the visual system (pp. 227–237). Chichester *(Ciba Foundation Symposium No. 184)*: Wiley.

Wilson, H. R., Ferrara, V. P. & Yo, C. (1992). A psychophysically motivated model for two-dimensional motion perception. *Visual Neuroscience, 9*, 79–97.

Wilson, H. R. & Kim, J. (1994). Perceived motion in the vector-sum direction. *Vision Research, 34*, 1835–1842.

Wilson, H. R. & Richards, W. A. (1992). Curvature and separation discrimination at texture boundaries. *Journal of the Optical Society of America, 9*, 1653–1662.

Wolfson, S. S. & Landy, M. S. (1995). Discrimination of orientation-defined texture edges. *Vision Research, 35*, 2863–2877.

Acknowledgement—This research was supported by National Eye Institute Grant 1 RO1 EY08459. A preliminary account of some of this work was given at the 1992 annual meetings of the Optical Society of America.

APPENDIX I

Part I. Further comments on relationships among several models of texture segregation and similar perceptual tasks

As mentioned in the Introduction, some models of other types of texture discriminations (e.g. Bergen & Landy, 1991; Fogel & Sagi, 1989; Grossberg & Mingolla, 1985; Landy & Bergen, 1991; Malik & Perona, 1990; Rubenstein & Sagi, 1993; Wolfson & Landy, 1995) seem similar to complex channels invoked for non-Fourier textures. However, the second stage of these other models may be more akin to the pooling and decision stage of the complex-channels models (e.g. Figure 17) than to the second filter in the complex channels themselves. To make this point more concrete, we present a specific example in the next few paragraphs.

Suppose one tried to specify how the comparison-and-decision stage used here could be instantiated by a reasonable mechanism. The two stages of "Pooling across spatial position" and computing the "Within-channel differences" [see Fig. 4 and/or A(1) and A(2) in Part II of this Appendix I] could be computed by the following processes which others have invoked (e.g. Figure 13 in Wolfson & Landy, 1995) and have referred to as being similar to complex cells in the cortex:

(i) applying an even nonlinearity pointwise to the output of the channels, e.g. a squaring or other power as in A(1)

(ii) then filtering each channel's transformed output by an array of "edge-detecting" receptive fields [instantiating A(2)]

(iii) and then taking the maximum of these fields to both identify the correct place in the visual field (that is, to detect the edge between the two texture regions) and to simultaneously measure the difference between the two regions.

Now suppose that one considers a "computing unit" (for lack of a better word) composed of a simple linear channel followed by steps (i) and (ii) above. The simple linear channel can be thought of as a first-stage filter, and the array of edge-detecting receptive fields in step (ii) is a second-stage filter, and there is an intermediate nonlinearity in step (i). Thus, this "computing unit" looks very much like the complex channels in Fig. 1. One minor difference is that, in this "computing unit", the receptive field of the second-stage filter has odd-symmetric (edge-detecting) receptive fields rather than the even-symmetric fields shown for the complex channels in Fig. 1. Another minor difference is the explicit presence of first-stage filters of both odd and even symmetry. A more significant difference between this computing unit and the complex channels is as follows, however:

- The "computing unit" just described is being used to tell the difference *between* two texture regions where the textures in the two regions produce different responses in simple channels (or, roughly speaking, the two textures have different Fourier amplitude spectra, or, as has sometimes been said, are "first-order" or Fourier textures).

- The complex channels in Figs 1 and 17 are being used to characterize the texture *within* any one region so that two textures which produce identical magnitudes of response in all simple channels (i.e., have very similar Fourier amplitude spectra, and thus are second-order or Fourier textures) but nonetheless differ (as in the checkerboard vs striped-arrangement textures in Fig. 2) can be told apart by higher processes and hence the observer.

Whether the "computing units" used to tell the difference between regions of first-order texture and the complex channels used to characterize second-order texture regions are in fact distinct entities (e.g. different neurons in different places in the brain) is a difficult question even to ask cleanly at this point in history, and it seems impossible to answer now.

If these entities are not the same, then the results of this study apply only to complex channels, that is, to entities used to characterize textures within a region. (This is so because the expansive nonlinearity revealed here was in the complex channels used to characterize the checkerboard texture and the striped textures in a case where both of them have extremely similar Fourier amplitude spectra and thus produce very similar output magnitudes from all simple channels.)

A similar point holds for the "higher-order mechanisms" mentioned in the Introduction, where the second stage pools across spatial frequency and/or orientation and/or phase rather than spatial position. Again, this second stage might better be compared with the pooling and decision stage (e.g. Figure 17) than to the second stage of the complex channels. For example, in the case of Olzak and Thomas (in preparation, 1996) their second filtering stage is followed by a simple differencing to form the decision variable, and these operations together are very close to those of our comparison and decision stage's spatial pooling followed by a within-channel difference.

Part II. The comparison and decision rule

This is a brief description of the comparison and decision rule stage in our models (the rightmost box in Figs 4 and 17). More extensive description and some discussion appeared in earlier publications (Sutter *et al.*, 1989; Graham, 1991; Graham *et al.*, 1992; the Appendix of Graham *et al.*, 1993). Our practice has been to compute various measures of the degree to which there are gross differences in overall modulated activity between the outputs of the channels to the checkerboard region and the outputs to the striped regions. These various measures all have the structure diagrammed in Fig. 4 but differ in the parameter values.

As Fig. 4 indicates, a spatially pooled response is computed from each channel (both simple and complex) in both the checked region and the striped region. The spatially pooled response of the jth channel to one region is taken to equal:

$$R_j(region) = \sqrt[k_{sp}]{\sum_{(x,y)\ in\ region}^{N_x}\sum^{N_y}\frac{|O_j(x,y) - M_j|^{k_{sp}}}{N_x \cdot N_y}} \qquad (A1)$$

where k_{sp} is the exponent characterizing spatial pooling, $O_j(x,y)$ is the output at position (x,y) of the jth channel, N_x and N_y are the numbers of positions in the region (usually just in one period in the appropriate region as that is sufficient). Then M_j is the average value of $O_j(x,y)$ over this one period and is very close to zero for these patterns and channels.

The observer's differential sensitivities to different spatial frequencies, orientations, and mean luminances can be incorporated into this model at several places, e.g. into the channels themselves by allowing different channels to have different peak sensitivities or by having an early sensitivity-setting stage as in Fig. 17. In either case, the sensitivity factor would be incorporated into the channel outputs $O_j(x,y)$ in the above A(1). In our earliest presentation (Sutter *et al.*, 1989), however, we acted as if the differential sensitivity were incorporated at a much later stage [corresponding to A(3) below], but that was done primarily for convenience in our computer programs and does not seem the most likely candidate. The results here are not affected by this problem in any case.

When the exponent k_{sp} is set equal to 2, the above measure is equal to the standard deviation of the outputs at different positions in one period of the given region and, by crude analogy to other situations, sometimes described as energy.

Then, a within-channel difference is computed:

$$D_j = |R_j(check) - R_j(stripe)| \qquad (A2)$$

which gives the difference between the spatially pooled response in the two regions.

Finally, the within-channel differences from the N_{ch} different channels are combined in a power-summation with an exponent k_{ch} to form D

$$D = \sqrt[k_{ch}]{\sum_{j=1}^{N_{ch}} D_j^{k_{ch}}} \qquad (A3)$$

where N_{ch} is the total number of channels. When the exponent $k_{ch} = 2$, the D is the root-mean-square difference between regions. An exponent of infinity corresponds to taking the channel that best discriminates the two regions.

These pooling rules across all spatial positions and channels can reduce to relatively simple statements involving pooling across types of channels (as in Graham & Sutter, 1996) when the exponents associated with different channels are the same. This reduction is possible because power-summation rules like the pooling rules above have the following convenient property: one can first pool over subsets of the whole set and then pool over these intermediate quantities and the answer is the same as if one had pooled over the whole set to begin with; see Eq. (4) in Graham *et al.* (1992).

The degree to which two regions (textures) segregate perceptually (as reflected by the observer's ratings of perceived segregation) is assumed to be *a monotonic function F of D*. If for no other reason, one must include this final monotonic transformation between D and the ratings because the observer's use of a bounded rating scale introduces a ceiling that does not occur for D.

If there is only one channel contributing to an observer's response, then the observer's response will be a monotonic function of the within-channel difference, and that will just equal the amplitude of the channel's output in one region minus its output in the other.

Part III. Equivalent forms of complex channels

In the text we ask a question about the intermediate nonlinearity in the complex channels because that is usually the only stage in the channels assumed to be nonlinear. More generally, however, these experiments measure the spatial-summation nonlinearity in the complex channel as a whole, and it would be possible (although at this moment in time it seems odd to do so) to place the nonlinearity elsewhere in the complex channel. To prevent future confusion about exactly what the results in this paper prove, some of these alternative complex-channel models (equivalent in this context) are described below. First, however, to set the stage for these descriptions as well as for the reader desiring further explanation of the prediction in the main text, that prediction [equation (3)] is informally proved in the next few paragraphs.

Further explanation of predictions from complex-channel model in main text

Consider the tuned complex-channel responses as shown in Fig. 6:

(i) Since the first-stage filter is tuned to the grating patches' spatial frequency and orientation, the filter's output to an element mimics the element itself. And thus the output to a large element mimics that to a small, except it occupies area A_1, whereas the output to the small occupies area A_2 (second row of Fig. 6);

(ii) The outputs then pass pointwise through a power nonlinearity. If the power nonlinearity is of the full-wave type, the output magnitude at each point is now proportional to $|C_1|^k$ and $|C_2|^k$ rather than to C_1 and C_2.

If it is a half-wave power nonlinearity, the output magnitude at each point in the positive halfwaves will be proportional to $|C_1|^k$ and $|C_2|^k$ while that in the negative halfwaves will be zero.

(iii) The modulation amplitude (call it R) in the second-stage filter output in the striped region (e.g. the final row of Fig. 6) is proportional to the maximum of that output. That maximum comes from the receptive field centered at the middle of a large element (the receptive field shown in the second to last row of Fig. 6). R is therefore approximately equal to excitatory center's response to the large element minus the inhibitory flank's response to a small element, or:

$$R \approx |A_1 \cdot |C_1|^k - A_2 \cdot |C_2|^k| \tag{A4}$$

To compute R exactly, one would need to explicitly apply a hypothetical first-stage filter, then the pointwise nonlinearity, and then the second-stage filter. But since the elements in these element-arrangement texture patterns are spatially separate, the differences between the exact computation from a tuned channel and that embodied in A(4) are insignificant for the purposes of comparison with data as noisy as psychophysical data. Also note that the verbal description just above A(4) assumes an even-symmetric receptive field for simplicity, but actually the phase characteristics of the filter are irrelevant.

(iv) The response of this tuned channel in the checkerboard region is approximately zero. Hence the overall response of this channel as it contributes to perceived segregatability (that is, the difference between its response in one region and that in another) is also given by A(4).

(v) Finally, since all the tuned complex channels will act analogously to that in Fig. 6, the total contribution of the complex channels to perceived segregatability is also given approximately by A(4).

Rearranging A(4) gives equation (3) and equation (4) of the main text.

Nonlinear second filter (with linear first filter and piecewise-linear intermediate stage)

Another possible model of a complex channel is one in which the intermediate nonlinearity is always piecewise-linear and the first-stage filter remains a linear filter, but the second-stage filter is not a linear filter; instead, the second stage does some nonlinear pooling of responses from different points in space (as if each receptive field in the second-stage filter were modeled nonlinearly). There is an indefinitely large set of possible nonlinear second stages. We consider three possible versions. We will describe them in the context of Fig. 6 following the same five steps used in the previous section. The three versions will differ only at the third step. In the end, they make the same predictions for the positions of the dips in the tradeoff curves as does the original model.

Step (i) The first-stage filter's response will be as before for all three versions.

Step (ii) For all three versions, the first-stage responses are passed pointwise through the full-wave or half-wave *piecewise-linear* nonlinearity so the response magnitude at each point is proportional to C_1 and C_2;

Step (iii) It is at this point that the three versions we are considering differ. They make different assumptions about the exact action of the second-stage filter, as subsections below will describe individually.

First, however, we will give a general description of their action at this third step.

For all three versions, the amplitude of the second-stage filter response (call it R—see the final row of Fig. 6) will be approximately proportional to the response of the receptive field centered at the middle of a large element (the receptive field shown in the second to last row of Fig. 6)

However, now, unlike the linear second-stage filter case considered in the main text, the response E of the excitatory center of each of these three versions may come from a nonlinear pooling of its inputs, as may the response I of the inhibitory surround. Further, the combination of E and I may not be linear but nonlinear in ways described below.

Steps (iv) and (v). For all three versions, the fourth and fifth steps are like that for the original model (in the preceding subsection) except that the appropriate A(5), A(6), or A(7) will be substituted for A(4).

Version #1 (power-law applied at second-stage filter rather than at intermediate nonlinearity). If the second-stage filter is nonlinear rather than linear and its nonlinearity can be described by an integration over a power function of the input, then one finds the following.

$$E \approx A_1 \cdot |C_1|^k$$

$$I \approx A_2 \cdot |C_2|^k$$

and letting

$$R = |E - I|$$

$$R \approx |A_1 \cdot |C_1|^k - A_2 \cdot |C_2|^k| \tag{A5}$$

A(5) is the same as A(4) for the model in the main text. Indeed, one has the feeling that the difference between that model and this one is close to verbal quibbling. In a physiological context, however, where particular cells were taken to correspond to particular parts of the model, it might be more natural to use one rather than the other.

Version #2 (quick pooling or Minkowski metric done separately in excitatory and inhibitory regions of second filter's receptive field). However, if E and I pool as follows,

$$E \approx \{A_1 \cdot |C_1|^k\}^{1/k}$$

$$I \approx \{A_2 \cdot |C_2|^k\}^{1/k}$$

then, letting $R = |E - I|$

$$\begin{aligned} R &\approx |\{A_1 \cdot |C_1|^k\}^{1/k} - \{A_2 \cdot |C_2|^k\}^{1/k}| \\ &= |A_1^{1/k} \cdot |C_1| - A_2^{1/k} \cdot |C_2|| \end{aligned} \tag{A6}$$

The minimum of R still occurs when the relationship in equation (3) holds. But the functions R versus C_2 will be straight lines, not curved as in Fig. 7. However, since the final monotonic transformation between the predicted segregation and the observer's segregation rating introduces curvature, it is unlikely that this difference between the predictions of this model and the others could be detected in the experimental results.

Version #3 (quick pooling or Minkowski metric done over whole receptive field). Still a third logical possibility is to let

$$E \approx A_1 \cdot |C_1|^k$$

$$I \approx A_2 \cdot |C_2|^k$$

but now let $R = |E - I|^{1/k}$ so that

$$R \approx \{A_1 \cdot |C_1|^k - A_2 \cdot |C_2|^k\}^{1/k} \tag{A7}$$

Again R will reach its minimum when the relationship in equation (3) holds. However, the precise functions for R versus C_2 are different than those in the preceding two cases. Again, though, it is unlikely that the difference between the predictions of models could be detected in these experimental results.

Nonlinear first filter (with piecewise-linear intermediate function and a linear second filter)

A compressive or expansive nonlinearity can be incorporated in the first-stage filter (instead of at the intermediate stage) by decreeing that the first-stage filter should contain two substages: one is a pointwise compressive or expansive nonlinearity; the other is a linear filter. Which of these substages comes first does have some effects on the predictions, but to a first approximation either order can mimic the effects of a nonlinear intermediate stage. If the pointwise nonlinearity comes first, the model seems significantly different from Fig. 1. If the pointwise nonlinearity comes second, then one is just quibbling about whether it should be considered to be part of a nonlinear first-stage linear filter or, as in Fig. 1, part of the intermediate

rectification stage. (At least quibbling as far as applications to psychophysics go. Again one can imagine that some exploration of the physiological substrate might make one description more desirable than the other.)

Note further, the pointwise nonlinearity of the first substage must be compressive or expansive in the following sense: it must be a symmetric function of its input centered at that value of the input produced by a blank field at the same background luminance as the patterns. Thus, it is just as compressive (or expansive) for excursions below the background luminance as for excursions above.

To see that this scheme will work for these area experiments described here, you might consider Fig. 6 again and notice that the compressive or expansive effect (that determines what contrasts are matched for effect on the complex channel) can occur either at the first-stage filter or at the intermediate nonlinearity without undermining the basic logic.