Vision Res. Vol. 32, No. 4, pp. 719–743, 1992
Printed in Great Britain.

0042-6989/92 $5.00 + 0.00

Nonlinear Processes in Spatial-frequency Channel Models of Perceived Texture Segregation: Effects of Sign and Amount of Contrast

NORMA GRAHAM,*† JACOB BECK,‡ ANNE SUTTER*

Received 23 July 1990; in revised form 3 June 1991

Observers rated the degree of segregation between two textures, each composed of the same two element types but in differing arrangements (a checkerboard arrangement in the middle region of the pattern and a striped arrangement in the top and bottom regions). The two element types in a given pattern were either both solid squares or both center-surround elements. In center-surround elements the average luminance equaled the background luminance. The two element types were identical in size but differed in sign and/or amount of contrast. Discrepancies between the observers' ratings of perceived segregation and the predictions of simple (linear) spatial-frequency and orientation channels models of texture segregation suggested adding nonlinear processes to the model. Complex channels (a rectification-type nonlinearity between two linear-filtering stages) can explain why some patterns made of center-surround elements segregate even though there is little energy at the spatial frequencies that differentiate the two textures. Complex channels cannot, however, explain the very poor segregation of "same-sign-of-contrast" patterns (where the luminances of the two element types were both far above or both far below the background). This second result might arise from a local nonlinearity preceding the channels and might be ascribed to retinal light adaptation except that it occurs at contrasts ⩽25%! Alternatively, it might arise from normalization, which may result from intracortical inhibition. Some preliminary quantitative predictions were computed from two models, one incorporating complex channels and an early local nonlinearity, the other complex channels and normalization. With suitable choices of parameters, either model could account for the results.

Texture perception Spatial-frequency channels Nonlinearity Intracortical inhibition

INTRODUCTION

Two-dimensional spatial-frequency content (spatial frequency and orientation) is a critical factor in "perceived texture segregation", that is, in how well differently patterned or textured regions of the visual field are perceived to segregate into obviously separate regions. (This has been known since the early studies of Beck and Julesz and their colleagues; for a review of the texture segregation literature, see Bergen, 1992). Analyzers that are selectively sensitive along the spatial-frequency and orientation dimensions (also called channels, detectors, pathways, mechanisms, units, neurons by various writers) are thought to exist at a relatively low level in the visual system. (For reviews of the concepts and the evidence see, e.g. DeValois & DeValois, 1988; Graham, 1989b.) Thus a number of investigators have asked how well perceived texture segregation can be explained on the basis of these low-level spatial-frequency- and orientation-selective analyzers (e.g. Beck, Sutter & Ivry, 1987; Bergen & Landy, 1991; Bovik, Clark & Geisler, 1987, 1990; Caelli, 1982, 1985, 1988; Chubb & Sperling, 1988; Clark, Bovik & Geisler, 1987; Coggins & Jain, 1985; Daugman, 1987, 1988; Fogel & Sagi, 1989; Klein & Tyler, 1986; Landy & Bergen, 1988, 1989, 1991; Malik & Perona, 1989a,b; Nothdurft, 1985a,b; Turner, 1986; Victor & Conte, 1987, 1989a,b; Victor, 1988). [The relationship between models containing spatial-frequency- and orientation-selective analyzers and Julesz's original statistical conjecture has been discussed in some detail by Klein and Tyler (1986) and Victor (1988).] Only recently, however, has there been much emphasis on quantitative comparisons between psychophysical data and quantitative predictions from models containing low-level spatial-frequency- and orientation-selective analyzers (e.g. Beck, Rosenfeld &

*Department of Psychology, Columbia University, New York, NY 10027, U.S.A.
†To whom all correspondence should be addressed.
‡University of Oregon, Eugene, Oregon, OR 97403, U.S.A.

Ivry, 1989; Bergen & Landy, 1991; Fogel & Sagi, 1989; Landy & Bergen, 1988, 1989, 1992; Malik & Perona, 1990; Sutter, Beck & Graham, 1989).

The work described here continues the quantitative comparison begun in Sutter *et al.* (1989) between the predictions of a simple spatial-frequency analyzer model and the results of texture segregation experiments using patterns like those in Fig. 1. These patterns are members of the class of *element-arrangement* texture patterns used originally by Beck, Prazdny and Rosenfeld (1983) and explored further by Beck *et al.* (1987). These patterns are composed of a uniform background on which are superimposed two types of elements; the elements are arranged in stripes in the bottom and top regions of the pattern and arranged in a checkerboard in the middle region of the pattern. (In the right half of Fig. 1, one type of element is equivalent to the background and thus not apparent.) The observer's task is to indicate (on a scale from 0 to 4) to what degree the whole pattern seems effortlessly and immediately to contain two different kinds of region (striped vs checkerboard).

The spatial-frequency channels in our simple model are linear systems, much like arrays of simple cortical cells. This model is the kind known to explain much of texture segregation (see discussions, e.g. in Bergen, 1992; Chubb & Landy, 1991; Victor, 1988). However, certain counter-demonstrations to this general class of models—based on constructing stimuli with special mathematical properties—have also been known for some time. We knew before beginning this work, therefore, that our simple model would work well but probably not perfectly. Our aim was to find out precisely what a realistic simple spatial-frequency channels model (with parameters based on physiological and psychophysical evidence) would quantitatively predict and, on the basis of systematic discrepancies between these predictions and the data, to add further visual processes (either low or higher level) to the model and then again to test the enhanced model. Similar aims seem to be propelling others (e.g. Bergen & Landy, 1991; Fogel & Sagi, 1989; Landy & Bergen, 1988, 1989, 1991; Malik & Perona, 1990; Rubenstein & Sagi, 1990; Victor, 1988; Victor & Conte, 1989a,b).

The comparisons between the predictions of the simple spatial-frequency channels model and the data from texture-element arrangement patterns like those in Fig. 1 have turned out to be very revealing, showing both the model's strengths and its weaknesses. In Sutter *et al.* (1989) we varied both the contrast and area of the individual elements. The two types of elements in a given pattern could differ in spatial characteristics (e.g. large vs small squares or long vs short lines) and in magnitude of contrast but not in sign of contrast (i.e. the luminances of the two element types were both higher or both lower than the background luminance). The simple model correctly predicted the major phenomenon in these results, namely a tradeoff between contrast and area. There were systematic deviations, however, between the simple model's predictions and the experimental results. These deviations suggested substituting nonlinear channels (that act more like complex cells in the cortex) for the simple linear channels of the original model. These complex channels were introduced in Sutter *et al.* (1989) and are discussed further below.

Here we report further studies using element-arrangement patterns. The two types of element in a given pattern were identical in spatial characteristics but differed in the sign and/or magnitude of their contrasts. Further, we studied not only solid-square elements (Fig. 1 top) but center-surround elements (where the outer square ring and the inner square had identical areas but opposite contrasts, as in Fig. 1 bottom). Using center-surround elements removes the energy at the fundamental frequency of the texture regions and thus weakens the ability of simple linear channels to signal the differences between the texture regions. Some deviations between the experimental results reported here and the simple spatial-frequency channels model's predictions can again be accounted for by introducing complex channels. But other deviations suggest the necessity of a second kind of nonlinearity, which we will call an "intensity-dependent" nonlinearity. We suggest two alternative candidates for this second kind of nonlinearity; either alternative could result from low-level visual processes occurring in cortical areas V1 or V2 or below. We then present predictions from two models, both models assuming complex spatial-frequency channels but each assuming a different candidate for the intensity-dependent nonlinearity. Both models fit the data quite well, and further work will be required to decide between them.

Simple spatial-frequency channels model

The simple spatial-frequency channels model consists both of assumptions about the channels themselves and assumptions about the relationship between the channels' outputs and the observer's response in an experiment (Sutter *et al.*, 1989). These assumptions are reviewed here.

Characteristics of simple channels (linear filters). Each simple channel is assumed to be a discretely-sampled, linear, translation-invariant filter. The physiological substrate for such a channel could be a set of linearly-acting neurons, perhaps simple cortical neurons, all sensitive to the same spatial frequency and orientation (that is, all having receptive fields of the same shape, size and sensitivity) but having receptive fields at a number of different spatial positions. The spatial positions are close enough together so that receptive fields of neighboring neurons overlap substantially. The spatial weighting function characterizing a channel (the physiological substrate for which might be the receptive field sensitivity profile of an individual neuron) was assumed to be a two-dimensional Gabor function (as used, e.g. by Bovik *et al.*, 1987, 1990; Clark *et al.*, 1987; Daugman, 1987, 1988; Field, 1987; Fogel & Sagi, 1989; Rubenstein & Sagi, 1990; Turner, 1986; Watson, 1983). The particular function makes little difference to these predictions, however. The spatial-frequency and orientation half-amplitude full-bandwidths of each filter were 1 octave

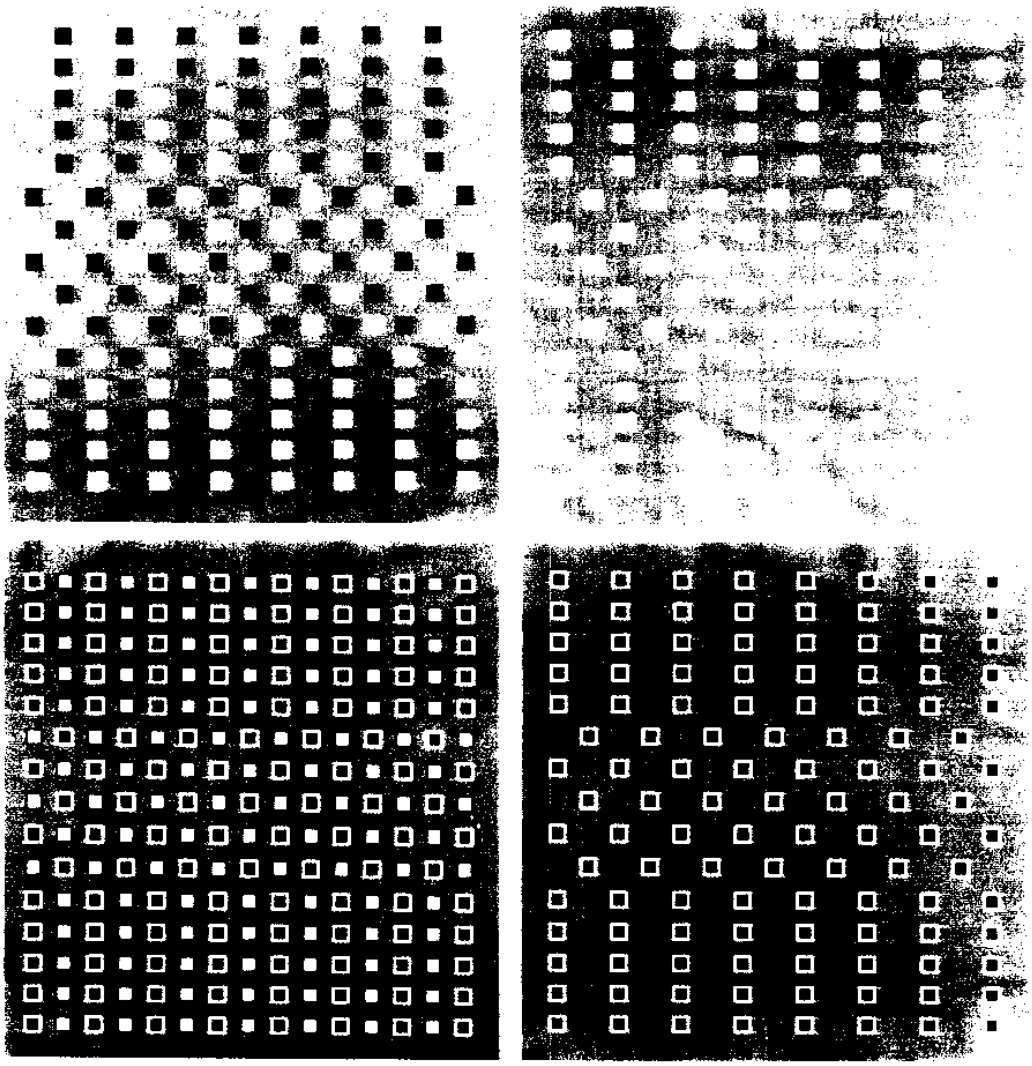

FIGURE 1. Reproductions of four patterns used in experiments. The top two have solid-square elements. The bottom two have center-surround elements. The two on the left show opposite-sign-of-contrast patterns. The two on the right show one-element-only patterns.

and 38 deg of rotation respectively (as in Watson, 1983) and in good agreement with results from near-threshold psychophysics (reviewed in Graham, 1989b). Again, within rather a large range, the precise bandwidth makes little difference for anything said here. The filters' center spatial frequencies were separated by a half-octave (each one a factor of $\sqrt{2}$ greater than the one before) from 0.25 to 16 c/deg. Three orientations of filter—vertical, horizontal, and oblique (45 deg)—were used. These three were sufficient to capture the significant features of the patterns. (Note that left oblique and right oblique act identically for the patterns used here.) The spatial weighting functions were all even-symmetric. Given our full model, however, using multiple phases at each position would make little or no difference. We assumed translation-invariance for computational simplicity. (That is, in terms of the physiological substrate, we assumed that the receptive fields at all spatial positions are identical.) It would be preferable to instead include the known changes in sensitivity with spatial position (as, e.g. Watson, 1983 does) but it would make little if any difference for the predictions reported here.

Figure 2 shows one pattern (left panel) and two simple channels' outputs to it (middle and right panel). In the left panel, there is a small square superimposed in the middle of each texture region. The width and height of this square equal one period of repetition in the horizontal and vertical directions respectively. Thus the fundamental frequency of these texture regions for vertically-oriented and horizontally-oriented receptive fields is the reciprocal of the width or height of these small squares.

The brightness at a particular point in the middle and right panels of Fig. 2 represents the magnitude of the channel's ouptut at that position (of the neuron having a receptive field center at that position) where the mid-gray represents zero, lighter grays represent positive responses, and darker grays represent negative responses. (Since neurons cannot actually yield negative responses, one needs to think slightly about what the physiological substrate for such a channel really is. The conventional answer is to use both *on* and *off* cells where on cells signal responses above zero and off cells responses below zero. See, e.g. Heeger, 1991.) The middle panel shows the output of a vertically-oriented channel with peak sensitivity to a spatial frequency near the fundamental frequency of the texture regions. One of the receptive fields from this channel is sketched as a trio of ellipses superimposed on the pattern in the left panel; the center ellipse represents the excitatory center and the side ellipses the inhibitory flanks of the receptive field. Notice that when one column of elements is stimulating the excitatory center of receptive fields of this size, neighboring columns are stimulating the inhibitory flanks. Hence the response of this channel (middle panel Fig. 2) is strong in the striped region and absent in the checkerboard region. (Further explanation of these responses can be found in Sutter *et al.*, 1989; Graham, 1989a.)

The right panel of Fig. 2 shows the output of a much higher spatial-frequency channel (still vertically-oriented). The receptive field that characterizes this channel is too small to sketch clearly in this figure. Notice in the right panel that this channel's output includes positive and negative responses along the vertical edges of all the individual elements. Since there are many elements of each type both in the striped and in the checkerboard region, this channel responds well in both regions. Indeed, there is very little (if any) difference overall between the amount of responding by this channel in the striped region and that in the checkerboard region. (There is a difference in the spatial arrangement of responses since there was a difference in the spatial arrangement of elements. This difference in spatial arrangement will be exploited by the complex channels presented subsequently.)

Figure 3 shows portions of the outputs of all 39 simple channels used in our calculations (in response to the pattern from Fig. 2 left panel). The 13 different spatial frequencies of filter are shown in 13 different columns and the 3 different orientations of filter in 3 different pairs of rows. Each small square in Fig. 3 shows a filter's output from the middle period of the checkerboard or from the middle period of the striped region (as indicated by "checked" or "striped" labels on the right); these middle periods are indicated by outline squares in the left panel of Fig. 2. The filter outputs shown in Fig. 2 appear in Fig. 3 in the upper pair of rows in the horizontal positions corresponding to 1 c/deg (for Fig. 2 middle) and 11.3 c/deg (for Fig. 2 right). The responses of the very low spatial-frequency channels, those at the left edge of Fig. 3, are responses to the outside edge of the pattern; and the differences between checked and striped regions in these very low channels reflect the fact that the striped regions were on the top and bottom of the pattern and the checked region was in the middle (rather than to characteristics of the checked or the striped regions themselves). Outputs of these 39 different channels to some other texture patterns are illustrated in Graham (1989a) and Sutter *et al.* (1989).

From channels' outputs to an observer's response. Next we discuss the assumptions that convert the channels' outputs into a quantitative prediction of the observer's rating of perceived texture segregation. As we had no good *a priori* reason to select any one way of doing this computation, we considered a whole family. As it turned out, for the conclusions presented here, all members of the family acted equivalently.

Figure 4 illustrates the several steps in the computation. The first step is to compute a *spatially-pooled response* from each channel's output in the checked region and from each channel's output in the striped region. Let the output at position (x,y) of the channel for the ith frequency and jth orientation be called $O_{ij}(x,y)$. Then the spatially-pooled response computed from the output of the ijth channel to the checked region equals

$$R_{ij}(ch) = \left[\sum_{\substack{(x,y) \text{ in one period} \\ \text{of checked region}}}^{N_x} \sum^{N_y} \frac{|O_{ij}(x,y) - A_{ij}|^k}{N_x \cdot N_y} \right]^{\frac{1}{k}} \quad (1)$$

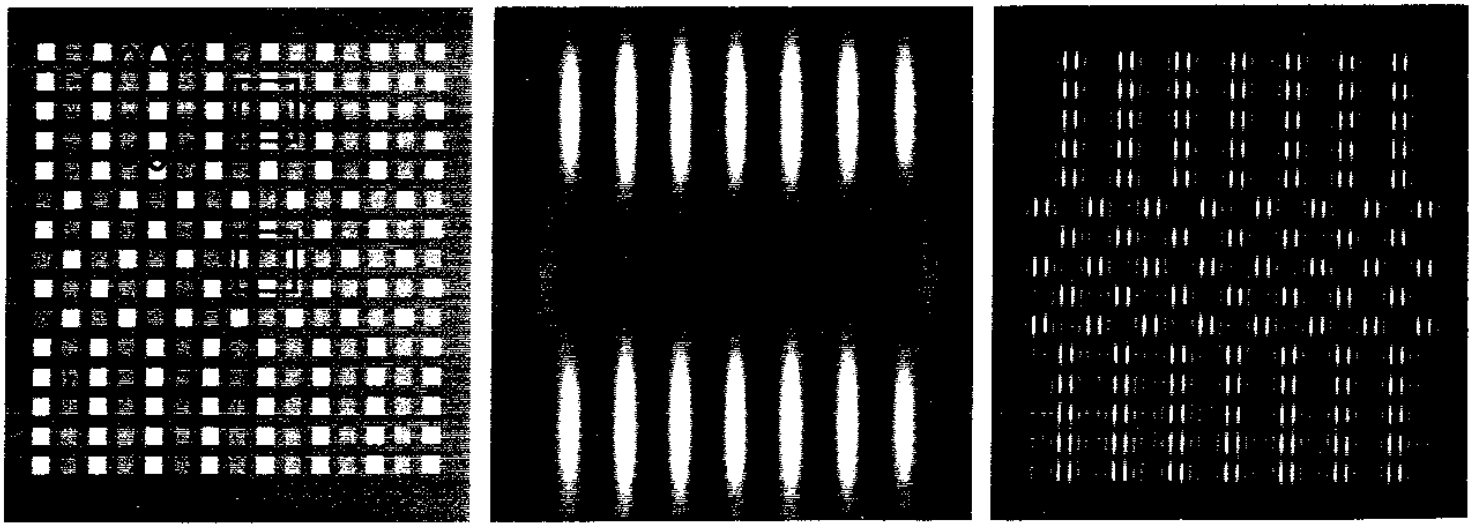

FIGURE 2. Left panel shows a same-sign-of-contrast pattern made with solid-square elements. The two outline squares superimposed on it indicate the middle period from the checkerboard region and that from the striped region. The three ellipses sketched on it represent the receptive field of one mechanism from one simple channel (one linear translation-invariant filter. This channel is most sensitive to vertical orientations and to spatial frequencies near the fundamental frequency of the texture region. The middle panel represents the output of that simple channel. The right panel represents the output of a channel tuned to spatial frequencies much higher than the fundamental frequency (about 11 times, and again to vertical orientations). Its receptive field would look like that sketched on the left panel but reduced by a factor of about 11. In both the middle and right panels, brightness indicates the magnitude of channel response at each point, where gray represents a response of zero, whites represent positive responses (excitation) and blacks represent negative responses (inhibition). (The reproductive processes have distorted the outputs somewhat.)

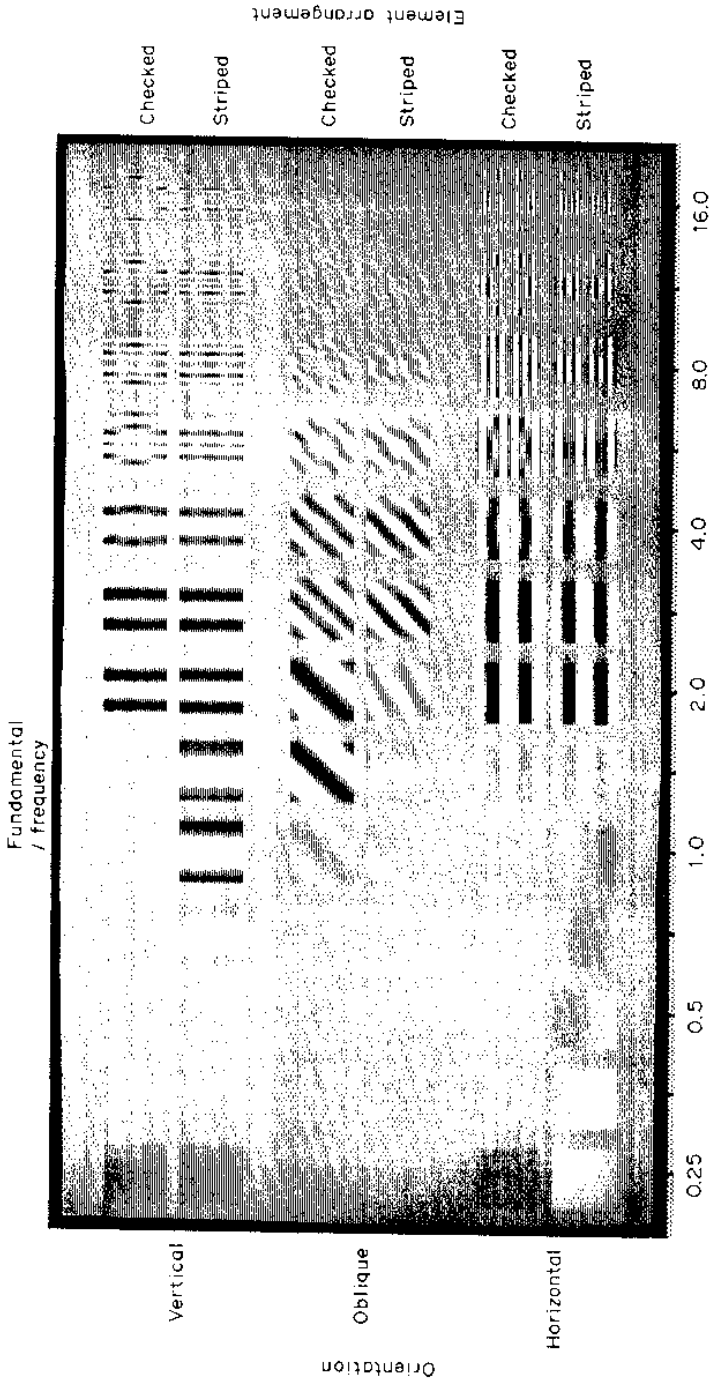

FIGURE 3. Portions from the outputs of all 39 simple channels (linear, translation-invariant filters) used in our calculations. Each small square shows a filter's output from the middle period of the checkerboard or the striped region (labels on right edge). Thirteen different spatial frequencies are shown in the different columns (horizontal labels), and three different orientations of filter in three different pairs of rows (labels on left edge).

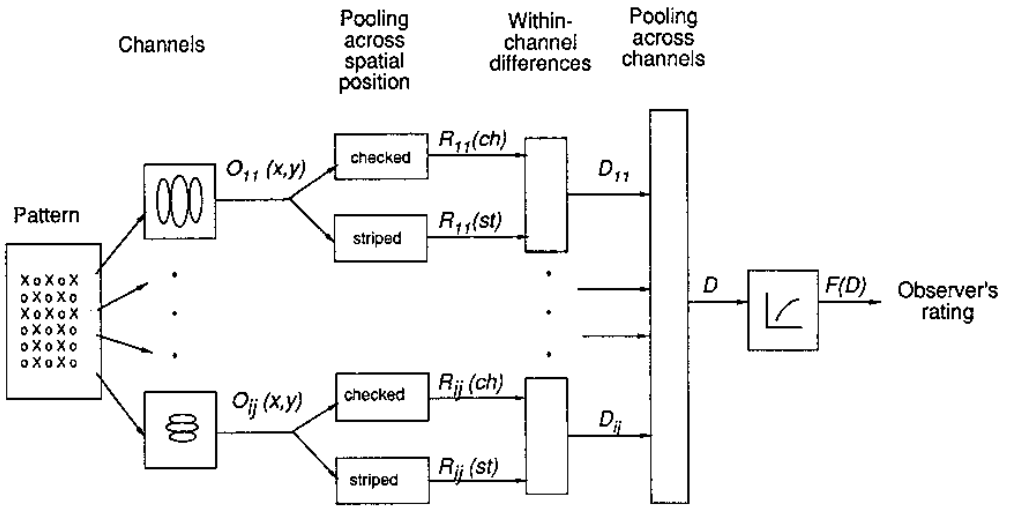

FIGURE 4. Diagram of the computations involved in converting the channels' outputs to an observer's response.

where N_x and N_y are the numbers of spatial positions in the x and y directions in one period of the pattern, and the summation is done over one period in the checked region. A_{ij} is the average value of O_{ij} over these spatial positions, and is close to zero for bandpass channels like these. When the exponent $k = 2$, the above measure is equal to the *standard deviation* of the outputs at different positions in one period of the given region. This measure is also sometimes described as *energy* by crude analogy to other situations using the same kind of mathematics. We used exponents $k = 1, 2, 3, 4$, as well as the maximum output, the minimum output, and the maximum–minimum difference between the outputs at different positions. (For the patterns here, this last measure is very close to using $k = \infty$ in the above equation.) All the general conclusions given below held for all choices of k.

The *within-channel difference* computed from the ij th channel's output is the difference between the channel's spatially-pooled response to the checked and to the striped regions. That is, for the ij th channel:

$$D_{ij} = |R_{ij}(\text{ch}) - R_{ij}(\text{st})|. \tag{2}$$

For a concrete example, consider Fig. 3. It shows large within-channel differences for four channels: the two vertically-oriented channels at spatial frequencies of 1 and 1.4 c/deg (top pair of rows, near-middle columns—which show a good deal of activity in the striped but not in the checked region) and the two obliquely-oriented channels at spatial frequencies of 1.4 and 2 c/deg (middle pair of rows, near-middle columns—which show a good deal of activity in the checked but not in the striped region). For all channels other than these four, the difference in spatially-pooled responses between the checked and striped region is very small or zero.

Next, the within-channel differences, after weighting according to the observer's sensitivities to different orientations and spatial frequencies, are pooled into an overall difference denoted by the symbol D.

$$D = \left\{ \sum_{i}^{N_{\text{freq}}} \sum_{j}^{N_{\text{orien}}} [D_{ij} \cdot S_{\text{obs}}(i,j)]^{k'} \right\}^{\frac{1}{k'}} \tag{3}$$

where $S_{\text{obs}}(i,j)$ is the observer's contrast sensitivity to the ith frequency and jth orientation, N_{freq} is the number of frequencies (13—from 2 to 128 cycles/screen in steps of $\sqrt{2}$), and N_{orien} is the number of orientations (3—horizontal, vertical and oblique). When the exponent $k' = 2$, the overall difference D is the root-mean-square difference between regions. We also used exponents 1, 3 and 4 as well as taking the maximum of all the differences. The general conclusions below held for all choices of k'. (Victor, 1988, also considers the whole family of rules produced by different exponents, but most authors have considered only the value 2.)

The sensitivity function we generally used for $S_{\text{obs}}(i,j)$ is described in Sutter *et al.* (1989) but changing it in any reasonable manner would have almost no effect on the predictions discussed in this paper. The weighting by sensitivity $S_{\text{obs}}(i,j)$ in equation (3) could be done equivalently at any earlier stage of the computation. Indeed, incorporating the sensitivity weighting in the channels' outputs themselves (the O_{ij}'s) seems the most plausible model of the visual process. (We incorporated the weighting at the later stage to facilitate comparison of different sensitivity functions.)

Finally, the observer's rating of perceived segregation is assumed to be a monotonic function of the predicted overall difference 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 4-point rating scale introduces a ceiling that does not occur for D. This monotonic transformation will be denoted $F(D)$ below.

The above assumption relating channels' outputs to observers' responses ignores the problem of how the perceived boundary between regions is actually extracted since it simply compares filter outputs on either

side of an experimenter-defined boundary. However, our measures of pooled within-channel differences are closely related to the local-energy and sum-of-quadrature-pair measures used by others (e.g. Adelson & Bergen, 1985; Bergen, 1992; Bergen & Adelson, 1986, 1988; Bergen & Landy, 1991; Bovik *et al.*, 1987, 1990; Clark *et al.*, 1987; Fogel & Sagi, 1989; Landy & Bergen, 1988, 1989, 1991; Turner, 1986). The relationship is particularly close when the exponents = 2. As these others have shown, it is possible to use such measures in a natural way to find the actual position of the boundary. (There is further discussion in Graham, 1992.) Since the computations involved are time-consuming, however, and little would be gained in predictive power for the experiments here, we have not done so.

As is clear from the above equations, the assumptions relating the channels' outputs to the observer's response do introduce some nonlinearity into the predictions of the simple model even though the individual channels are simple channels (linear systems). This introduced nonlinearity is of the kind typically found in decision rules that compress a multidimensional internal representation into the sparse set of responses given by an observer in a psychophysical experiment. It is quite different in tone from the nonlinearities discussed below and does not mimic their effects in these experiments (although it could in some situations). In any case, because of the importance of the assumptions relating the channels' outputs to an observer's response, we do feel it is important to investigate a variety of such assumptions. The families of pooling rules discussed above certainly do not exhaust all possible assumptions for this stage, but they encompass a wide and interesting variety.

Pooling over subsets. Families of pooling rules like the ones above have the following convenient property (used in a number of places below):

$$\left(\sum_{i=1}^{n}\sum_{j=1}^{n_i} X_{ij}^{k}\right)^{\frac{1}{k}} = \left(\sum_{i=1}^{n} X_i^{k}\right)^{\frac{1}{k}} \tag{4}$$

where

$$X_i = \left(\sum_{j=1}^{n_i} X_{ij}^{k}\right)^{\frac{1}{k}}.$$

That is, 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.

THE EXPERIMENTS: STIMULI AND PROCEDURES

Stimuli

In each pattern the two element types were always the same size and shape (either solid squares or center-surround elements) but the squares differed in sign of contrast (i.e. lighter or darker than the background) and/or in amount of contrast. (The element type with the greater absolute value of contrast comprised the leftmost and rightmost stripes.) The mean luminance of all the patterns was 16 ft-L and thus the appearance of the background of all patterns in this experiment was much the same, a mid-gray. Some examples are shown in Fig. 1 although they have, of course, been distorted by the reproduction processes.

We used sets of stimuli like that represented in Fig. 5 or else subsets of this full set. In Fig. 5 each solid circle represents a stimulus. (The squares and diamonds around some of the solid circles will become relevant later.) The horizontal axis gives the incremental luminance of one element type and the vertical axis the incremental luminance of the other element type. (Incremental luminance $\Delta L = L - L_0$ where L is the luminance of the point in question and L_0 is the luminance of the background. The axes in Fig. 5 show the square's incremental luminance in the case of solid-square elements and the center's incremental luminance in the case of center-surround elements.) Incremental luminances are shown in arbitrary units. Since the background luminance was kept constant, the axes could be replotted in terms of luminance. Similarly the axes could be replotted in terms of contrast where the contrast of any point in these patterns (e.g. the center of a center-surround element) will be defined relative to the background luminance as $c = (L - L_0)/L_0 = \Delta L/L_0$ where L is the luminance at the point and L_0 is the luminance of the background.

Note that the set represented in Fig. 5 contains the texture patterns defined by all possible combinations of a number of different positive and negative contrasts. Only half a matrix of solid circles is shown because the other half was presumably redundant. (Which of the two sets of elements has which luminance presumably matters very little since the elements are all identical in size and shape.)

The contrast of any individual element relative to the background was never greater than 25%; exact values will be given later with the two individual experiments. It is worth pointing out that 25% is very low for the literature on texture segregation where black/white patterns, i.e. patterns of 100% contrast, dominate; but 25% is rather large for the psychophysical near-threshold literature from which most of the best evidence for spatial-frequency channels has emerged.

Constant-difference series

From the set of stimuli diagrammed in Fig. 5, any series of stimuli occupying the same positive diagonal (see inset lower right) is of particular interest. In such a series the *difference* between the luminances of the two element types ($L_2 - L_1 = \Delta L_2 - \Delta L_1$) is held constant, while the absolute luminances of the two element types vary together. Examples of two such series are illustrated in the next figure (Fig. 6) for solid-square elements (top row) and center-surround elements (middle row). Each little drawing in the figure represents the luminance profiles of the two element types in one pattern. The luminance profiles for a constant-difference series of center-surround patterns (middle row) are analogous to

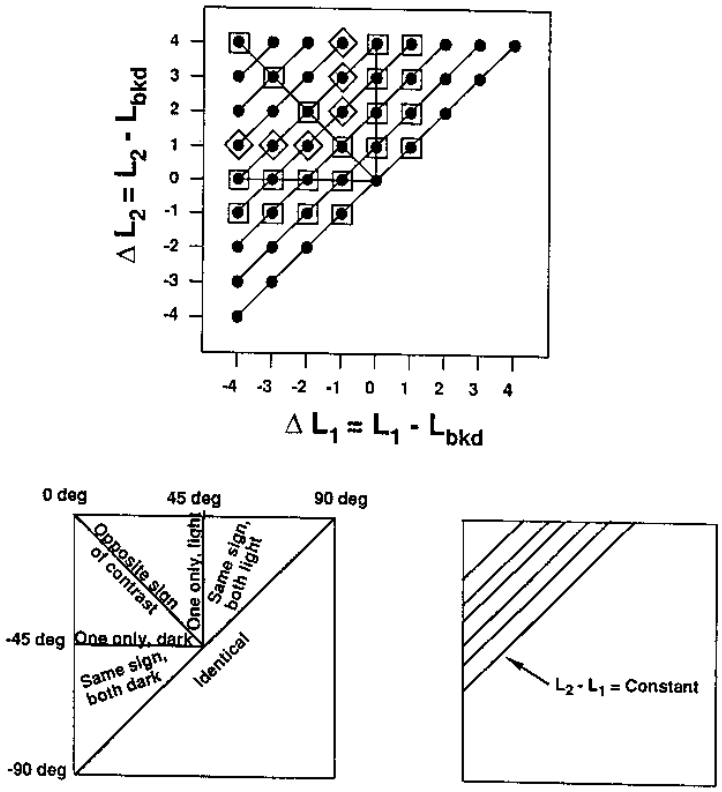

FIGURE 5. Diagram of the set of stimuli. Incremental luminance of one element type is plotted on the horizontal axis and that of the other element type on the vertical axis. Incremental luminance here is given in arbitrary units. The sizes and shapes of elements remain constant throughout a set of stimuli. Each symbol in the top panel represents one stimulus in the set. (The different symbols are explained in the text.) Lines along which the difference between the luminances of the two element types is constant are shown in the lower right. Lines along which the radio of the incremental luminances (equivalently, the contrasts) of the two element types is constant are shown in the lower-left diagram, and the corresponding contrast-ratio angle is labeled on the outside of the square.

those for the solid squares (top row) except that it is the difference between the centers' luminances that stays constant. In a *constant-difference* series, which generally contains more patterns than the five shown here, there can be patterns where both element types have the *same sign of contrast* (both darker or both lighter than the background—an example is shown in Fig. 2 left panel), patterns where one element type has the same luminance as the background so there is *one element type only* apparent (which can be either dark or light—examples are shown in Fig. 1 right) and patterns where the two element types are of *opposite sign of contrast* (examples are shown in Fig. 1 left).

Notice also that all patterns on a line through the origin in diagrams like Fig. 5 have the same incremental luminance ratio ($\Delta L_2/\Delta L_1$) and thus the same contrast ratio. For example, as diagrammed further in the lower left of Fig. 5, the patterns on the negative diagonal have element types with opposite (but equal) contrasts; those on the vertical ray upwards from the origin have one element type only, and that element type is bright (or with a bright center). Rather than using the pair of luminances (L_2 and L_1) to describe a stimulus pattern, therefore, another pair of values has proven very useful: (1) the luminance difference $L_2 - L_1 = \Delta L_2 - \Delta L_1$; and (2) the angle, which we will call the contrast-ratio angle, between the negative diagonal through the origin and the line going through the origin and the point representing a stimulus pattern. The value of this angle is shown labeling the end of several rays in Fig. 5 lower left. (This angle can be computed using the four-quadrant arctangent of $\Delta L_2/\Delta L_1$.) See also the bottom labels of Fig. 6.

The exact stimulus sets used in the two experiments will be described along with the results from those experiments.

Some details of the patterns

The viewing distance was 6 ft throughout the experiments reported here. At that distance 1 pixel subtended 1.08 min of visual angle. The solid-square elements and the center-surround elements had the same outer dimension (14 pixels = 15.12 min on a side). The inner square in the center-surround element (10 pixels = 10.8 min) was chosen so that the area of the center (the inner square) and that of the surround (the border around the square) were as close as possible to being equal. (They were not quite equal, the center area being 100 pixels^2 and the surround area being 96 pixels^2). The contrast in the center of the center-surround element (relative to the background luminance) was always opposite and equal

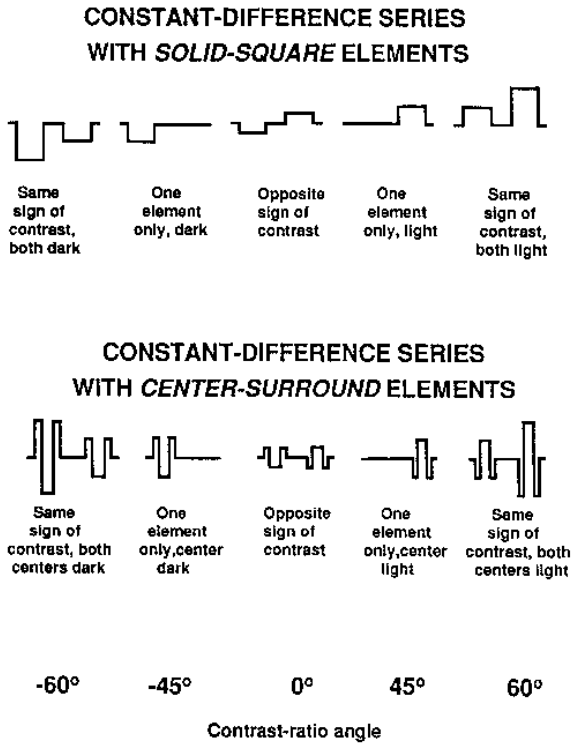

FIGURE 6. Each small diagram shows the luminance profile of the two element types in a single pattern within a constant-difference series of patterns having either solid-square elements (top) or center-surround elements (bottom).

to that in the surround (relative to the background luminance). Thus the average luminance across a center-surround element was almost exactly equal to the background luminance (although not quite because of the 4% difference between the center's and surround's areas and the discreteness of the luminance levels available on the display). The inter-element spacing was always the same in both horizontal and vertical directions (12 pixels = 12.96 min). There were always 15 rows and columns of elements as shown in the examples of Figs 1 and 2.

A period over which repetition occurs in both the checked or the striped region is a square region with its horizontal extent being two columns of elements and two columns of inter-element spaces (52 pixels = 56.16 min of arc in these experiments) and its vertical extent being two rows of elements and two rows of interelement spaces (again 52 pixels = 56.16 min of arc). The square outlines in the left panel of Fig. 2 indicate single periods in the middle of the checkerboard and striped regions.

The background luminance of the patterns, and also the luminance of the blank screen that was presented whenever a pattern was absent, was 16 ft-L. As shown in Fig. 1, there were 15 columns and 15 rows of elements, with the top third and bottom third being in a striped arrangement and the middle third in a checkerboard arrangement. The full spatial extent of the patterns was 454 pixels (8.2 deg) high × 576 pixels (10.4 deg) wide. Outside this extent the screen and room were very dimly illuminated (0.05 ft-c). The patterns were presented for 1 sec with abrupt onsets and offsets. The subject sat in a chair with unrestrained viewing.

Due to the discrete number of gray levels on the display, the luminance levels used could not be made exactly equally spaced although we have been talking as if they were. The display was carefully calibrated, however, and the discrete levels chosen to make the calibrated luminances as close as possible to evenly spaced. The biggest discrepancy was 0.1 ft-L which is not big enough to affect any of the conclusions presented below.

Further details of the equipment are given in Sutter *et al.* (1989).

Procedure

The subject initiated each trial by pressing a button. A fixation mark (a blue X) appeared in the center of the screen for 1 sec before the pattern onset. After each trial, the subject rated the degree of perceived segregation on a scale from 0 to 4. They were instructed that 0 meant that the three regions of the pattern were not distinguishable from each other without scrutiny, and 4 meant that the three regions were very distinct and that segregation was "immediate." Subjects wrote their ratings on a form supplied for that purpose.*

There are two experiments reported below. Each experiment used ten subjects for one session each. Experiment 1 involved 96 patterns. Experiment 2 involved 66 patterns. Each subject rated each pattern five times in a session.

Further details of the instructions to the subject and the rating procedure are given in Sutter *et al.* (1989).

SIMPLE-MODEL PREDICTIONS

Constant-difference series of patterns are of particular interest because models of texture segregation involving linear, spatial-frequency channels (in particular, our simple model) make simple predictions for such series. The predictions from our simple model are shown in Fig. 7. Contrast-ratio angle is given on the horizontal axis. Each curve connects the points for patterns in a constant-difference series (patterns on an $L_1 - L_2 = constant$ line in the diagram of Fig. 5). And the model's predicted value of the overall difference D (which is assumed to be monotonic with the observer's rating in the experiment) is given on the vertical axis. The predictions shown in Fig. 7 are for exponents of 2 in both the

*In subsequent experiments (not reported here) we automated the rating procedure, with subjects pressing buttons instead of writing down their ratings. The results of these experiments showed some important differences from those reported here. We believe that the differences are due to a marked increase in the speed with which the subjects ran through the trials under the automated procedure, thus encouraging ratings based on local rather than global information, and possibly contaminated by strong afterimages. When we imposed a 1 sec delay between the stimulus offset and the subject's response in the automated procedure, we were able to replicate the results reported here.

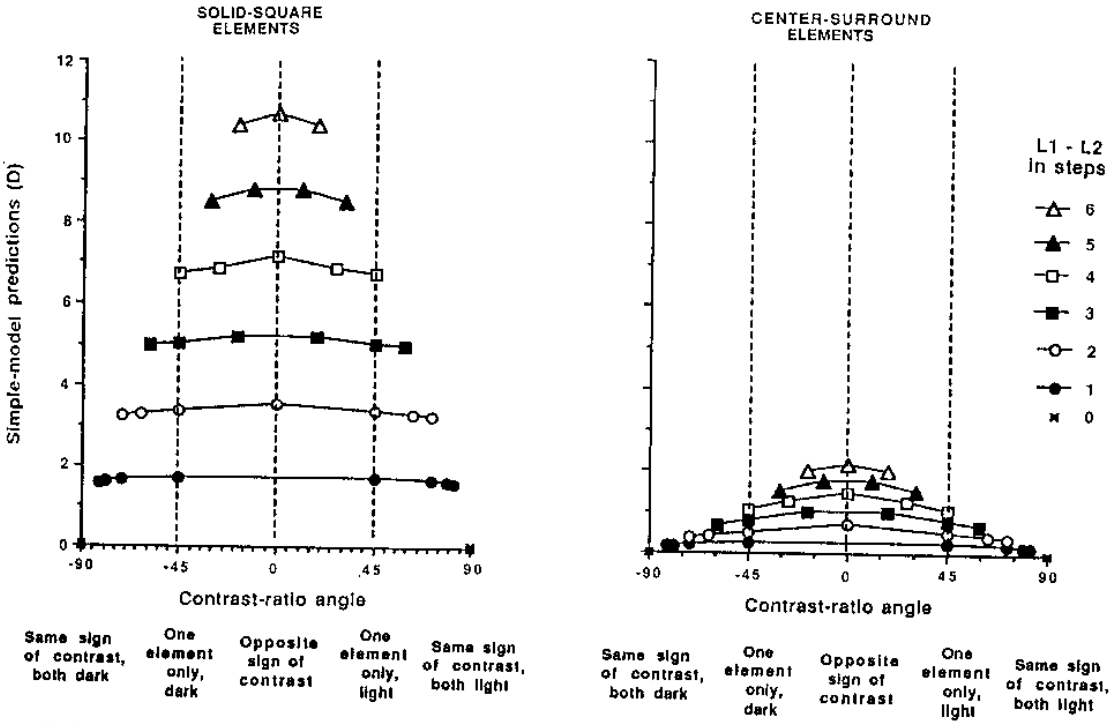

FIGURE 7. Predictions of the *simple-channels model* for patterns made from solid-square elements (left panel) or center-surround elements (right panel). The predicted value plotted on the vertical axis here is assumed to be monotonic with the observers' ratings of perceived segregation. The contrast-ratio angle characterizing a pattern is plotted on the horizontal axis. Each curve in the figure represents patterns in a constant-difference series (indicated in terms of numbers of "steps" of luminance in the right legend).

pooling rules, but varying these exponents makes little difference.

As shown in Fig. 7, the simple model makes three clear predictions for these stimuli: (i) all members of a constant-difference series of patterns have approximately the same value of the overall difference D (the curves are approximately horizontal lines in Fig. 7) and thus should segregate to approximately the same extent; (ii) the overall difference D is approximately proportional to the absolute value of the difference between the elements' luminances, $|L_1 - L_2|$ as reflected in the even vertical spacing of the set of surves in either panel of Fig. 7 (although the observer's rating may not reflect this direct proportionality since there is a monotonic transformation between D and the observer's ratings); and (iii) patterns with solid-square elements should segregate much more than those with center-surround elements (the curves are much higher in the left than in the right panel of Fig. 7).

The simple model makes these predictions because the only simple channels that can signal the difference between the checkerboard and the striped regions are the channels sensitive to frequencies near the fundamental (e.g. the channel having receptive fields like that sketched on the left panel of Fig. 2 and having the output shown in the middle panel). Since the background stimulates both the excitatory center and inhibitory surround of these channels' receptive fields to approximately the same extent, it is only the two element types that matters. These two element types have identical spatial characteristics and are spaced so that when one type is in the receptive field center, the other is in the surround. Thus, these receptive fields essentially subtract the luminance of one type of element from that of the other; that is, they respond proportionally to the difference between the two element types' luminances [as in (ii) above]. Also, the average luminance across each center-surround element approximately equals the background luminance and each element is smaller than either the excitatory or inhibitory region of the fundamental channels' receptive fields. Thus a center-surround element has the same effect on these receptive fields as the background has, and there is no net response from these receptive fields either in the checked or in the striped region. Hence the center-surround element patterns are predicted to segregate little if at all [see (iii) above].

A detail. Given these explanations, you might wonder why the curves in Fig. 7 are not perfectly horizontal and why the curves for the center-surround-element patterns are not at zero. Inspection of the computations shows that the deviation from horizontal for both the solid-square and the center-surround element patterns results primarily from channels tuned to frequencies about twice the fundamental. (For the vertically-oriented channel of this frequency, the receptive field is of such a size that when its center is stimulated by a column of elements, its suround is stimulated

by columns of inter-element spaces.) The predicted non-zero segregation of the center-surround patterns results not only from these channels at twice the fundamental but also from channels tuned to the fundamental since the average luminance over the center-surround elements was not quite equal to the background luminance. The predictions shown in Fig. 7 are for exponents of 2 in both the pooling across spatial position and the pooling across channels. When the exponents are 1 for both, the curves are somewhat more peaked in the center than those shown. When the exponents are greater than two, the curves are somewhat flatter.

Approximation

The following approximation to the actual prediction of the simple model will be useful later. Let's represent by D_A the overall difference predicted by the simple model, that is, the value of D that is predicted from the simple model and equations (1)–(3) and plotted on the vertical axis of Fig. 7. Then, the following is approximately true for the simple model (because it is true of the channels at the fundamental and they dominate the predictions—as we have just seen in Fig. 7):

$$D_A = w_A \cdot |L_1 - L_2| = w_A \cdot |\Delta L_1 - \Delta L_2| \qquad (5)$$

where w_A is a constant that may depend on the spatial parameters of the pattern (e.g. center-surround vs solid-square elements) but not on the luminances of the elements. Notice that equation (5) incorporates directly the first two predictions mentioned above. Prediction (iii) is incorporated by making the constant w_A much smaller for center-surround element patterns than for solid-square element patterns.

Results of Experiment 1

Experiment 1 included both solid-square element and center-surround element patterns, but it only included subsets of the set shown in Fig. 5 (in order to be able to intermix all the patterns we wished to compare within a single session). It included 20 patterns made of solid-square elements (symbols enclosed in open squares in Fig. 5) and 26 patterns made of center-surround elements (symbols enclosed in either squares or diamonds in Fig. 5). Experiment 1 also included some other patterns which are not described here as they yielded no extra information. The size of the incremental-luminance step ΔL in this experiment was 0.90 ft-L. Since the background was 16 ft-L, this is a step in contrast $\Delta L/L$ of 5.6%. The largest contrast (4 steps) was, therefore, 22.5%.

The results from this experiment are shown in Fig. 8 in the same format as the predictions in Fig. 7. (The average across all observers and all presentations of each stimulus is shown in this and subsequent figures. Standard error bars are omitted for visual clarity, but perusal of the regularities in the results will make it clear that the sampling variability is small relative to all the effects discussed here.)

As is obvious, the results in Fig. 8 do not agree very well with the simple model's predictions in Fig. 7. No monotonic transformation of the predicted curves in Fig. 7 (no vertical expansion or contraction, even if

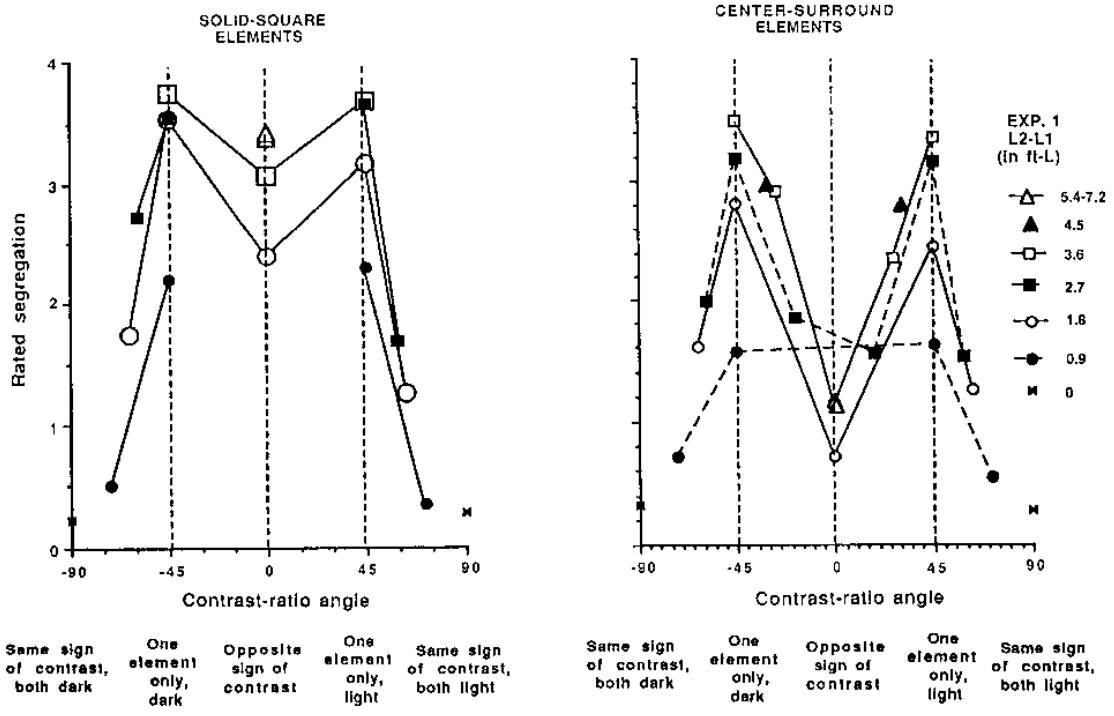

FIGURE 8. Results from Experiment 1. Observer's ratings of perceived segregation are plotted on the vertical axis. Other conventions like previous figure except that the luminance difference is given in ft-L.

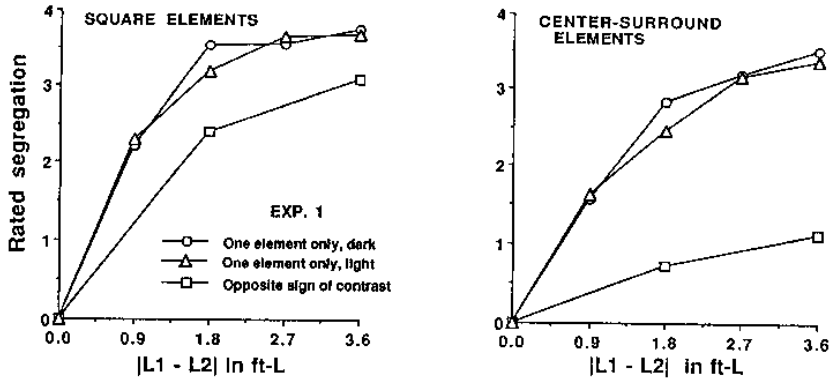

FIGURE 9. A different plot of the results in Fig. 8 for the opposite-sign-of-contrast (with equal magnitudes) and one-element-only patterns. Rated segregation is plotted against the absolute value of the difference between the luminances of the two element types.

done non-uniformly) will bring them into satisfactory correspondence with the experimental results of Fig. 8. First, many of the center-surround patterns segregate as well as corresponding solid-square patterns. Second, the curves are far from being flat. Each curve sinks dramatically at its center as well as at both its ends producing two "ears" at the one-element-only patterns. In other words, there is maximal segregation for the one-element-type-only patterns with somewhat less segregation for the opposite-sign-of-contrast patterns (in the middle of the curves) as well as much less segregation for the same-sign-of-contrast patterns (at the ends of the curves).

Figure 9 is an alternative view of some of the results in Fig. 8. It shows the observers' ratings plotted vs the absolute value of the luminance difference $|L_1 - L_2|$ for one-element-only patterns and for pure opposite-sign-of-contrast patterns. According to the simple model, all three curves in either panel of Fig. 9 should be superimposed since the observer's ratings of perceived segregation should depend only on $|L_1 - L_2|$ for a given set of spatial parameters (e.g. for the solid-square patterns or for the center-surround patterns). If the curves were superimposed, thus confirming the simple model, the shape of the single curve formed by their superpositions would reflect the shape of the monotonic transformation F that relates the overall difference D to the observers' ratings. Clearly, however, the curves are not superimposed in either panel.

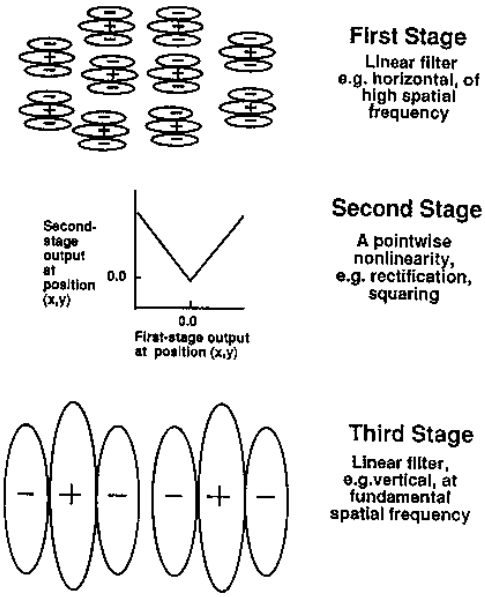

FIGURE 10. Diagram of the three stages in a complex channel.

COMPLEX CHANNELS

Some of the discrepancies between the simple model predictions and the experimental results for constant-difference series of experiments may be resolved by replacing or supplementing the simple channels with *complex channels*. As illustrated in Fig. 10 each complex channel has three stages: two stages of linear filters separated by a pointwise nonlinearity that is dramatic near zero. The nonlinearity might be a full-wave rectification (taking the absolute value of the first filter's output at each point in space), a half-wave rectification (substituting zero for the negative values and leaving the positive values untouched), or perhaps a squaring operation (squaring the output at each point). Here we assume a full-wave rectification but, if we assumed a half-wave rectification, the conclusions would remain the same. The assumptions relating the channels' outputs to the observer's responses remain the same as in the simple model.

We introduced complex channels earlier (Sutter *et al.*, 1989), and simple and complex channels here correspond closely to the first- and second-order systems of Chubb and Sperling (1988) (also Sperling, 1989; Sperling & Chubb, 1989; Sutter, Sperling & Chubb, 1991). Others have also suggested the use of similar calculations in tasks like texture segregation (e.g. Robson, 1980; Grossberg & Mingolla, 1985; Fogel & Sagi, 1989). These complex channels are more complicated than the simple channels above in much the same way that complex cortical cells seem to be more complicated than simple cortical cells (e.g. Hochstein & Spitzer, 1985) although

the post-rectification filtering in complex cells is often thought to be lowpass rather than bandpass. In any case, it is premature to take this possible physiological analog too seriously. Others in the texture literature have used the analogy to complex cells in a somewhat different way (e.g. Bergen & Adelson, 1986; Landy & Bergen, 1991). They have noted that the combination of simple linear filters followed by spatial pooling as in equation (1) is like the operation performed by complex cells. We will reserve the word "complex" here, however, to refer to a rectification-like nonlinearity imbedded between two stages of bandpass filtering, all of which comes before spatial pooling (and the subsequent pooling across channels).

A pointwise nonlinearity before a single stage of linear filtering (rather than a pointwise nonlinearity sandwiched between two layers of linear filtering) works in principle to explain many previously-noted failures of simple linear channels (e.g. Peli, 1987) and, indeed, it will explain some failures in our experiments. We will return to this point below in the section on an early pointwise nonlinearity. However, by itself it cannot fully explain our present results nor those of Sutter *et al.* (1989) and Malik and Perona (1990).

The complex channels needed to explain the results in Figs 8 and 9 have (1) a first filter that is sensitive to relatively high spatial frequencies (for example, the filter in Fig. 2 right panel) and (2) a second filter, after the rectification-type nonlinearity, that is sensitive to relatively low spatial frequencies at or near the fundamental frequency of the texture pattern. For convenience we will call such a channel a *high–low complex channel* or sometimes simply a *high–low channel*. That such complex channels exist is supported by evidence obtained with other texture patterns which suggests that over the whole population of complex channels (second-order texture systems) the first filter's preferred frequency tends to be about 3 or 4 octaves higher than the second filter's (Sutter *et al.*, 1991).

Figure 11 illustrates a high–low channel's outputs to portions of the striped region from both opposite-sign-of-contrast (left) and one-element-only (right) patterns. The patterns sketched in the figure contain square elements, but substituting center-surround elements make little difference to the following argument. In this example, this first-stage high spatial-frequency filter is characterized by the small horizontal receptive field superimposed on the stimulus patterns (top row). This filter responds at the appropriately-oriented edges of elements yielding positive and negative outputs at all element edges. A caricature of such outputs is shown in the second row of Fig. 11. (See Fig. 2 right panel for a more veridical illustration of a similar response.) These outputs are then rectified (third row Fig. 11) leaving positive outputs wherever there is an element edge in the original pattern. The third stage is a filtering done at the fundamental frequency of the texture region, and this filter is characterized by the large vertical receptive field sketched on these rectified outputs in the third row. This filter does not respond at all in the striped region of opposite-sign-of-contrast patterns since both the excitatory and inhibitory regions of the receptive fields are stimulated to the same degree (as shown in the left panel of the bottom row of Fig. 11. For the same reason, this filter also does not respond in the checkerboard region. Thus this high–low channel *cannot* signal the difference between regions in the opposite-sign-of-contrast pattern. The one-element-only pattern produces quite a different result. The third-stage vertical filter will respond strongly to the rectified outputs from the striped region of such patterns (bottom right of Fig. 11). It does not, however, respond in the checkerboard region because similar amounts of rectified output fall into its center and surround. Thus this high–low channel *can* signal the difference between regions in one-element-only patterns. Similarly it can signal the difference in same-sign-of-contrast patterns because the magnitude of the rectified outputs to one element type is greater than that to the other.

High–low complex channels might be briefly described by saying they respond to *low spatial-frequency arrangements of high spatial-frequency elements*. For those aided by radio analogies, one might say these channels demodulate the responses at high spatial-frequencies to allow recognition of low spatial-frequency modulating signals, as in AM radios (see Sperling, 1989 for example). With a full-wave rectification, such channels lose information about the sign of contrast.

The empirically observed segregation of the center-surround-element patterns (Figs 8 and 9 right panels) is as predicted by the activity of high–low complex channels: the opposite-sign-of contrast patterns do not segregate at all but the one-element-only patterns do. The solid-square element patterns, however, act somewhat differently: in particular, the opposite-sign-of-contrast element types do segregate to some extent (although substantially less than the one-element-only patterns). Taken together, therefore, these results cannot be explained with just high–low complex channels.

The results are consistent with a model in which high–low complex channels exist alongside another kind of channel. This other kind can include: (i) simple channels at the fundamental frequency; and/or (ii) complex channels whose first and third stages are both filters at the fundamental frequency. These two types of channels are both well described by equation (1). Since they respond very similarly to all patterns considered here, we will not distinguish between them, referring to either or both simply as *channels at the fundamental* or *fundamental channels*. To repeat what was said earlier, these fundamental channels respond approximately equally to all the patterns in a constant-different series, responding well to solid-square element patterns but poorly to center-surround element patterns. Thus, fundamental and high–low channels acting together will, at least qualitatively, account for the opposite-sign-of-contrast and one-element-only results in Figs 8 and 9. Some quantitative predictions are presented below.

Straightforward computation of such a model's predictions is tedious and time-consuming due to the large amount of spatial filtering involved. For the experiments

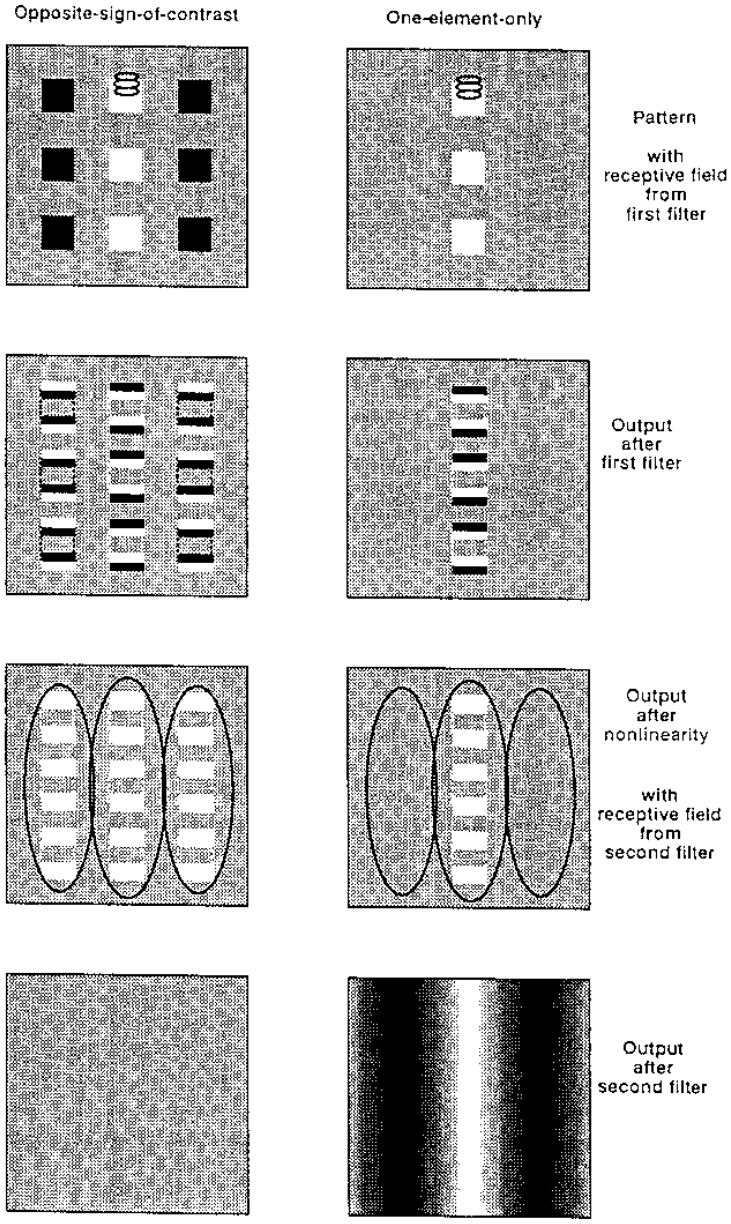

FIGURE 11. Diagram of outputs from a complex channel's three stages to the striped region of opposite-sign-of-contrast and one-element-only patterns.

described here, however, the spatial characteristics of the patterns do not vary much (the only change being from solid-square elements to center-surround elements). Thus it is possible to avoid computing the outputs of the spatial filters by using two equations that are good approximations to the full model: equation (1) above and another similar equation described below. In these equations the effects of the spatial properties are incorporated into a constant, the value of which can be estimated directly from the experimental results.

Approximation to complex + simple model

The following equation is a good approximation to the pooled effect of the high–low complex channels, a quantity we will call D_B. The number and relative sensitivities of different channels are absorbed into the parameter w_B

$$D_B = w_B \cdot |\,|\Delta L_1| - |\Delta L_2|\,|. \tag{6}$$

The parameter w_B depends on the spatial parameters of the pattern (e.g. center-surround vs solid-square elements) but not on the luminances of the elements.

This is just like equation (5) for the simple channel at the fundamental frequency except that $|\Delta L_i|$ has replaced ΔL_i. To see why, consider the following: the first filter in a high–low complex channel responds to higher harmonics, i.e. edges in these patterns, and therefore responds in proportion to the edge contrast ΔL_i.

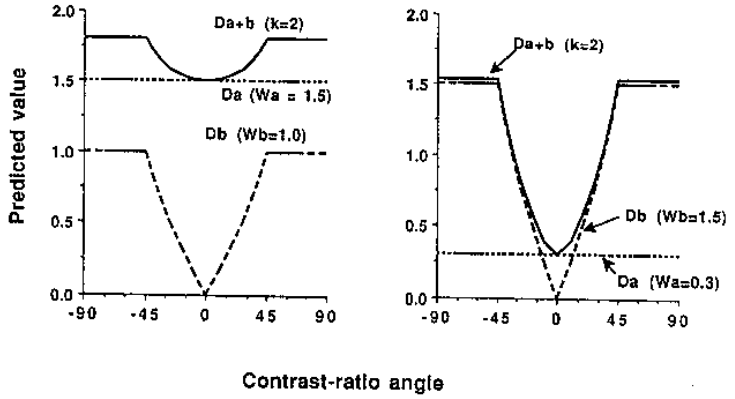

FIGURE 12. Predictions of channels at the fundamental frequency (curve marked Da), or high–low complex channels (curve Db), and of both pooled together using an exponent of 2 (curve marked Da + b) for a constant-difference series of patterns. Contrast-ratio angle is shown on the horizontal axis and predicted value (which should be monotonic with the observer's rated segregation) is shown on the vertical axis. The values chosen for the weights (Wa and Wb) in the left panel are appropriate for patterns having solid-square elements and those in the right panel for patterns having center-surround elements.

The second-stage nonlinearity then rectifies ΔL_i to $|\Delta L_i|$ before sending it on as an input to the third-stage filtering. Then the output of the third stage, which is a linear filtering done at the fundamental frequency of the texture, will depend on its inputs, $|\Delta L_1|$ and $|\Delta L_2|$, in just the same way that the output of a simple channel at the fundamental depended on its inputs, ΔL_1 and ΔL_2.

The dashed curve in each panel of Fig. 12 is a plot of D_B vs contrast-ratio angle for one constant-difference series. The value of w_B is different for the two panels. Each dashed curve is flat for contrast-ratio angles $> +45°$ or $< -45°$, that is, for all same sign-of-contrast patterns and for one-element-only patterns. But as the contrast-ratio angle gets still closer to 0°, the value of D_B drops, reaching zero at a contrast-ratio-angle of 0° (for opposite-sign-of-contrast patterns).

Also in each panel of Fig. 12 there is a dotted curve showing D_A from equation (1), that is, the prediction from all channels at the fundamental. The value of w_A is different for the two panels.

To predict segregation of constant-difference series of patterns assuming there are both high–low complex channels and channels at the fundamental, one can pool over these two subgroups of channels rather than pooling over individual channels [see equation (4)]. This yields

$$D_{A+B} = (D_A^{k'} + D_B^{k'})^{\frac{1}{k'}} \tag{7}$$

where k' is the exponent for pooling across channels, D_{A+B} is the predicted overall difference, and, as before,

$$D_A = w_A \cdot |\Delta L_1 - \Delta L_2|$$

$$D_B = w_B \cdot ||\Delta L_1| - |\Delta L_2||.$$

Then observers' ratings are presumed to be a monotonic function of the overall difference D_{A+B}.

In a full model involving all possible complex and/or simple channels, there are channels other than the fundamental and high–low channels. For example, there would be complex channels in which both stages of filtering are tuned to frequencies higher than the fundamental. Until one has done more computations, it is not clear that these other channels can be completely ignored as we are doing here. But qualitative arguments and some calculations suggest that their effects should be a good deal smaller than the effects of these two kinds of channel.

The value of w_A, the weighting constant on the fundamental channels, is expected to be higher for solid-square element patterns than for center-surround element patterns since the latter have little energy at the fundamental. The value of w_B, the weighting constant for the high–low complex channels, could be quite high for both kinds of patterns, although not necessarily identical. The values for the left and right panels of Fig. 12 were chosen accordingly to make the left panel more appropriate for solid-square element patterns and the right panel for center-surround element patterns. The solid curve in each panel in Fig. 12 shows the predictions from equation (7) using an exponent $k' = 2$. Notice that, while these predicted curves are rather like the experimental results in the region from $-45°$ to $+45°$, these predicted curves are flat for all angles outside that region (for same-sign-of-contrast patterns). This flatness—which is quite unlike the experimental results shown in Fig. 8—will occur in all predictions from this simple + complex model regardless of the values of w_A, w_B, and k'. Before further discussing this discrepancy between the complex + simple channels model and the experimental results, let us look at results from a second experiment, which explored a wider range of same-sign-of-contrast patterns and thus allows this discrepancy to appear in more detail.

Results of Experiment 2

Experiment 2 included only solid-square element patterns in order to allow more stimuli and thus a wider range of same-sign-of-contrast angles (angles further toward $-90°$ or $+90°$). The set of stimuli was like that in Fig. 5 except it contained 5 ΔL steps rather than 4

steps and thus was made up of 66 patterns. The size of the incremental-luminance step ΔL in this experiment was 0.75 ft-L. Since the background was 16 ft-L, this is a step in contrast $\Delta L/L$ of 4.7%. The largest contrast (5 steps) was, therefore, 23.4%.

The results from Experiment 2 are shown in Figs 13 and 14. Like the results of Experiment 1, they show maximal segregation for the one-element-only patterns. They also show clearly that same-sign-of-contrast patterns segregate much less well than the other types. Indeed, the curves for constant-difference series plotted against contrast-ratio angle sink dramatically at their ends with different curves actually converging (Fig. 13). In other words, for the same-sign-of-contrast patterns where both elements' luminances are rather far from the background, only the ratio of the contrasts in the two element types matters for segregation; the difference between the luminances and the absolute values of the luminances does not matter. Such a result is not predicted either by the simple-channels model or by the simple-plus-complex-channels model (as was shown in Figs 7 and 12).

THE INTENSITY-DEPENDENT NONLINEARITY—TWO CANDIDATE PROCESSES

Thus, a second nonlinear process is needed to explain the convergence of the curves for same-sign-of-contrast patterns. (*A priori*, one might find that, in adding a second nonlinear process, the first becomes unnecessary. In the present case, however, the complex-channels remain essential, as will be described further below.) When a general name is needed, we will call this the *intensity-dependent nonlinearity* and contrast it with a *spatial nonlinearity* like that embodied in the complex channels. We will discuss two possible candidates for this intensity-dependent nonlinearity, either of which can account for the results from this study: (i) a local (pointwise) nonlinearity occurring early (before the channels); and (ii) a normalization process operating among the channels, perhaps a result of intracortical inhibition.

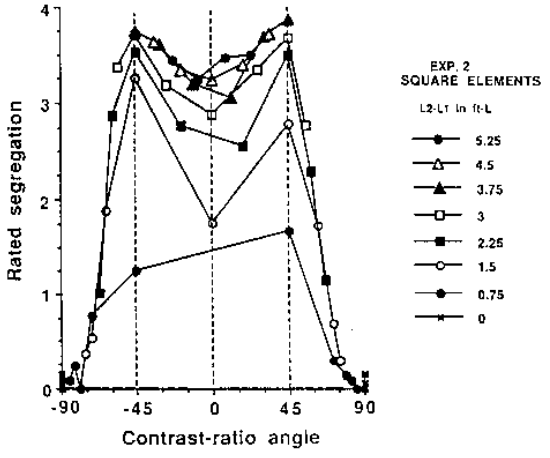

FIGURE 13. Results from Experiment 2, which only used patterns having solid-square elements. Conventions as in Fig. 8.

VR 32/4—G

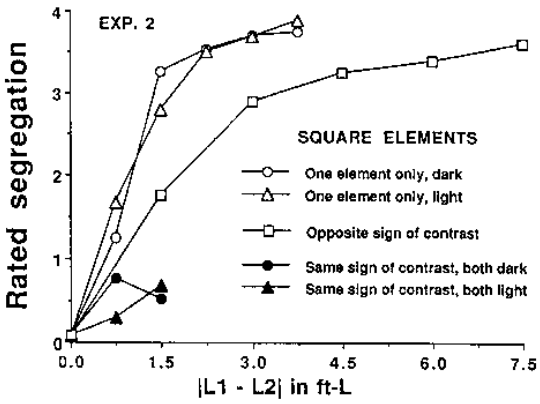

FIGURE 14. Some results from Experiment 2 plotted as a function of luminance difference. The results shown are for the opposite-sign-of-contrast (equal magnitude), the two one-element-only cases, and two cases of same-sign-of-contrast (those where the contrast-ratio was 2, and the contrast-ratio-angles +/− 72 deg).

An early local nonlinearity—one possibility

The decrease in segregation for same-sign-of-contrast patterns where the elements are far from the background might result from an early local nonlinearity preceding the channels, that is, presuming the channels are cortical, from something happening in the retina or LGN. Suppose, for example, as illustrated in Fig. 15, that early local light adaptation processes readjust the operating range of the visual system to be centered on the recent mean luminance—the background luminance in this case. This maximizes discriminability between luminances near that level and sacrifices discriminability for luminances far away. (Recent reviews of light adaptation appear in Hood & Finkelstein, 1986; Shapley & Enroth-Cugell, 1985; Walraven, Enroth-Cugell, Hood, MacLeod & Schnapf, 1989). Indicated on the horizontal axis of Fig. 15 are five pairs of luminances corresponding to five different patterns in a constant-difference series.

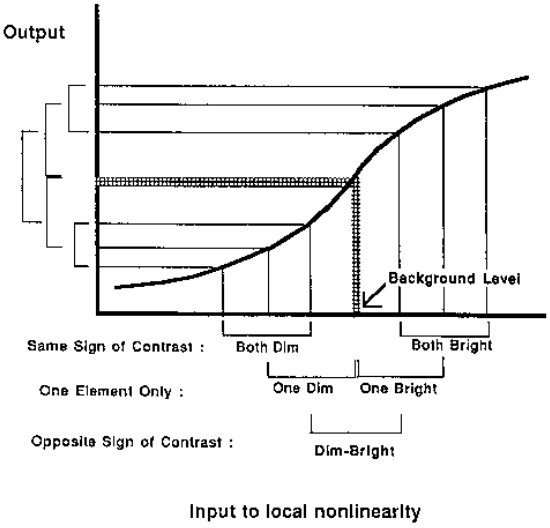

FIGURE 15. Diagram of early local nonlinearity acting between the luminance and the input to the simple or complex channels.

Vertical lines extend from these luminances up to the nonlinear curve, and then horizontal lines extend over to the vertical axis to show the outputs of the early local nonlinearity for each pair of luminances. Note that the difference between the two outputs for the same-sign-of-contrast patterns is smaller than that for the one-element-type-only patterns, which in turn is smaller than that for the opposite-sign-of-contrast patterns. As the luminances get further from the background in either direction, in short, this difference gets smaller.

To calculate approximate predictions from such an early local nonlinearity applied before multiple channels is quite easy. One uses equations (5) and (6) above, but substitutes the outputs of an early local pointwise nonlinearity [which will be called $r(\Delta L_1)$] for the incremental luminances in those equations. The resulting overall differences will be called D_{A*} and D_{B*}, from the fundamental and the high–low complex channels, respectively, to distinguish them from the corresponding quantities in equations (5) and (6):

$$D_{A*} = w_A \cdot |r(\Delta L_1) - r(\Delta L_2)| \tag{8}$$

$$D_{B*} = w_B \cdot \big||r(\Delta L_1)| - |r(\Delta L_2)|\big|. \tag{9}$$

Next combine the two overall differences as in equation (7) but now call the result D_{ELN}, where ELN stands for early local nonlinearity:

$$D_{ELN} = (D_{A*}^{k'} + D_{B*}^{k'})^{\frac{1}{k'}}. \tag{10}$$

A monotonic transformation F is then applied to D_{ELN}. Figure 16 illustrates this process. Its left panel shows some predictions from equation (10) plotted against contrast-ratio angle. Note that, as a result of the early local nonlinearity, the ends of the curves drop and different curves converge. As a result of the action of complex channels there are peaks in the curves at one-element-only patterns. The middle panel of Fig. 16 shows the monotonic transformation F applied to the D_{ELN} values in order to produce the displayed fit to the experimental results. An equation for the family of monotonic transformations we considered is given in the Appendix. It is a 4-parameter family, always having an S-shape. The right panel of Fig. 16 shows D_{ELN} transformed by the monotonic transformation F from the middle panel and plotted against contrast-ratio-angle in order to be compared to the experimental results.

The two top panels in Fig. 17 show predictions $F(D_{ELN})$ with parameters chosen to give a good fit to the experimental results using solid-square element patterns (left) and center-surround elements (right). The method of choosing the parameters is described in the Appendix. As visual comparison of these predictions with the experimental results indicates, these predictions fit the data quite well. More formally, for the parameter values used in the top of Fig. 17, the correlation r^2 between the values $F(D_{ELN})$ and the average observer ratings was 0.89, 0.93 and 0.98 for the solid-square stimuli in Experiment 1, the center-surround stimuli in Experiment 1, and the solid-square stimuli in Experiment 2, respectively. Allowing the parameter values to be different for the three different data sets, the highest r^2 values that appeared in our crude grid searches were 0.92, 0.93 and 0.99 for the three data sets, respectively.

The pooling exponent k' was 1.0 for the predictions in both panels of Fig. 17; the ratio of w_A (the weight on the channels at the fundamental) to w_B (the weight of the high–low complex channels) was 1 for the solid-square elements and 0.4 for the center-surround elements. Figure 18 illustrates the early local nonlinearity used for Fig. 17. These parameter values should not be overinterpreted, however, as there are strong interactions among parameters and a wide range of values for any one parameter can work well given the appropriate changes in values of other parameters. Some of these interactions are discussed in the Appendix.

It is interesting to consider just what visual process this early local nonlinearity might correspond to. That the experimental results plotted against contrast-ratio angle curves in Fig. 13 converge at the ends means that perceived segregation depends primarily on contrast ratio once both elements' luminances are far above or far below the mean luminance. This in turn means that the

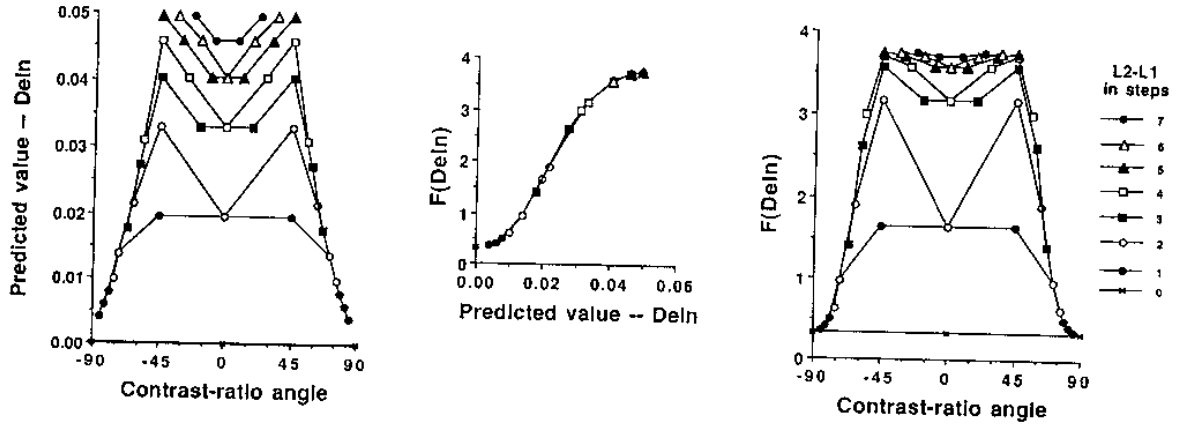

FIGURE 16. Some predictions from equation (10) for D_{ELN} are shown in the left panel plotted against contrast-ratio angle. The kind of monotonic transformations F between the predicted values and the observer's ratings that was allowed in fitting the models' predictions to observers' ratings is illustrated in the middle panel. The right panel replots the points in the left panel, this time with $F(D_{ELN})$ on the vertical axis. The right panel is shown again in the top left of Fig. 17.

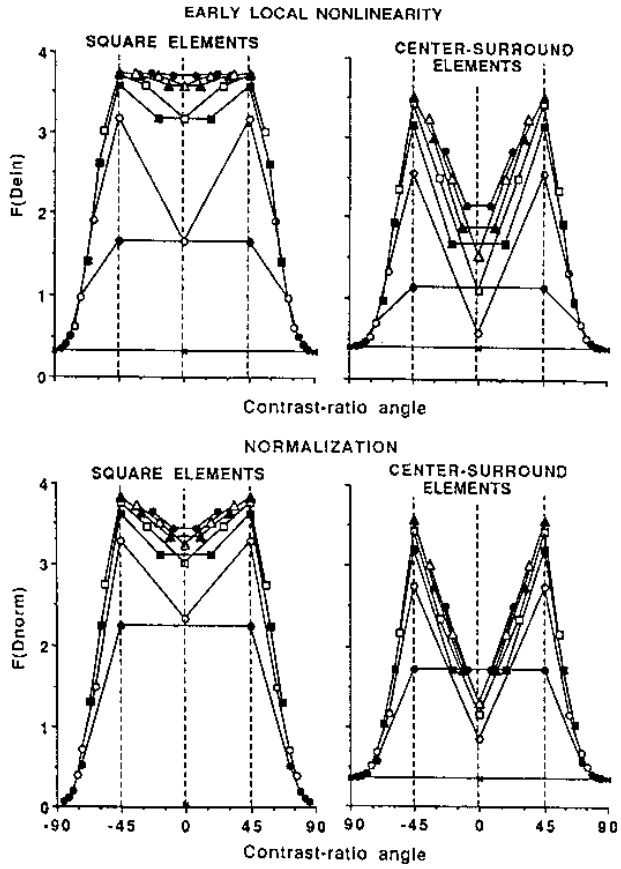

FIGURE 17. Predictions of models containing complex channels and either an early local nonlinearity (top half) or normalization among channels (bottom half). A monotonic transformation of the predicted overall difference between regions is plotted on the vertical axis. The parameter values for the predictions shown in the left (respectively right) panels were chosen to produce a good fit to the observers' ratings for solid-square (respectively center-surround) elements. The upper left panel repeats the right panel of Fig. 16. See text and Section 1 of the Appendix for further details.

early local nonlinearity needs to be approximately a logarithmic function of ΔL for values far from the background luminance. The heavy solid curve in Fig. 18 shows the early local nonlinearity used for the predictions in Fig. 17 top, and the dotted curves show alternatives that were clearly worse. Notice that the heavy solid curve is already clearly compressing at 18 or 14 ft-L, which—on a background of 16 ft-L—is a contrast of 13%. The light adaptation processes in the retina do not seem to show this much compression (e.g. in the model of Sperling & Sondhi, 1968; also see reviews of light adaptation referenced above), but some M-cells in the lateral geniculate nucleus might (e.g. Derrington & Lennie, 1984; Sclar, Maunsell & Lennie, 1990; Shapley & Perry, 1986; Spekreijse, van Norren & van den Berg, 1971). Thus, LGN M-cells are the most likely substrate currently for the early local nonlinearity discussed here, if such a nonlinearity is, in fact, the correct explanation of these results.

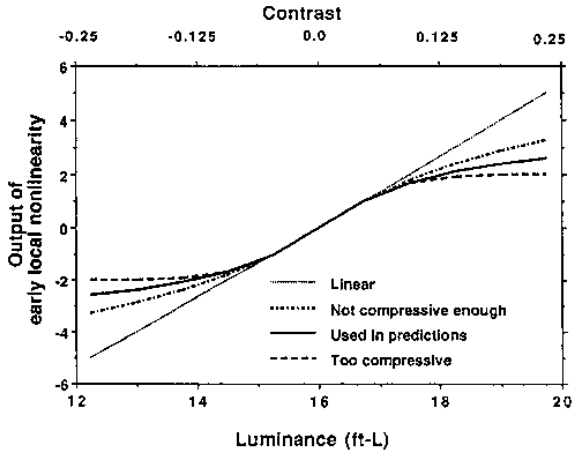

FIGURE 18. Several possible early-local transformation between the luminance at each spatial position and the input at each spatial position to the channels.

However, the next section examines a nonlinearity that occurs in V1 itself and might also explain the convergence of the curves for the same-sign-of-contrast patterns.

Normalization across channels (intracortical inhibition)—another possibility

The relationship between stimulus contrast and cortical cells' responses is known to be very compressive,

a good deal more compressive than at earlier stages in the visual system. Some cortical cells show compression at 10 or 20% contrast (e.g. Albrecht & Hamilton, 1982; Ohzawa, Sclar & Freeman, 1982; Sclar, Lennie & DePriest, 1989; Sclar *et al.*, 1990). Further, neurophysiological recordings from cortical cells produce results that are often described as cross-orientation or cross-frequency inhibition (e.g. Bonds, 1989; DeValois & Tootell, 1983; Morrone, Burr & Maffei, 1982). As has been recently pointed out (Robson, 1988a,b; Heeger & Adelson, 1989; Heeger, 1991), both the intracortical inhibition and the response compression may result from the same process, a *normalization* process which keeps the total response from some set of neurons at or below a ceiling. It accomplishes this by doing something like dividing (normalizing) the response of each individual neuron by the total response from a set of neurons (once thresholds have been exceeded). Heeger (1991) shows a possible wiring diagram for such a process and compares its predictions to data from cat striate cortex cells. Normalization-type processes have also been used in models of human texture segregation or other visual psychophysical tasks (e.g. Bergen & Landy, 1991; Grossberg, 1987; Grossberg & Mingolla, 1985; Heeger, 1987; Lubin, 1989; Lubin & Nachmias, 1990). A different but probably closely-related model of intracortical inhibition has been used to explain some other texture-segregation results (Malik & Perona, 1990).

For normalization to predict our results, we need to assume that the set of cortical neurons over which normalization occurs contains neurons with a wide range of different peak spatial frequencies. In fact, we will assume that the normalization occurs over all neurons having receptive fields that overlap in spatial position with the neuron in question no matter what spatial frequency and orientation they are most sensitive to. [Robson's (1988a,b) ideas would limit the normalization set more, but that limit has little effect in the following argument.]

Notice (e.g. in Fig. 6) that the total amount of contrast at element edges (e.g. $|\Delta L_1| + |\Delta L_2|$) is greater for the same-sign-of-contrast patterns than for the other patterns in a constant-difference series. This is true in both the checkerboard and striped regions of the pattern. The greater contrast at element edges is reflected in a larger amplitude of higher-harmonics in the Fourier transform of the stimulus.

Consider the effect of between-pattern differences in higher-harmonic amplitude on a group of channels we will call *high-only channels*. High-only channels include (i) simple channels tuned to higher harmonics and/or (ii) complex channels having both filters tuned to higher harmonics. Since these channels respond only to the edges of the individual elements and therefore respond to the same degree in the checkerboard and striped regions, none of these high-only channels can contribute to perceived region segregation. Because the contrast at element edges is greater in the same-sign-of-contrast patterns, however, the high-only channels' responses in both the checked and the striped regions are *larger* to same-sign-of-contrast patterns than to the other patterns in a constant-difference series. These *larger* responses from the high-only channels enter into the denominator of the normalization process not only for neurons in the high-only channels but also for neurons in other channels. Their entry produces *smaller* post-normalization responses to the same-sign-of-contrast patterns than to the other patterns. These smaller post-normalization responses occur in both the checkerboard and striped regions and in all channels, including the channels that do contribute to segregation of texture regions, namely the fundamental and high-low channels. These smaller responses in both regions lead in turn to smaller within-channel differences between the regions and therefore to less predicted segregation (for the same sign-of-contrast patterns relative to the other patterns).

In short, the high-only channels respond very similarly to both the striped and the checkerboard texture regions, and, therefore, these high-only channels do not contribute to region segregation. These high-only channels do, however, enter into the normalization process and thereby decrease the effectiveness of the fundamental and high–low channels that do contribute to region segregation. And these high-only channels are activated more by same-sign-of-contrast patterns than by the others in a constant-difference series.

To write an equation representing the approximate predictions of the normalization model, we need an expression for the responses of the high-only channels in the checked (or striped) region. Since these high-only channels respond primarily at the edges of elements and in proportion to the luminance increments at the edges, their response is in approximate proportion to ΔL_1 or ΔL_2 depending on element type. Since the two element types occur in equal proportions in each region, the spatially-pooled response of a high-only channel in either region will depend on spatial pooling across responses to both element types in equal proportions (that is, across both ΔL_1 and ΔL_2) in equal proportions. Thus R_0, the spatially-pooled regional responses of the high-only channels (in either the checked or the striped region), ought to be approximately equal to:

$$R_0 = w_0 \cdot (|\Delta L_1|^k + |\Delta L_2|^k)^{\frac{1}{k}} \qquad (11)$$

where k is the exponent describing spatial pooling as in equation (1).

The curves in Fig. 19 show predictions from equation (11) for several values of k. To demonstrate that this is indeed a good approximation, the points in Fig. 19 show numerical calculations of the spatially-pooled responses from one high-only channel. As expected, the high-only channels' spatially-pooled response increases precipitously as the contrast-ratio angle gets smaller than $-45°$ or larger than $+45°$, that is, as one moves from one-element-only patterns out to same-sign-of-contrast patterns further and further from the background.

To finish the prediction of perceived texture segregation, one puts the spatially-pooled regional responses of these *high-only* channels into a denominator that normalizes the responses of all the channels; one does

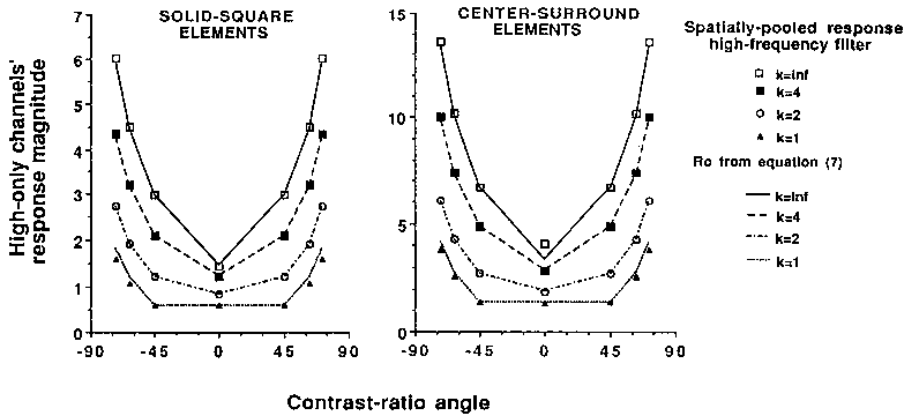

FIGURE 19. Spatially pooled responses of the high-only channels to a constant-difference series of patterns having solid-square elements (left panel) or center-surround elements (right panel). The curves show values calculated using equation (11). The symbols show values computed from full calculations. For these full calculations, the channel was a simple linear channel of vertical orientation with peak sensitivity at approx. 11 times the fundamental frequency of the texture regions. The spatial pooling was done over the checked region using equation (1). Very similar results were obtained for all other high-spatial-frequency channels and for the striped region. The symbols marked $k = \infty$ show the maxima of the absolute values of the outputs divided by 2 for convenience. The values of w_0 used in equation (11) were chosen separately for each curve.

not put it into the numerator (reflecting the fact that the high–low channels do not respond differentially in the two texture regions). As it turns out (interested readers can see the derivation in the appendix), the same expressions D_A and D_B that we used before go into both the numerator and denominator of the normalization equation to represent the action of the fundamental and high–low channels. For the high-only channels the spatially-pooled regional response R_0 goes into the denominator. Thus,

$$D_{NORM} = \frac{(D_A^{k'} + D_B^{k'})^{\frac{1}{k'}}}{\sigma + (D_A^{k'} + D_B^{k'} + R_0^k)^{\frac{1}{k'}}} \qquad (12)$$

where

$$D_A = w_A \cdot |\Delta L_1 - \Delta L_2|$$

$$D_B = w_B \cdot ||\Delta L_1| - |\Delta L_2||$$

$$R_0 = w_0 \cdot (|\Delta L_1|^k + |\Delta L_2|^k)^{\frac{1}{k}}$$

and σ is a constant which, among other things, serves to keep the expression from going to infinity as the contrast in the pattern is reduced to zero. Note that when the contrast of both elements is low enough, this equation becomes equivalent to equation (7) above except that the weights have been divided by σ. The maximal value of this equation is 1.0.

The bottom panels of Fig. 17 shows some monotonically-transformed predictions from equation (12) where the parameters were chosen to predict solid-square element results (left panel) and center-surround element results (right panel). These predictions again fit the data very well, as well or better than those in Fig. 17 top from the early local nonlinearity. More formally, the correlation r^2 between the monotonically-transformed predicted values and the average ratings was 0.97, 0.96 and 0.97 for the solid-square stimuli in Experiment 1, the center-surround stimuli in Experiment 1, and the solid-square stimuli in Experiment 2, respectively. Allowing the parameter values to be different for the three different data sets, the highest r^2 values that appeared in our crude grid searches were 0.97, 0.96 and 0.99 for the three data sets.

The values used in both panels of Fig. 17 for both pooling exponents k' and k were 1.0. The ratio of w_0 (the high-only channels' weight) to σ (the constant in the denominator of the normalization equation) was set at 1.0. The ratio of w_A (the weight on the channels at the fundamental) to w_B (the weight of the high–low complex channels) was 3 for the solid-square elements and 1 for the center-surround elements. Again, however, these parameter values should not be overinterpreted as there are strong interactions among parameters and a wide range of values of any one parameter can work well given the appropriate changes in values of other parameters. More information about the fits and the fitting procedure is given in the Appendix.

Thus, either an early, local compressive nonlinearity occurring before the level of the spatial-frequency channels themselves or a normalization network at the level of the channels could account satisfactorily for the results seen in these experiments. The fit of the normalization predictions was slightly better in general, but it would be unwise to take this superiority very seriously at this point. The fact that we have not done a very refined grid search and that we have used approximations rather than complete models makes small difference in fit unconvincing.

DISCUSSION

Although the effortless perceptual segregation of different regions in the visual field is well accounted for by spatial-frequency and orientation channels acting in a linear manner, there is clear nonlinear behavior as well. The nonlinear behavior exhibited in our experiments can

probably be explained by nonlinear processes known to exist at relatively low levels in the visual pathways: (1) a rectification nonlinearity or some other similar nonlinearity that occurs between two stages of linear filtering—this linear-nonlinear-linear sequence is quite like that used to describe complex cell behavior in the cortical area V1; (2) an intensity-dependent nonlinearity such that response as a function of contrast shows dramatic compression occurring at contrasts far less than 25%. Neither of these types by itself explains our results.

The intensity-dependent nonlinear behavior in our experimental results can be quantitatively predicted either by an early local nonlinearity occurring before the channels or by normalization among the channels (perhaps intracortical inhibition). As mentioned above, to decide between these processes on the basis of their predictions for the experimental results reported here is unreasonable. To decide between them on the basis of plausible physiological substrates is premature. However, the processes differ in other properties not tested by the experimental results reported here (e.g. their spatial characteristics), and further psychophysical experiments should be able to decide between them. Of course, both processes may operate.

Future experimental and theoretical work should also be able to clarify the relationship of the nonlinear processes suggested here to the nonlinear processes that have been suggested by other work on texture segregation, e.g. the normalized opponency in the model of Bergen and Landy (1991); the inhibition in the model of Malik and Perona (1990), the thresholding and suggestion of still further more complicated nonlinearities in Victor and Conte (1987, 1989a,b), the normalization across spatial position in Guernsey and Browse (1989) and the two stages of filtering with an embedded rectification in Sagi (1990) that bear at least superficial resemblance to our complex channels although may differ in important ways.

Notice that one might interpret the normalization equation (12) as a formulation of "masking" of the responses at the fundamental frequency by responses to the higher harmonics. It is not clear that this last sentence has much content but it does suggest searching for analogies between these psychophysical results and others commonly called "masking." Indeed, inhibition among spatial-frequency and orientation-selective channels has several times been proposed as the physiological substrate for masking and adaptation effects (see brief review in section 3.7 and 3.8 of Graham, 1989b).

Although superior segregation of opposite-sign-of-contrast relative to same-sign-of-contrast patterns has sometimes been taken as evidence for the existence of separate *on* and *off* feature maps or channels (e.g. Beck *et al.*, 1987), these results can be explained, as we did here, without such separate channels. (There are definitely *on* and *off* neurons but they need not be separated into different channels. Indeed, they might be used together to form one linear channel.) On the other hand, the existence of such separate *on* and *off* channels could easily be incorporated into the approach here by using a half-rectification either in the complex channel or in the spatial pooling stage. As you can probably imagine, a region containing only light-centered center-surround elements segregates rather well from one containing only dark-centered center-surround elements. (See the similar demonstration and accompanying discussion in Malik & Perona, 1990.) This demonstration does suggest that *on* and *off* neurons might be separated into different channels. It turns out to be quite difficult, however, to rule out the possibility that both kinds of neurons are in the same low-level channel if you allow higher stages of processing to include processes like looking for the maximum or the minimum. In any case, the experiments reported here certainly cannot decide whether or not there are separate *on* and *off* channels.

In summary, a great deal (perhaps all) of perceived region or texture segregation can be quantitatively accounted for by models involving only visual processes known to occur at realtively low levels in the visual system. Perhaps region segregation is a cheap and quick computation done early in visual processing to ease the overload on higher processes by delimiting regions within which computations can be sensibly concentrated. Of course, higher-level processes may turn out to play a substantial role in region segregation but such processes should not be invoked until they are needed (cf. Beck *et al.*, 1989; Bergen, 1992).

REFERENCES

Adelson, E. H. & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America, A2*, 284–299.

Albrecht, D. G. & Hamilton, D. B. (1982). Striate cortex of monkey and cat: Contrast response function. *Journal of Neurophysiology, 48*, 217–237.

Beck, J., Graham, N. & Sutter, A. (1991). Lightness differences and the perceived segregation of regions and populations. *Perception and Psychophysics, 49*, 257–269.

Beck, J., Prazdny, K. & Rosenfeld, A. (1983). A theory of textural segmentation. In Beck, J., Hope, B. & Rosenfeld, A. (Eds), *Human and machine vision* (pp. 1–38). New York: Academic Press.

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

Beck, J., Sutter, A. & Ivry, R. (1987). Spatial frequency channels and perceptual grouping in texture segregation. *Computer Vision, Graphics, and Image Processing, 37*, 299–325.

Bergen, J. R. (1991). Theories of visual texture perception. In Regan, D. (Ed.), *Vision and visual dysfunction, Vol. 10B: Spatial vision*. New York: Macmillan.

Bergen, J. R. & Adelson, E. H. (1986). Visual texture segmentation based on energy measures. *Journal of the Optical Society of America, A 3*, 98.

Bergen, J. R. & Adelson, E. H. (1988). Early vision and texture perception. *Nature, 333*, 363–364.

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

Bonds, A. B. (1989). Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex. *Visual Neuroscience, 2*, 41–55.

Bovik, A. C., Clark, M. & Geisler, W. S. (1987). Computational texture analysis using localized spatial filtering. *Proceedings of*

Workshop on Computer Vision, Miami Beach, Fla, 1987 (pp. 201–206). The IEEE Computer Society Press.

Bovik, A. C., Clark, M. & Geisler, W. S. (1990). Multichannel texture analysis using localized spatial filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 12,* 55–73.

Caelli, T. M. (1982). On discriminating visual textures and images. *Perception and Psychophysics, 31,* 149–159.

Caelli, T. M. (1985). Three processing characteristics of visual texture segregation. *Spatial Vision, 1,* 19–30.

Caelli, T. M. (1988). An adaptive computational model for texture segmentation. *IEEE Transactions on Systems, Man, and Cybernetics, 18,* 1.

Chubb, C. & Landy, M. S. (1991). Orthogonal distribution analysis: A new approach to the study of texture perception. In Landy, M. S. & Movshon, J. A. (Eds), *Computation models of visual processing.* Cambridge, Mass.: MIT Press.

Chubb, C. & Sperling, G. (1988). Processing stages in non-Fourier motion perception. *Investigative Ophthalmology and Visual Science (Suppl.), 29,* 266.

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.*

Coggins, J. M. & Jain, A. K. (1985). A spatial filtering approach to texture analysis. *Pattern Recognition Letters, 3,* 195–203.

Daugman, J. G. (1987). Image analysis and compact coding by oriented 2D Gabor primitives. *SPIE Proceedings, 758,* 19–30.

Daugman, J. G. (1988). Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 36,* 1169–1179.

Derrington, A. M. & Lennie, P. (1984). Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. *Journal of Physiology, 357,* 219–240.

DeValois, R. L. & DeValois, K. K. (1988). *Spatial vision.* New York: Oxford University Press.

DeValois, K. K. & Tootell, R. B. (1983). Spatial-frequency-specific inhibition in cat striate cortex cells. *Journal of Physiology, 336,* 339–376.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America, A4,* 2379–2394.

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

Graham, N. (1989a). Low-level visual processes and texture segregation. *Physica Scripta, 39,* 147–152.

Graham, N. (1989b). *Visual pattern analyzers.* New York: Oxford University Press.

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

Graham, N., Beck, J. & Sutter, A. K. (1989). Two nonlinearities in texture segregation. *Investigative Ophthalmology and Visual Science (Suppl.), 30,* 161.

Grossberg, S. (1987). Cortical dynamics of three-dimensional form, color, and brightness perception: Monocular theory. *Perception and Psychophysics, 41,* 87–116.

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

Guernsey, R. & Browse, R. A. (1989). Asymmetries in visual texture discrimination. *Spatial Vision, 4,* 1.

Heeger, D. J. (1987). Model for the extraction of image flow. *Journal of the Optical Society of America, 4,* 1455–1471.

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

Heeger, D. J. & Adelson, E. H. (1989). Mechanisms for extracting local orientation. *Investigative Ophthalmology and Visual Science (Suppl.), 30,* 110.

Hochstein, S. & Spitzer, H. (1985). One, few, infinity: Linear and nonlinear processing in the visual cortex. In Rose, D. & Dobson, V. G. (Eds), *Models of the visual cortex* (pp. 341–350). New York: Wiley.

Hood, D. C. & Finkelstein, M. A. (1986). Sensitivity to light. In Boff, K. R., Kaufman, L. & Thomas, J. P. (Eds), *Handbook of perception and peformance, Volume I. Sensory processes and perception* (Chap. 5). New York: Wiley.

Klein, S. A. & Tyler, C. W. (1986). Phase discrimination of compound gratings: Generalized autocorrelation analysis. *Journal of the Optical Society of America, A3,* 868–879.

Landy, M. S. & Bergen, J. R. (1988). Texture segregation by multiresolution energy or by structure gradient? Presented at the Annual Meeting of the Optical Society of America. *Technical Digest Series, 11,* 162.

Landy, M. S. & Bergen, J. R. (1989). Texture segregation for filtered noise patterns. *Investigative Ophthalmology and Visual Science (Suppl.), 30,* 160.

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

Lubin, J. (1989). Discrimination contours in an opponent motion stimulus space. *Investigative Ophthalmology and Visual Science (Suppl.), 30,* 426.

Lubin, J. & Nachmias, J. (1990). Discrimination contours in an F/3F stimulus space. *Investigative Ophthalmology and Visual Science (Suppl.), 31,* 409.

Malik, J. & Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America, A7,* 923–932.

Morrone, M. C., Burr, D. C. & Maffei, L. (1982). Functional implications of cross-orientation inhibition of cortical visual cells. I. Neurophysiological evidence. *Proceedings of the Royal Society, B, 216,* 335–354.

Nothdurft, H. C. (1985a). Orientation sensitivity and texture segmentation in patterns with different line orientation. *Vision Research, 25,* 551–560.

Nothdurft, H. C. (1985b). Sensitivity for structure gradient in texture discrimination tasks. *Vision Research, 25,* 1957–1968.

Ohzawa, I., Sclar, G. & Freeman, R. D. (1982). Contrast gain control in the cat visual cortex. *Nature, 298,* 266–268.

Peli, E. (1987). Seeing the forest for the trees: The role of nonlinearity. *Investigative Ophthalmology and Visual Science (Suppl.), 28,* 365.

Press, W. H., Flannery, B. T., Teukolsky, S. A. & Vetterling, W. T. (1986). *Numerical recipes.* Cambridge: Cambridge University Press.

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

Robson, J. G. (1988a). Linear and non-linear operations in the visual system. *Investigative Ophthalmology and Visual Science (Suppl.), 29,* 117.

Robson, J. G. (1988b). Linear and non-linear behavior of neurones in the visual cortex of the cat. Presented at *New Insights on Visual Cortex*, the Sixteenth Symposium of the Center for Visual Science, University of Rochester, Rochester, New York, June, 1988. Abstract p. 5.

Rubenstein, B. S. & Sagi, D. (1990). Spatial variability as a limiting factor in texture discrimination tasks: Implications for performance asymmetries. *Journal of the Optical Society of America A, 7,* 1632–1643.

Sagi, D. (1990). Detection of an orientation singularity in Gabor textures: Effects of signal density and spatial frequency. *Vision Research, 30,* 1377–1388.

Sclar, G., Lennie, P. & DePriest, D. D. (1989). Contrast adaptation in striate cortex of macaque. *Vision Research, 29,* 747–755.

Sclar, G., Maunsell, J. H. R. & Lennie, P. (1990). Coding of image contrast in central visual pathways of the macaque monkey. *Vision Research, 30,* 1–10.

Shapley, R. & Enroth-Cugell, C. (1985). Visual adaptation and retinal gain controls. In Osborne, N. N. & Chadler, G. F. (Eds), *Progress in retinal research* (Vol. 3, pp. 263–343). Oxford: Pergamon Press.

Shapley, R. & Perry, V. H. (1986). Cat and monkey retinal ganglion cells and their visual functional roles. *Trends in Neuroscience, 9,* 229–235.

Spekreijse, H., van Norren, D. & van den Berg, T. J. T. P. (1971). Flicker responses in monkey lateral geniculate nucleus and human perception of flicker. *Proceedings of the National Academy of Science, U.S.A., 68*, 2802–2805.

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

Sperling, G. & Chubb, C. (1989). Apparent motion derived from spatial texture. *Investigative Ophthalmology and Visual Science (Suppl.), 30*, 161.

Sperling, G. & Sondhi, M. M. (1968). Model for visual luminance discrimination and flicker detection. *Journal of the Optical Society of America, 58*, 1133–1145.

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., Sperling, G. & Chubb, C. (1991). Further measurements of the spatial frequency selectivity of second-order texture mechanisms. *Investigative Ophthalmology and Visual Science (Suppl.), 32*, 1039.

Turner, M. R. (1986). Texture discrimination by Gabor functions. *Biological Cybernetics, 55*, 71–82.

Victor, J. D. (1988). Models for preattentive texture discrimination: Fourier analysis and local feature processing in a unified framework. *Spatial Vision, 3*, 263–280.

Victor, J. D. & Conte, M. (1987). Local and long-range interactions in pattern processing. *Investigative Ophthalmology and Visual Science (Suppl.), 28*, 362.

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

Victor, J. D. & Conte, M. (1989b). What kinds of high-order correlation structure are readily visible? *Investigative Ophthalmology and Visual Science (Suppl.), 30*, 254.

Walraven, J., Enroth-Cugell, C., Hood, D. C., MacLeod, D. I. A. & Schnapf, J. (1989). Control of visual sensitivity. In Spillman, L. & Werner, J. (Eds), *Physiological foundations of perception*. Berlin: Springer.

Watson, A. B. (1983). Detection and recognition of simple spatial forms. In Braddick, O. J. & Sleigh, A. C. (Eds), *Physiological and biological preprocessing of images* (pp. 110–114). New York: Springer.

Acknowledgements—The research was supported by Air Force Office of Scientific Research Grants AFOSR 85-039 and 88-0323 and National Eye Institute Grant 1 RO1 EYO8459-01. Preliminary accounts of some of this work appear in Graham, Beck and Sutter (1989) and Graham (1991).

APPENDIX

Some Details of the Models and Fits

Section 1. Details of fitting models to data

Although counting free parameters is at best an approximate description of the constraints in the data-fitting process, it often gives some feel for those constraints. The prediction embodied in equation (10) for the early-local nonlinearity has the same 3 free parameters (k', w_A and w_B) as in the earlier prediction of equation (7) plus the several free parameters that go into the description of the early-local transformation $r(\Delta L)$. For the latter, we did not specify a parametrized form for $r(\Delta L)$, but rather simply choose a few examples, those shown in Fig. 18 plus two other more extreme ones. (The nonlinearity was assumed to be odd-symmetric around the background luminance.) Thus we certainly might have failed to find the best-fitting early-local nonlinearity. This procedure for choosing the early-local nonlinearity could be described as having at maximum 10 free parameters, the 10 values of $r(\Delta L)$ for different ΔL, or at minimum 1 parameter (the parameter actually used in the program), namely a parameter indexing which of the several possible early-local nonlinearities shown in Fig. 18 was chosen. Something intermediate is probably the best description.

Since the observer's response is assumed to be a monotonic function F of this predicted overall difference, there are also as many free parameters as were used in specifying this monotonic function. (In the fits shown here we used four such parameters as described below.)

In fitting the normalization model's prediction in equation (12) to the data, we did a crude search over the following parameters: w_A, w_B, w_0, σ and k'. The prediction embodied in equation (12) has the same 3 free parameters (k', w_A and w_B) as in the earlier simple-channels model prediction of equation (7) plus the 3 parameters used in the normalization process (k, w_0 and σ) although, in fact, we only tried one value for the spatial pooling exponent ($k = 1$). Again, since the observer's response is assumed to be a monotonic function F of this predicted overall difference, there are also as many free parameters as were used in specifying this monotonic function (4 in the fits shown here as described below).

Both for practical reasons and to see if there were systematic differences between data sets, we did the fits for three sets of data separately: (i) the solid-squares from Experiment 1; (ii) the center-surround elements from Experiment 1; and (iii) all the 66 stimuli (which were all solid-square element patterns) from Experiment 2. There were systematic differences among the best fits to all three data sets (which differed from one another either in type of element or in the particular set of contrast-ratio angles and luminance differences chosen), but describing these differences at this point in detail does not seem worth the space as it will take more experiments before we understand the significance (if any other than sampling error) of these systematic differences. (A brief description of one of these is given in the next paragraph.) The predictions shown in Fig. 17 are an informal compromise among those that, in three crude grid searches, produced good fits to the three data sets as measured both by r^2 and the appearance of plots like those in Fig. 17. (The best predictions found for the results from Experiment 2 alone are shown in Graham, 1991.)

For solid-square element patterns, the predictions were not very sensitive to the value of the exponential for pooling across channels (k') for either model, and, to the extent they are sensitive, $k' = 1$ (the value used in the fits shown here) was actually least good. For center-surround element patterns, however, the fits were quite sensitive to the value of k', and a value of $k' = 1$ was much better than higher values.

There are strong interactions between the parameters involved in making any of the fits. Trivially, making all the weights w larger by some factor or making the outputs of the early-local nonlinearity larger by some factor will be exactly compensated by making c [in the monotonic transformation of equation (A1) described below] smaller by that factor. Of particular note for the normalization model, it is generally only the ratio of w_0 to σ that matters rather than the values of either one.

The family of monotonic functions and the fitting procedure. The observer's rating of perceived segregation is assumed to equal $F(D)$ where F is a monotonic function otherwise unspecified and D is either D_{ELN} [equation (7)] or D_{NORM} [equation (12)]. Ideally for our purposes this value would have been computed from the best-fitting function F of all possible monotonic functions. For practical purposes, however, we had to limit ourselves to monotonic functions that can be described in simple equations.

In the fits shown here, we assumed that F was a member of a particular 4-parameter family of functions. (This family is a slight generalization of the Weibull distribution function, the Quick psychometric function, and the asymptotic regression function. It was picked merely because it contains S-shaped functions that can be varied in slope, horizontal position, minimal value at, and maximum.) The particular algebraic form we used was:

$$F(x) = a \cdot [b - 2^{-(c \cdot x)^d}] \quad \text{for } x \geq 0. \tag{A1}$$

The Nelder–Meade algorithm (see, e.g. Press, Flannery, Teukolsky & Vetterling, 1986) as instantiated in MATLAB (available from The Math Works Inc., 21 Eliot Street, South Natrick, MA 01760, U.S.A.) was used to find the 4 parameters for F producing the smallest mean-square error (over all the stimuli in a given set) between $F(D)$ and the average ratings of the observers.

Note that using the average ratings of the observers as we did is not ideal. The variance of average ratings near 0 and 4 is less than for those

intermediate. It would presumably be better to transform the ratings first onto a scale where the variance is close to constant before using a mean-square error criterion. Then ratings near 4 and 0 in the experimental results would matter more in the fit of the model. In the case shown in Fig. 16, for example, which attempts to fit the data in Fig. 13, the best-fitting monotonic transformation would not compress the predictions at the higher end as much as the transformation shown in Fig. 16, does. At this point in the investigation, however, such niceties produce much more work than benefit.

It may be of some interest to point out that, in the fits shown in Fig. 17, the value of the exponent d in the monotonic transformation of equation (A1) was between 2.4 and 3.4, and the minimum value of the function F (which equals $a \cdot b$ in the above formula) varied from 0.024 to 0.35 while the maximum value (which equals $ab + a$ in the above formula) varied from about 3.4 to 4.4. The value of c directly depends on the values of the weights in the model. For the early-local-nonlinearity (respectively normalization) model, when w_A was equal to 1 for the solid-square case, then c varied from 0.01 to 0.03 (respectively 0.13–0.23).

Section 2. Deriving the normalization prediction in equation (12)

First, we consider some properties of the unnormalized (open-loop, before-normalization) responses of the three types of channels under consideration (fundamental, high–low, and high-only). Second, we present the equations giving the with-normalization (closed-loop, after-normalization) responses of those channels in terms of the unnormalized responses. Thirdly, we simplify the with-normalization equations using the properties of the unnormalized responses. We will continue to use the symbol R for the unnormalized (open-loop, before-normalization) regional responses and the symbol D for the within-channel differences between unnormalized regional responses; we will use R^{**} and D^{**} for the corresponding quantities with normalization (closed loop, after-normalization).

Some properties of the unnormalized responses. In the following derivation we consider, of all the orientations of fundamental channels (indicated by subscript A) and high–low channels (indicated by subscript B), only those that are vertically-oriented. These channels contribute to texture segregation because they respond very well in the striped region of the pattern and not at all in the checkerboard region. To a very good first approximation, the vertically-oriented A and B type channels have the following spatially pooled responses:

$$R_A(\text{ch}) = 0 \quad \text{and} \quad R_A(\text{st}) \gg 0 \tag{A2}$$

$$R_B(\text{ch}) = 0 \quad \text{and} \quad R_B(\text{st}) \gg 0. \tag{A3}$$

Thus the within-channel difference equals the regional response in the striped region, that is:

$$D_A = R_A(\text{st}) \quad \text{and} \quad D_B = R_B(\text{st}). \tag{A4}$$

It is easy to verify that the derivation below for these vertically-oriented channels holds also for the obliquely-oriented channels except that the spatially pooled responses in the striped region will be zero while those in the checkerboard region are considerable. The horizontally-oriented fundamental and high-low channels do not contribute substantially to texture segregation and thus can be ignored here or else considered as grouped with the high-only channels. (We are also ignoring many other varieties of intermediate channels. Doing so is probably justified because these other varieties either act like one of the varieties we are considering or should contribute to segregation much less than they do.)

For the high-only channels (of any orientation), the regional responses are the same for the two regions:

$$R_0(\text{st}) = R_0(\text{ch}) = R_0. \tag{A5}$$

The equations for the with-normalization responses. We will compute the result of the normalization process by an equation in which the numerator is the unnormalized response of the channel under consideration and the denominator is the total of the unnormalized responses from all the channels over which the normalization occurs (plus a parameter σ). Heeger (1991) presents this equation as that for a single cell in the context of a feedback wiring diagram of cortical cells. We could start by presenting it here for single neurons in the appropriate channel and then explicitly doing the spatial pooling over the region in question, but the necessary terminology seems to obscure more than it reveals. Thus, for brevity's sake, we skip the single neuron stage. Here we also generalize from Heeger's assumption of a linear pooling across channels (an exponent of 1.0) to a family of summations characterized by different exponents k' for pooling across channels. (Note, however, that Heeger does explicitly contrast rectification to squaring in the complex cell and opts for squaring. This squaring in the middle of the complex channels followed by an exponent of 1 at the pooling-across-channels stage would seem to have many of the properties of rectification in the middle of the complex channels followed by an exponent of 2 at the pooling-across-channels stage.) Then the normalization equations become:

$$R_A^{**}(\text{st}) = \frac{R_A(\text{st})}{\sigma + (R_A(\text{st})^{k'} + R_B(\text{st})^{k'} + R_0(\text{st})^{k'})^{\frac{1}{k'}}} \tag{A6}$$

$$R_B^{**}(\text{st}) = \frac{R_B(\text{st})}{\sigma + (R_A(\text{st})^{k'} + R_B(\text{st})^{k'} + R_0(\text{st})^{k'})^{\frac{1}{k'}}}. \tag{A7}$$

The two equations for the regional responses in the checkerboard region, $R_A^{**}(\text{ch})$ and $R_B^{**}(\text{ch})$, are exactly like those for the striped region except that ch is substituted for st.

The with-normalization within-channel differences are given by:

$$D_A^{**} = |R_A^{**}(\text{ch}) - R_A^{**}(\text{st})| \tag{A8}$$

$$D_B^{**} = |R_B^{**}(\text{ch}) - R_B^{**}\text{st})|. \tag{A9}$$

Finally, to compute the predicted with-normalization overall difference, D_{NORM}, one pools together these within-channel differences over all channels. Since the within-channel differences for channels other than those of the A or B type are being ignored (i.e. being assumed approximately equal to zero):

$$D_{\text{NORM}} = [(D_A^{**})^{k'} + (D_B^{**})^{k'}]^{\frac{1}{k'}}. \tag{A10}$$

The observer's rating is assumed to be a monotonic function of this overall difference D_{NORM}.

Simplifying the equations for the with-normalization responses. Now we can simplify equations for the normalized spatially pooled regional responses R^{**} by using equations (A2) and (A3) to show:

$$R_A^{**}(\text{ch}) = 0.0 \quad \text{and} \quad R_B^{**}(\text{ch}) = 0.0. \tag{A11}$$

Combining this equation (A11) with equations (A8) and (A9), gives:

$$(D_A^{**}) = |R_A^{**}(\text{st})| \quad \text{and} \quad (D_B^{**}) = |R_B^{**}(\text{st})|. \tag{A12}$$

Also by equation (A4) substituted into (A6):

$$R_A^{**}(\text{st}) = \frac{D_A}{\sigma + (D_A^{k'} + D_B^{k'} + R_0^{k'})^{\frac{1}{k'}}}. \tag{A13}$$

And by equation (A5) substituted into equation (A7), the expression for $R_B^{**}(\text{st})$ has the same denominator as the above but the numerator is D_B.

Substituting into equation (A10), the expressions for D_A^{**} and D_B^{**} obtained by combining equations (A12) and (A13) gives the following which is identical to equation (12) presented in the main text:

$$D_{\text{NORM}} = \frac{(D_A^{k'} + D_B^{k'})^{\frac{1}{k'}}}{\sigma + (D_A^{k'} + D_B^{k'} + R_0^{k'})^{\frac{1}{k'}}}. \tag{A14}$$