

TEXTURE SEGREGATION AND ORIENTATION GRADIENT

MICHAEL S. LANDY¹ and JAMES R. BERGEN²

¹Psychology Department, New York University, New York, NY 10003 and ²SRI/David Sarnoff Research Center, Princeton, NJ 08540, U.S.A.

(Received 2 October 1989; in revised form 6 June 1990)

Abstract—Rapid texture segregation is examined using filtered noise textures. The stimuli consist of a foreground region of filtered noise with one dominant texture orientation against a background region with a different dominant orientation. Shape discrimination of the foreground region is measured as a function of the difference in orientation between the two regions ($\Delta\theta$), the distance over which the dominant orientation rotates from the background to the foreground value (Δx), and the dominant spatial frequency of the textures (f). Performance declines with smaller $\Delta\theta$, larger Δx , and lower f . These effects are partially independent of viewing distance, which implies that it is the *relative* or *object* spatial frequency, not *retinal* spatial frequency, which determines performance in this task. We present a model consisting of channels tuned for orientation and spatial frequency which compute local *oriented energy*, followed by (texture) edge detection and a cross-correlator which performs the shape discrimination. Monte Carlo simulations of this model are in accord with the degradation in performance with increased Δx and decreased $\Delta\theta$.

Texture Texture gradient Spatial filtering

INTRODUCTION

An image region with one texture appears as a distinct perceptual entity within a background region with contrasting texture if the difference between the two textures is sufficiently great. The phenomenon of *texture segregation* will occur with differences between the foreground and background in mean local contrast, local spatial frequency, orientation and other local image properties (e.g. Beck, 1982; Beck, Sutter & Ivry, 1987; Bergen & Julesz, 1983a,b; Caelli, 1982; Julesz, 1981; Julesz & Bergen, 1983; Sutter, Beck & Graham, 1989; for a review see Bergen, 1991). Most of these studies utilized *micropattern textures*, in which a foreground region contains a number of copies of a given micropattern and a background region contains a contrasting micropattern (such as randomly oriented L's on a background of randomly oriented X's, Bergen & Julesz, 1983a).

Nothdurft (1985c; see also 1985a,b) studied texture segregation using textures consisting of a random collection of short line segments with identical orientation. The line segments in the foreground texture had a different orientation than those in the background. It was shown that performance in a shape discrimination task improved with increasing difference in orien-

tation of the lines in the two image regions. In addition, a *spacing effect* was observed, wherein the effectiveness of a given orientation difference dropped when the line segments were spaced further apart. Nothdurft took this as evidence for the computation of a *structure gradient*. In other words, for a given textural property (in this case, orientation θ), the determinant of segregation performance is not simply the difference in that property from foreground to background ($\Delta\theta$), but the spatial gradient of that property across the texture boundary ($\Delta\theta/\Delta x$).

There are some difficulties in interpreting Nothdurft's results. One problem is the issue of sampling. Given a foreground region of a particular shape and size on a larger background, when the spacing between foreground micropatterns is increased, the sampling density of each pattern is necessarily decreased (along with the dominant spatial frequency of the stimulus). Thus, it is unclear whether poor performance results from the decrease in structure gradient or from sparser sampling of the foreground shape by the micropatterns. The stimuli used by Nothdurft confound spatial sampling of the textures and the visible gradient of spatial structure between the micropatterns of differing orientation. Nothdurft did examine this issue by

varying the size of the objects and the viewing distance, but these manipulations change other characteristics of the display in addition to those of interest.

We investigate texture segregation and the structure gradient using filtered noise textures. The foreground and background again differ in orientation; in this case the difference is in the *dominant orientation* of the filtered noise, which is determined by the orientation of the filter that is applied. With this type of stimulus it is possible to vary the change in structure ($\Delta\theta$) independently from the distance over which that change occurs (Δx) by using a different filter kernel in different image locations (nonisoplanatic filtering) to generate the textured image. The filters share the same dominant spatial frequency, which makes it possible to vary Δx without varying the spatial frequency content of the stimulus (unlike the micropattern textures used by Nothdurft).

In the stimuli studied by Nothdurft, the orientation of each micropattern (a small line segment) is well-defined and easily measured. For noise textures, as well as most naturally occurring textures, *local orientation* is not so easily defined, nor may it be specified with complete precision. However, in order to test the structure gradient hypothesis using these stimuli, we must define an appropriate local orientation measure. One way to do this is based on the outputs of filters with different preferred orientations applied to a given image region (Freeman & Adelson, 1989; Knutsson & Granlund, 1983). In order to analyze a texture for local orientation and scale, one can begin with a set of channels tuned for various orientations and spatial frequencies in much the same way as has been done in modeling threshold spatial vision (Watson, 1983; Wilson & Bergen, 1979). Multiple channels models of texture segregation (Bergen & Adelson, 1988; Bovik, Clark & Geisler, 1990; Caelli, 1985; Clark & Bovik, 1989; Fogel & Sagi, 1989; Malik & Perona, 1990) may be used to analyze the artificial micropattern textures which have been popular in empirical texture segregation research. However, these stimuli are broadband, and hence stimulate channels with different preferred spatial frequencies as well as multiple orientation channels. One added benefit of the filtered noise stimuli used in this paper is that these displays are narrowband in spatial frequency. In analyzing a texture segregation model we can concentrate on the interactions between different

orientations, since image information is restricted to a single scale.

In this paper, we are interested in texture segregation performance, which we define as the ability to make different image regions distinct based solely upon a difference in a textural property. As such, texture segregation succeeds only when differently textured regions are perceived as having a well-defined shape. Hence, a shape discrimination task is used to gauge performance. By varying $\Delta\theta$ and Δx independently, we are able to examine the structure gradient effect without the covariation of sampling of the displays found in the work of Nothdurft (1985c). We vary the dominant spatial frequency of the noise textures, as an analogy to the change in dominant spatial frequency which accompanies the change in raster width in Nothdurft (1985c). We also vary viewing distance, which allows us to determine whether the effects of varying spatial frequency are best described in retinal terms (cycles per deg) or in object terms (cycles per object). Finally, we show that human performance can be successfully modeled using a texture analyzer which consists of a set of parallel channels tuned for different orientations, followed by an edge detector which extracts the edges between differently textured regions.

METHODS

Subjects

There were two subjects in these experiments. One was an author, and the other was a student naive to the purposes of the experiments. All had normal or corrected-to-normal vision.

Displays

The displays consisted of areas of oriented bandpass filtered noise. A region with one dominant orientation was set against a background of a different orientation. The shape of the foreground region was a square with a diagonally truncated corner, and the subject's task was to identify which corner of the square was missing. (In these respects the task is very similar to that used by Bergen & Julesz, 1983b.) We investigated the importance of orientation gradient by varying the difference in orientation between the central foreground square and the background, and also by varying the distance over which that change in orientation occurred.

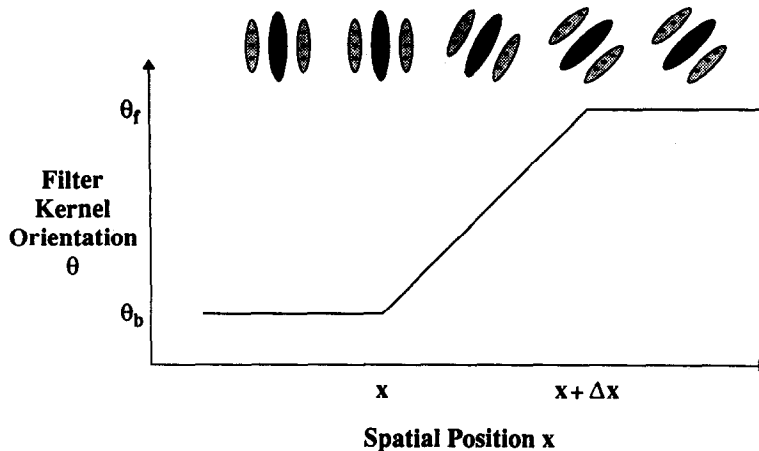


Fig. 1. A representation of the way in which the filtered noise textures were generated. A noise stimulus (Gaussian, IID) was filtered using oriented filters with Gabor impulse response. The preferred spatial frequency of the Gabor (θ) varied from the background (with value θ_b) to the foreground shape (with value θ_f). Across the boundary between the foreground and background, the dominant orientation rotated continuously from θ_b to θ_f across a distance Δx .

The oriented noise textures were generated by filtering Gaussian noise images (where each pixel is an independent random sample from a zero-mean normally distributed random vari-

able) using filters with a Gabor-shaped impulse response function:

$$w(x, y; \theta) = \sin[2\pi f(x \cos \theta + y \sin \theta)]e^{-(x^2 + y^2)/s^2}.$$

*It is important to be careful about the particular choice of filtering procedure used in order to avoid artificial cues at the edges of the figure other than the change in dominant orientation. There are three images used in this process: the Gaussian noise image $G(x, y)$ (the input), an orientation map $\theta(x, y)$ (used to guide the filtering process so as to produce the square figure on the background), and the texture image $T(x, y)$ (the result). The stimulus generation was accomplished in two steps: (a) compute an orientation map, and (b) use the orientation map to guide the filtering of the Gaussian noise image.

First, we associate an orientation $\theta(x, y)$ with each pixel position (x, y) of the Gaussian noise image. For $\Delta x = 0$, the pixels lying inside the foreground figure (the square with the missing corner) are associated with an orientation of θ_f , and others with θ_b . This image of dominant orientations is the orientation map. For larger Δx , the orientation map is simply a blurred version of the $\Delta x = 0$ orientation map. We use a blurring filter with an impulse response which is a square with sides of length Δx . This causes the orientation to change linearly between the background and foreground values.

The texture image may now be computed. The texture image is first cleared (to zeroes). Then, for each pixel in the Gaussian noise image, the value of that noise pixel is used to scale a Gabor with the dominant orientation associated with that pixel. This scaled Gabor is added into the texture image centered on that same pixel:

$$T(x, y) = \sum_{x', y'} G(x', y') W[x - x', y - y'; \theta(x', y')].$$

This procedure effectively blurs the dominant texture changes across the edge of the figure, avoiding artificial edge cues. Note that this is true even for $\Delta x = 0$.

The peak orientation of the filter was θ deg. For a viewing distance of 30 cm, three dominant spatial frequencies were investigated: 0.75, 1.5 and 3 c/deg. The spread of the Gaussian envelope s was set so that the diameter at half height of the weighting function was 1.4 cycles of the peak frequency f .

The displays consisted of a foreground with one dominant orientation and a background with a different dominant orientation. Thus, the images were created using *nonisoplanatic* (non-shift-invariant) filtering. Figure 1 illustrates the manner in which we are able independently to vary $\Delta\theta$ ($\theta_f - \theta_b$, the difference between dominant foreground orientation θ_f and the dominant background orientation θ_b) and Δx (the distance over which that change in dominant orientation occurs). A cross-section across the edge of the square figure is shown. As the edge of the figure is traversed, a series of filters are used at each spatial position which slowly rotate in dominant orientation from θ_b to θ_f . We used three background orientations θ_b (0° , 60° and 120°), two values of $\Delta\theta$ (18° and 36°), and three values of Δx (0, 1.25 and 2.5 deg, from a viewing distance of 30 cm).* Note that we will use the unit "deg" for spatial distance (as in "degrees of visual angle"), and the symbol "°" for dominant orientations of filters used to generate or analyze a texture (as in "degrees relative to the

vertical"). For each display condition (combination of f , $\Delta\theta$ and Δx) there were 12 possible stimulus displays (three values of θ_0 for each of four missing corners of the foreground square). Representative displays are shown in Fig. 2.

The displays as stored on disk were 512×512 pixels, with a 200×200 square figure centered in each display. The displays (and modeling described below) were computed using the HIPS image processing software (Landy, Cohen & Sperling, 1984a,b). A randomly chosen 400×400 portion of a given display was shown on each trial centered in a 480×512 raster display. The subject was not aware in advance of the exact position of the square within the display area, and therefore could not use directed eye movement or other strategies involving the attending to localized areas in the display in order to improve performance on the task. The stimulus geometry is shown (for a viewing distance of 30 cm) in Fig. 3. In the second experiment, a viewing distance of 60 cm was used as well with the same CRT displays. This resulted in reduced stimulus size (by approximately one half), and increased dominant spatial frequency (1.5, 3 and 6 c/deg). These frequencies correspond to 16, 8 and 4 pixels per cycle, respectively.

Stimuli were displayed using an Adage RDS-3000 display system on a US Pixel PX15 monochrome monitor with a fast P4-like phosphor. The video format was 60 Hz, noninterlace, with 480×512 visible pixels. The mean luminance was 53 cd/m^2 , and the peak Weber

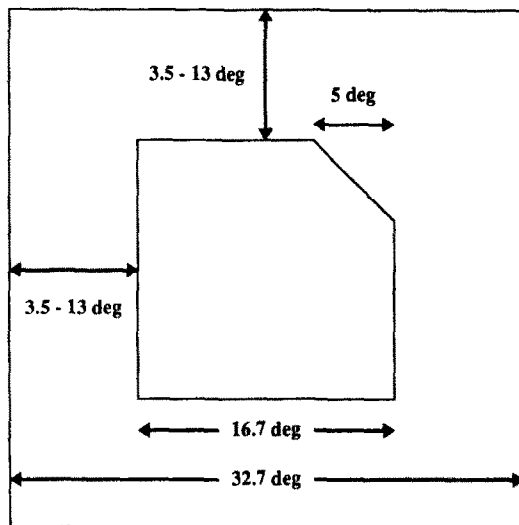


Fig. 3. Stimulus geometry for a viewing distance of 30 cm. For the 60 cm viewing distance all distances in retinal angle are halved (approximately).

contrast of the images was approx. 98%. The lookup tables of the display system were set so that the relationship between pixel value and display luminance was linear. The CRT display was kept at the same mean luminance between trials, and the area of the display outside of the visible 400×400 pixel portion of the texture display was also maintained at this same mean luminance. Subjects viewed the stimuli binocularly in a dimly lit room.

Procedure

The subject's task was to identify which corner of the foreground square was missing. The procedure was thus one-interval, four-alternative, forced choice, with a chance performance rate of 25%. Subjects initiated each trial with a response key, and keyed in their responses after the stimulus display. Feedback was provided after the subject's response which indicated which of the four corners was the correct answer.

Each trial (Fig. 4) consisted of a cue spot (which lasted 250 msec), a variable blank interval (250–500 msec), the stimulus texture (for two 60 Hz frames), and in some cases a variable blank interval and a post-masker (a 400×400 , 100% contrast, random-dot image, 50% white pixels and 50% black pixels). The variable blank interval before the stimulus image was intended to defeat subject strategies involving a well-timed blink or eye-movement prior to the post-masker. Subject performance (percent correct) was measured as a function of the SOA between the appearances of the stimulus texture and the post-masker, the viewing distance (30 or 60 cm), and the display condition (combination of f , $\Delta\theta$ and Δx).

The first experiment was performed entirely from a viewing distance of 30 cm. Each subject ran in 48 blocks of 144 trials. Within a block, $\Delta\theta$, Δx and f were fixed. The 48 blocks (4 replications of each combination of $\Delta\theta$, Δx and

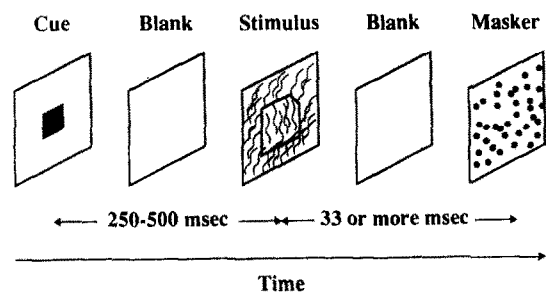


Fig. 4. A representation of the timing of the displays.



Fig. 2. Representative filtered noise stimuli. (A) $\theta_b = 0^\circ$, $\Delta\theta = 36^\circ$, $\Delta x = 0$ deg. The upper-right corner of the square figure is missing. (B) $\theta_b = 60^\circ$, $\Delta\theta = 36^\circ$, $\Delta x = 2.5$ deg at a viewing distance of 30 cm. The upper-right corner of the square figure is missing. (C) $\theta_b = 120^\circ$, $\Delta\theta = 18^\circ$, $\Delta x = 0$ deg. The lower-right corner of the square figure is missing. Note that segregation is poor with smaller $\Delta\theta$ and larger Δx .

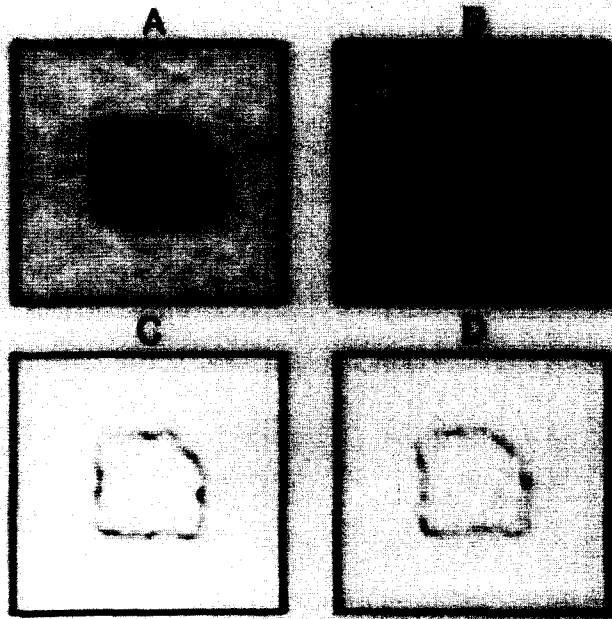


Fig. 8. Representative output from the texture model. (A) Normalized opponent energy of a filtered noise stimulus (vertical - horizontal). (B) Normalized opponent energy (right diagonal - left diagonal). (C) The final edge detected output in response to a stimulus with $\Delta\theta = 36^\circ$ and $\Delta x = 0$ deg. (D) The final edge detected output in response to a stimulus with $\Delta\theta = 36^\circ$ and $\Delta x = 3$ deg (from a viewing distance of 30 cm).

$f = 1.5$ or 3 c/deg) were run in random order. Within a block, 12 stimulus displays were used (4 corners of the square by 3 values of θ_b). For subject MSL, the 144 trials consisted of 2 replications of each display at each of 6 SOAs (33, 83, 133, 183 or 233 msec, or no post-masker). For subject EMB, there were 3 replications of each display at 4 SOAs (33, 83 or 133 msec, or no post-masker). Thus, for subject MSL there are 96 trials per data point (4 replications of each block, 2 replications of each display per block, 12 displays), and for subject EMB there are 144 trials per point (3 replications of each display per block).

A second experiment investigated the effect of viewing distance. $\Delta\theta$ was fixed at 36° , and Δx was fixed at 0 deg. Twelve stimulus blocks were run in random order (2 viewing distances, 3 values of f , and 2 replications). Each block consisted of 144 trials, with 4 replications of each of the 12 displays, and 3 SOAs (33 and 83 msec, or no post-masker). There were 96 trials per data point.

RESULTS

The results for the first experiment are given in Fig. 5. Each panel shows the data for one subject at one spatial frequency. The time course of these data is substantially different from what is seen in texture discrimination experiments using binary high contrast displays for which percent correct performance usually asymptotes near 100% as a function of SOA (e.g. Bergen & Julesz, 1983a,b). With filtered noise textures, the display does not appear to produce a useful long-lasting after-image, and any improvement with increasing SOA is complete by 133 msec or earlier. There appears to be little possibility of a serial search of an iconic image after the display has completed. The time course of easier conditions (e.g. $\Delta\theta = 36^\circ$, $\Delta x = 0$ deg), does not appear to be different from that for difficult conditions (e.g. $\Delta\theta = 36^\circ$, $\Delta x = 3$ deg). Thus, we may characterize the interesting variations in the data by the asymptotic performance value for each condition. This

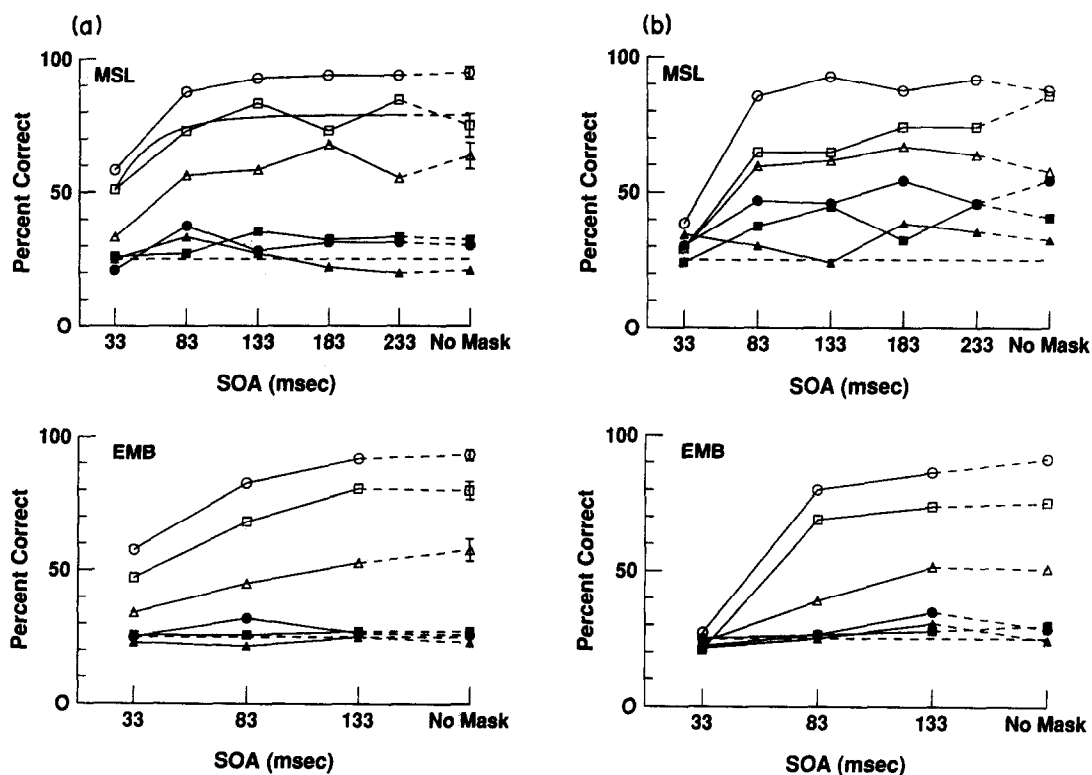


Fig. 5. Shape discrimination performance as a function of the SOA of a post-masker. Open symbols denote $\Delta\theta = 36^\circ$; solid symbols denote $\Delta\theta = 18^\circ$. $\Delta x = 0$ deg (circles), 1.5 deg (squares), or 3 deg (triangles). The viewing distance is 30 cm. Chance performance is given by the horizontal dashed line. Data for two subjects are shown. (a) $f = 1.5$ c/deg. (b) $f = 3$ c/deg. One representative fit is shown of an exponential function which is used to derive the asymptotic performance levels given in Fig. 10. Selected error bars indicate ± 1 standard error of the mean.

is done below (Fig. 10) by fitting an exponential function ($\alpha - \beta e^{-t/\gamma}$) to each data set and plotting the asymptotic value of the best fitting function (in the least squares sense).^{*} The best fit values of γ vary between 30 and 70 msec for the $\Delta\theta = 36^\circ$ conditions, and do not vary systematically. The γ values for the $\Delta\theta = 18^\circ$ conditions vary more, since the data are so flat and variable.

Nothdurft (1985c) suggested that the important variable for determining performance in a texture task requiring a discrimination of orientation is the spatial gradient of orientation from figure to background ($\Delta\theta/\Delta x$). If this theory is correct, then performance should improve with larger $\Delta\theta$ and smaller Δx . This is precisely the case with filtered noise textures. Thus, we have been able to confirm Nothdurft's basic observation using a stimulus for which there is no covariation of Δx with stimulus sampling (the grid of micropatterns used by Nothdurft). The asymptotic values reported here are comparable to the 30 msec data from Nothdurft, where no post-masker was used (1985c, Figs 5 and 7b). In Nothdurft's data, a substantial difference in texture segregation was found between $\Delta\theta$ values of 20 and 30 deg. Here, the drop in performance was a bit larger, and the $\Delta\theta$ values were 18 and 36 deg. Direct comparisons between "comparable" values of Δx are not possible given the different types of stimuli used. Nothdurft's hypothesis relating the "structure gradient" to texture segregation leads to the following quantitative prediction: conditions which have the same values of $\Delta\theta/\Delta x$ should lead to the same performance. Before this prediction can be tested, a model is needed for how the observer computes a local value of θ . We discuss this issue in the next section.

^{*}Confidence intervals for the estimated asymptotic performance level are difficult to define given that we fit a three parameter nonlinear model. We estimated such a confidence interval in two ways. Using the exact 95% confidence region described by Draper and Smith (1981, Chap. 10) based on the raw response data, and fixing the other two fit parameters at the values found by least squares, we determined the boundaries of the confidence region varying only the asymptotic performance parameter α . This yields an interval of $\pm 6\%$ correct or so for each curve. In addition, one can estimate this interval simply by averaging all of the data for a given curve once asymptote has been reached (the rightmost four data points in Fig. 5 for subject MSL, and the rightmost two data points for subject EMB) and use the binomial variance. This yields slightly smaller confidence intervals.

In this first experiment, there was only a small effect of dominant spatial frequency f (compare the two panels for each subject in Fig. 5). There is some hint that performance is better for 1.5 c/deg textures than 3 c/deg textures for short SOAs. The 33 msec SOA data for 1.5 c/deg are generally as good or better than the corresponding 3 c/deg condition.

The results of the second experiment are given in Fig. 6. Here, a wider span of spatial frequencies were used in order to better show any frequency dependent effects. It is clear that with a sufficiently low-frequency texture, performance suffers. Nothdurft (1985c) found that for a fixed difference in orientation between foreground and background elements (small oriented line segments), texture segregation was impaired by increasing the distance between adjacent elements (the raster width). That manipulation had several effects. It increased the distance across which structure is changed (our Δx), decreased the dominant spatial frequency, decreased the sampling rate of the texture, and decreased the total amount of visible structure.

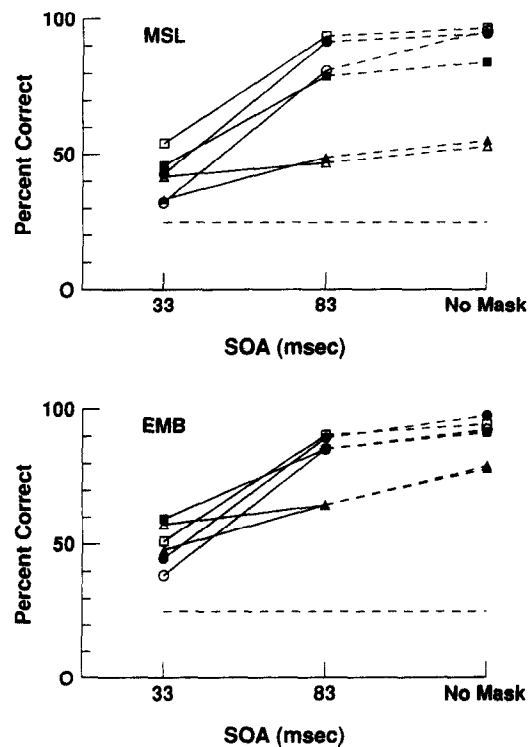


Fig. 6. Shape discrimination performance as a function of the SOA of a post-masker. $\Delta\theta = 36^\circ$. $\Delta x = 0$ deg. $f = 3$ c/deg (\bullet), 1.5 c/deg (\blacksquare), or 0.75 c/deg (\blacktriangle), viewed from a distance of 30 cm. The open symbols represent the same displays viewed from a distance of 60 cm (so that the retinal spatial frequencies were doubled). Data for two subjects are shown.

We have found that performance declines when either Δx is increased or dominant spatial frequency is lowered sufficiently, when these are varied independently of the other factors.

The entire experiment was run using the same displays from two viewing distances. The bulk of the variation in performance may be attributed to the stimulus that was displayed on the CRT, with little effect of viewing distance. In other words, the effects of spatial frequency that we have observed here are a function of *relative* or *object* spatial frequency (measured in cycles/object or cycles/cm), rather than of *retinal* spatial frequency (cycles/deg). This observation is consistent with other results involving supra-threshold spatial vision including studies of recognition of letters masked by noise (Parish & Sperling, 1987) and reading rate (Legge, Pelli, Rubin & Schleske, 1985). It is also consistent with the viewing distance manipulation used by Nothdurft (1985c) in his study of texture segregation with rasters of oriented line segments (and with Nothdurft, 1985b).

A MATCHED FILTER MODEL

From studies of threshold spatial pattern detection and discrimination, there is general agreement that visual processing includes a number of spatial channels tuned for a range of spatial frequencies and orientations, and models of spatial vision at threshold generally include a set of such filters (Watson, 1983; Wilson & Bergen, 1979). It has also been suggested that texture segregation results may be explained by models which initially analyze the retinal image

in a similar manner (Bergen & Adelson, 1988; Bovik et al., 1990; Caelli, 1985; Clark & Bovik, 1989; Fogel & Sagi, 1989; Malik & Perona, 1990), rather than resorting to specialized processing for texture (e.g. Beck, 1982; Julesz, 1981; Iwama & Maida, 1989).

We have simulated several variants of a model for texture segregation which compute local oriented energy (Adelson & Bergen, 1985). Here, we present a brief account of a particular model which can account for the data. A more extended account may be found in Bergen and Landy (1991). We have used the model to simulate to decline in performance with increased Δx and decreased $\Delta\theta$. We have not modeled the effects of varying spatial frequency or viewing distance. Also, because there is no temporal component to the model, we do not account for the complete time course as a function of SOA, but merely the asymptotic performance levels. The model is shown in Fig. 7.

The model first analyzes the retinal image using a bank of several spatial filters varying in spatial frequency and preferred orientation. As noted above, one of the benefits of using band-pass filtered noise stimuli is that such displays are only passed by spatial filters with a peak spatial frequency near the dominant spatial frequency of the stimulus. Thus, we do not need to consider the manner in which responses are pooled across spatial frequencies. The more traditional micropattern stimuli used in much of the texture segregation literature are broadband in spatial frequency content and more complicated to analyze using models such as

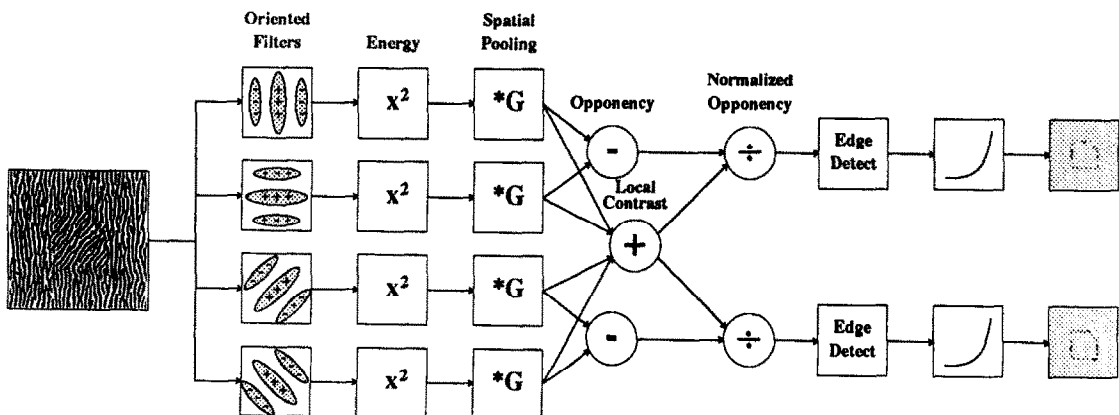


Fig. 7. A block diagram for a texture analysis model. The stimulus is analyzed by four oriented filters. The filter outputs are squared and pooled, resulting in local oriented energy maps. Orthogonal orientations are placed in opponency, and normalized by a local oriented contrast measure. The resulting normalized opponency images are edge enhanced, and then further processed by an accelerating nonlinearity which effectively discards weak texture edge responses.

this one. We include four filters, with preferred orientations spaced 45° apart, and with peak frequency centered on the dominant frequency of the stimuli. The orientation bandwidth is 40° , and the spatial frequency bandwidth (at half height) is 1.6 octaves. These produce four filtered images: vertical ("V"), horizontal ("H"), diagonal down-and-right ("R"), and diagonal down-and-left ("L").

The essence of a texture analysis (Bergen, 1991; Caelli, 1985) is to look for particular image qualities ("verticality," etc.) pooled across spatial phases within a small region of the image. This may be done by computing a pooled texture "energy" in much the same way that "motion energy" may be computed by squaring the output of a motion selective linear filter and summing across locations or phases (Adelson & Bergen, 1985). The four filtered images are each squared (point-wise) and averaged across a small spatial area. (Since averaging necessarily results in an image with less high-frequency content, we also subsample the image at this point.*) In order to produce responses that depend in a convenient way on local image orientation, two opponent signals are constructed by taking the difference of the images corresponding to orthogonal orientations ($H^2 - V^2$ and $R^2 - L^2$). Next, in order that the model measure a textural property independent of local image contrast, the opponent responses are normalized by the local oriented energy pooled over all orientations [e.g. $(H^2 - V^2) / (H^2 + V^2 + R^2 + L^2)$]. This procedure reduces the effects of local contrast variation, and thus yields a less noisy estimate of local structure. The resulting images for our textures consist of a noisy square region on a noisy background (Fig. 8A,B). These values are systematically related to local orientation. If texture segregation performance depends on the spatial gradient of local structure, then this corresponds in image processing terms to working with edge gradient information rather than with the original energy values. Consequently, for analysis of the shape of the square boundary, we apply the Sobel edge enhancer (see Gonzalez & Wintz, 1977) which computes the magnitude of the gradient at each location, followed by an expansive nonlinearity (a fourth power) in order to suppress weak edge responses. There is some

evidence that texture segregation proceeds from the extraction of texture edges, rather than grouping uniform texture regions (for example, the texture-based Craik-Cornsweet figure in Nothdurft, 1985c, Fig. 8; and the texture contrast-based Craik-Cornsweet figure in Sagi & Hochstein, 1985, Fig. 2).

The subject's task in the experiments reported above was to localize the square textured region and characterize its shape. We simulate this behavior by requiring the model to localize the figure in the resulting edge images (Fig. 8C,D) and determine which of the four corners is missing. The model's "performance" was determined by Monte Carlo simulation using a noisy cross-correlator to determine the model's "response" (Fig. 9). For any given condition (value of $\Delta\theta$ and Δx ; only the data for $f = 1.5$ c/deg at a viewing distance of 30 cm were analyzed), 12 displays were used in the experiments (four missing corners and three values of θ_b). For each such display, the texture analysis model results in two edge images (derived from $V^2 - H^2$ and $R^2 - L^2$). Noise was added to each of these images (Gaussian noise; the standard deviation of the noise was adjusted to lower simulated performance to the range of human performance). A cross-correlation was computed between each of these two noisy edge images and four idealized edge templates (squares with missing corners). The template that produced the maximum cross-correlation value determined the model's "response" on that "trial". For each condition, the model ran through 120 "trials" (10 replications with different additive noise fields for each of the 12 displays), and a percentage correct performance was computed.

The results of the Monte Carlo simulation are given in Fig. 10. Performance for the model and the asymptotic performance for the two subjects are shown as a function of Δx (the parameter is $\Delta\theta$). The asymptotic performance levels were derived from the exponential fits to the data described above (a sample fit is shown in Fig. 5). The model was able to provide a reasonable fit to the data.

Although the only parameter used to provide this fit was the amplitude of the additive noise, the model is quite elaborate. We believe that this complexity is justified because simpler versions of the model are unable to provide an adequate fit to the data (Bergen & Landy, 1991). If a noiseless cross-correlator is used, the model's performance is perfect. Also, it is not consistent

*For the blurring and subsampling two levels of a Gaussian pyramid "reduce" operation were used (Burt, 1981; Burt & Adelson, 1983).

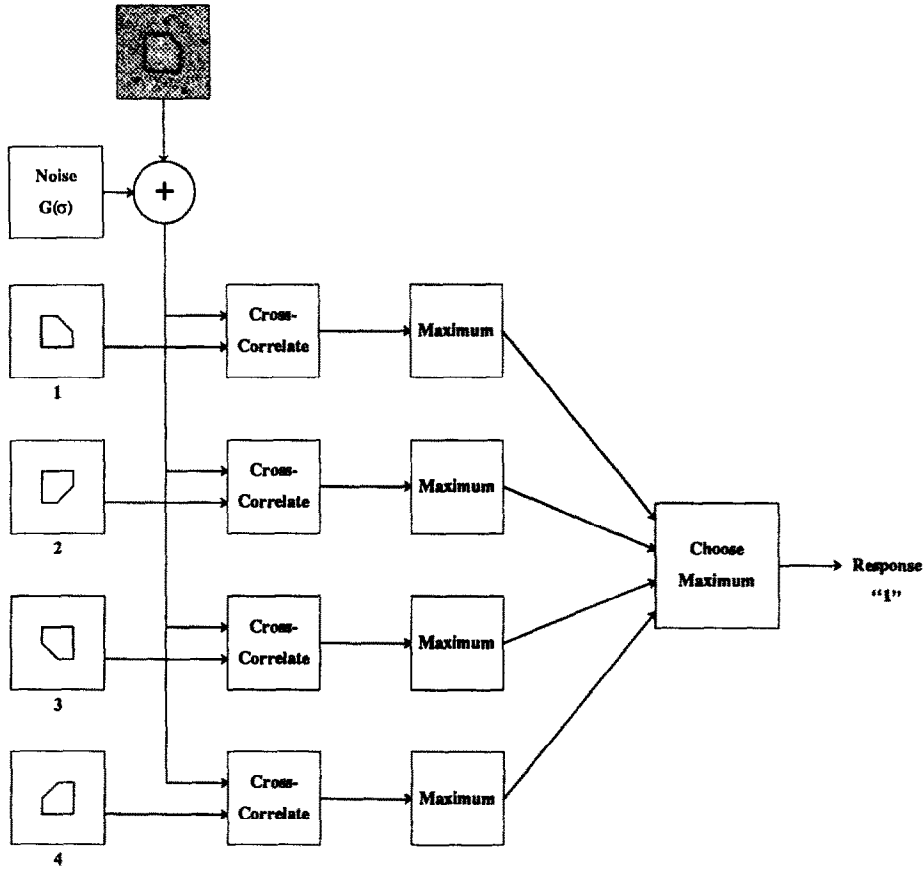


Fig. 9. A block diagram of the Monte Carlo procedure. An edge detected response of the model was corrupted by noise and then cross-correlated with idealized edge images of the patterns to be discriminated. The pattern yielding the largest maximal correlation value was chosen as the "response" on that "trial". In the simulations, this cross-correlation was performed between the templates and both the "vertical - horizontal" and "right diagonal - left diagonal" images, and the response was based on the largest correlation across all eight cross-correlations.

with the data to omit the nonlinearity applied to the edge responses, or to perform the cross-cor-

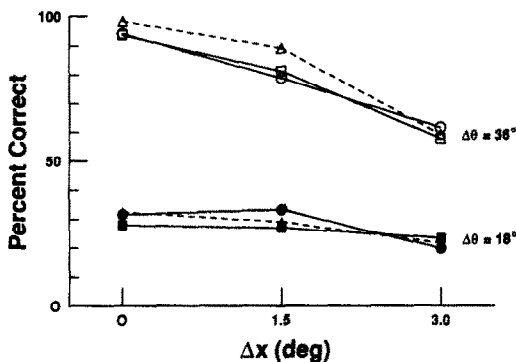


Fig. 10. Asymptotic performance in the shape discrimination task and performance predicted by the texture segregation model. The human performance data were determined by curve fits to the data for $f = 1.5$ c/deg from Fig. 5a (solid lines; circles for subject MSL, squares for subject EMB). The performance of the texture segregation model (dashed line and triangles) is the result of 120 "trials" per data point.

relation at any earlier stage of the model shown in Fig. 7. The qualitative reason for these failures is that at each of the earlier stages, for sufficient additive noise, increasing Δx has too little effect on the model's performance relative to that found in the human data, although performance is strongly degraded by decreased $\Delta\theta$. The effect of $\Delta\theta$ is more a function of the effective bandwidth of the filters; excessively narrowband filters would be required to segregate textures differing by only 18° , given that the noise stimulus itself has a fairly large orientation bandwidth. Displays with large values of Δx do not give the impression of having an edge between the foreground and background textures, even though these textures are easily discriminated. This accords with the notion that an edge response is computed between different textured regions, and that weak edge responses are discarded, as by the expansive nonlinearity in the model.

Nothdurft (1985c) suggested that texture segregation occurs when the "structure gradient" is sufficiently large. Our results and analysis are essentially in agreement with this idea. A measure of local orientation can be computed from the normalized opponent energy values computed in the model (Bergen & Landy, 1991; Freeman & Adelson, 1989; Knutsson & Granlund, 1983). Our computation of texture edge strength is certainly related to the structure gradient, although we do not explicitly compute local orientation (θ) in the model, nor do we compute the structure gradient ($\Delta\theta/\Delta x$).

We have described a model for texture segregation. The model has been used successfully to model a portion of our results: the degradation of human texture segregation performance with increased Δx and decreased $\Delta\theta$. We have not captured all of our results (variation with spatial frequency and viewing distance), although one would expect some amount of degradation with lower spatial frequencies simply because such a stimulus contains less information. But, we have shown that this approach to modeling can reproduce human behavior in a few circumstances. More empirical research would be required to support a particular block diagram for the model, particular filter characteristics, spatial pooling, and so on. Fortunately, from our own investigations of this model, it appears to be robust with respect to the type of filters employed. For example, similar results were produced using first-derivative filters with a much wider orientation bandwidth. Because of the orientation opponency, the precise details of the filters become less critical.

SUMMARY

Texture segregation performance was analyzed using filtered noise patterns. It was found that performance degrades with decreased orientation difference between the foreground and background textures ($\Delta\theta$) and with an increase in the distance over which that orientation difference occurs (Δx), which supports the results of Nothdurft (1985c). Performance is also poorer for low spatial frequencies, and this is an effect of object spatial frequency, not retinal spatial frequency. The results may be accounted for using a model with channels tuned to various scales and orientations, followed by an edge detection stage which responds to texture edges. This model is similar to current models of spatial pattern detection and

discrimination, and can account for human texture segregation performance. The model does not require more elaborate processing mechanisms such as the detection of textons, linking of edges, and region growing.

Acknowledgement—The work described in this paper was supported in part by a grant from the Office of Naval Research, Grant N00014-85-K-0077. We would like to thank Charlie Chubb and Anne Sutter for their helpful comments, and Robert Picardi for technical assistance. Portions of this work have been presented at the *Annual Meetings of the Optical Society of America*, 4 November 1988, and the *Association for Research on Vision and Ophthalmology*, Sarasota, Florida, 2 May 1989, and published in abstract form (Landy & Bergen, 1988 and 1989).

REFERENCES

- Adelson, E. H. & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, 2, 284–299.
- Beck, J. (1982). Textural segmentation. In Beck, J. (Ed.), *Organization and representation in perception* (pp. 285–317). Hillsdale, NJ: Erlbaum.
- Beck, J., Sutter, A. & Ivry, R. (1987). Spatial frequency channels and perceptual grouping in texture perception. *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). New York: Macmillan. In press.
- Bergen, J. R. & Adelson, E. H. (1988). Early vision and texture perception. *Nature, London*, 333, 363–364.
- Bergen, J. R. & Julesz, B. (1983a). Parallel versus serial processing in rapid pattern discrimination. *Nature, London*, 303, 696–698.
- Bergen, J. R. & Julesz, B. (1983b). Rapid discrimination of visual patterns. *IEEE Transactions on Systems, Man and Cybernetics, SMC-13*, 857–863.
- 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, MA: MIT Press. In press.
- Bovik, A. C., Clark, M. & Geisler, W. S. (1990). Multi-channel texture analysis using localized spatial filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12, 55–73.
- Burt, P. J. (1981). Fast filter transforms for image processing. *Computer Graphics and Image Processing*, 16, 20–51.
- Burt, P. J. & Adelson, E. H. (1983). The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications, COM-31*, 532–540.
- 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 segmentation. *Spatial Vision*, 1, 19–30.
- Clark, M. & Bovik, A. C. (1989). Experiments in segmenting texton patterns using localized spatial filters. *Pattern Recognition*, 6, 707–717.
- Draper, N. & Smith, H. (1981). *Applied regression analysis* (2nd edn). New York: Wiley.
- Fogel, I. & Sagi, D. (1989). Gabor filters as texture discriminator. *Biological Cybernetics*, 61, 103–113.

- Freeman, W. T. & Adelson, E. H. (1989). Steerable filters. *Topical meeting on image understanding and machine vision*. 1989 Technical Digest Series (Vol. 14). Washington, DC: Optical Society of America.
- Gonzalez, R. C. & Wintz, P. (1977). *Digital image processing*. Reading, MA: Addison-Wesley.
- Iwama, K. & Maida, A. S. (1989). Organizing and integrating edge segments for texture discrimination. *Journal of Experimental and Theoretical Artificial Intelligence*, 1, 4–30.
- Julesz, B. (1981). Textons, the elements of texture perception and their interactions. *Nature, London*, 290, 91–97.
- Julesz, B. & Bergen, J. R. (1983). Textons, the fundamental elements in preattentive vision and perception of textures. *Bell Systems Technical Journal*, 62(6), 1619–1645.
- Knutsson, H. & Granlund, G. H. (1983). Texture analysis using two-dimensional quadrature filters. *IEEE Computer Society workshop on computer architecture for pattern analysis and image database management—CAPAIDM* (pp. 206–213). Silver Spring, MD: IEEE Computer Society.
- Landy, M. S. & Bergen, J. R. (1988). Texture segregation by multiresolution energy or by structure gradient? *OSA Annual Meeting*. 1988 Technical Digest Series (Vol. 11, pp. 162–163). Washington, DC: Optical Society of America.
- Landy, M. S. & Bergen, J. R. (1989). Texture segregation for filtered noise patterns. *Investigative Ophthalmology and Visual Science (Suppl.)*, 30, 160.
- Landy, M. S., Cohen, Y. & Sperling, G. (1984a). HIPS: A Unix-based image processing system. *Computer Vision, Graphics and Image Processing*, 25, 331–347.
- Landy, M. S., Cohen, Y. & Sperling, G. (1984b). HIPS: Image processing under UNIX—Software and applications. *Behavior Research Methods, Instruments and Computers*, 16, 199–216.
- Legge, G. E., Pelli, D. G., Rubin, G. S. & Schleske, M. M. (1985). Psychophysics of reading—I. Normal vision. *Vision Research*, 25, 239–252.
- Malik, J. & Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A*, 7, 923–932.
- Nothdurft, H. C. (1985a). Texture discrimination: Representation of orientation and luminance differences in cells of the cat striate cortex. *Vision Research*, 25, 99–113.
- Nothdurft, H. C. (1985b). Orientation sensitivity and texture segmentation in patterns with different line orientation. *Vision Research*, 25, 551–560.
- Nothdurft, H. C. (1985c). Sensitivity for structure gradient in texture discrimination tasks. *Vision Research*, 25, 1957–1968.
- Parish, D. H. & Sperling, G. (1987). Object spatial frequency, not retinal spatial frequency, determines identification efficiency. *Investigative Ophthalmology and Visual Science (Suppl.)*, 28, 359.
- Sagi, D. & Hochstein, S. (1985). Lateral inhibition between spatially adjacent spatial-frequency channels? *Perception and Psychophysics*, 37, 315–322.
- 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.
- Watson, A. B. (1983). Detection and recognition of simple spatial forms. In Braddick, O. J. & Sleigh, A. C. (Eds.), *Physical and biological processing of images* (pp. 73–87). New York: Springer.
- Wilson, H. R. & Bergen, J. R. (1979). A four mechanism model for threshold spatial vision. *Vision Research*, 19, 19–32.