PERGAMON

Vision Research 39 (1999) 2191–2203

Vision Research

Target detection against narrow band noise backgrounds

Mark W. Cannon [a], Greg J. Reese [b,*], Steven C. Fullenkamp [b]

[a] *AirForce Research Laboratory, AFRL/HECV, 2255 H Street, Wright Patterson AFB, OH 45433, USA*
[b] *Logicon Technical Services Incorporated, PO Box 317258, Dayton, OH 45437, USA*

Received 24 October 1997; received in revised form 18 November 1998

Abstract

We studied the detectability of narrow band random noise targets embedded in narrow band random noise backgrounds as a function of differences in center frequency, spatial frequency bandwidth and orientation bandwidth between target and the immediately adjacent background. Unlike most target detection experiments the targets were not added to the background; they replaced the underlying background texture. Simulations showed that target detection probabilities could be accounted for by a simple transformation on the summed outputs of a two layer filter model similar to the complex channels model proposed by Graham, Beck and Sutter (Graham, N., Beck, J., & Sutter, A. (1992). *Vision Research*, *32*, 719–743). Subsequently, the model was tested on the detection of camouflaged vehicle targets with encouraging results. Published by Elsevier Science Ltd.

Keywords: Detection; Discrimination; Texture; Filtering

1. Introduction

Several visual discrimination models have been developed that produce estimates of visible differences between pairs of images (Daly, 1993; Lubin, 1993, 1995). These models are normally used to address image quality issues by producing spatial maps of visible differences between an original image and a distorted version of the same image (the distortion may be due to noise or to the application of some image processing algorithm).

Recently, it has been shown that discrimination models can also be used to predict the relative detectability of targets in natural scenes (Ahumada, Watson & Rohaly, 1995). In these experiments, model discrimination estimates were obtained from image pairs, where one image consisted of a natural scene and the other image contained the same scene with a superimposed target. Model-generated d' values for target discrimination correlated strongly with d' estimates from single look target detection experiments performed with human observers.

While the discrimination method is quite useful, it does not address the problem of how observers really detect targets. One obviously does not perform a two interval forced choice experiment when doing real world detection tasks. The portion of the background behind the target remains invisible. Apparently, the observer detects some difference between a potential target region and its immediately adjacent background. From the point of view of the visual scientist, the detectability of the target must be some function of the differences in contrast, brightness and the shape of the spatial frequency spectrum between target and background. Contrast and brightness discrimination thresholds have been measured by others for various types and sizes of targets (e.g. Blackwell, 1946; Legge & Foley, 1980). In this paper we primarily address the human observer's ability to detect target to background differences that are mediated only by differences in the shape of the spatial frequency spectrum. The research described in this paper is divided into three parts. First, we used filtered random noise targets embedded in filtered random noise backgrounds and determined the probability of target detection as a function of differences in spatial frequency bandwidth, center frequency and orientation bandwidth between target and background for two different background sets. Our noise

* Corresponding author. Present address: Qualia Computing Incorporation, 2372 Lakeview Drive, Ste. G, Beavercreek, OH 45431, USA.

0042-6989/99/$ - see front matter Published by Elsevier Science Ltd.
PII: S0042-6989(98)00330-7

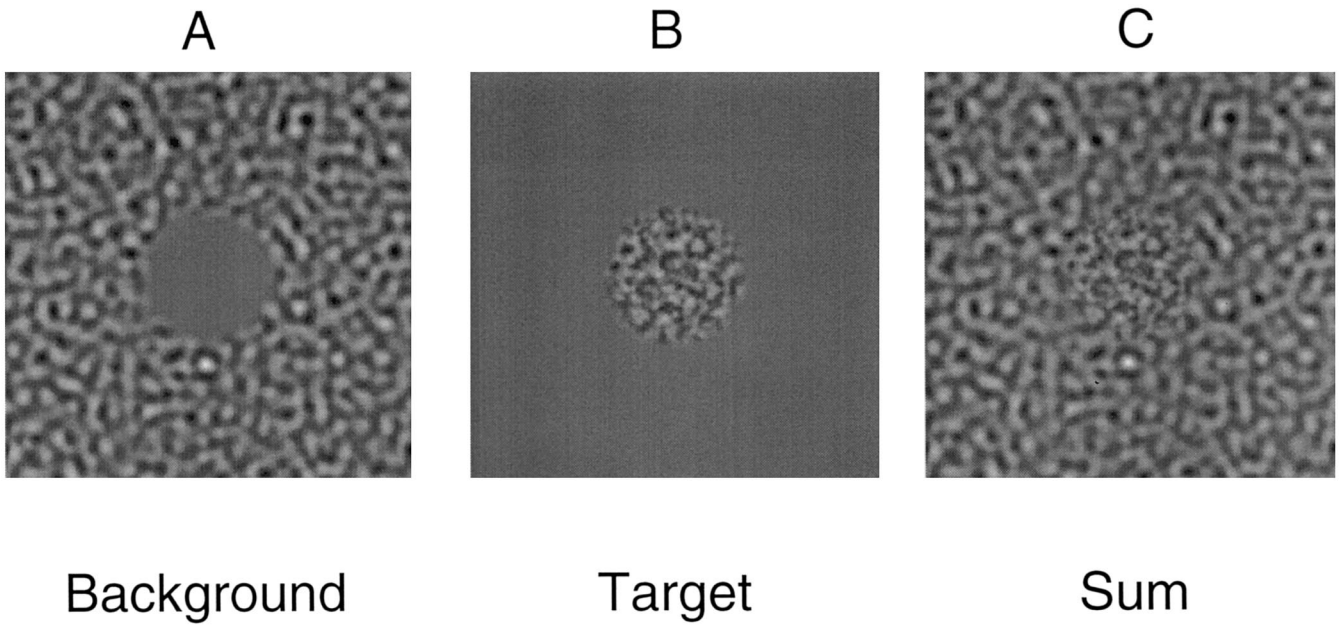

Fig. 1. Typical target and background patterns. Panel A shows a 1 octave spatial frequency bandwidth background pattern. Panel B shows a 1.75 octave spatial frequency bandwidth target. The two images are added in Panel C.

targets had the same mean luminance and contrast as the background, and edge clues were removed. Second, we developed a spatial processing model to account for our probability of detection data and calibrated it using one set of target data. We found that the model could satisfactorily account for four other data sets with the same calibration. Finally, we compared psychophysical data and model predictions for the detection of computer generated camouflaged vehicles on noise backgrounds.

2. Methods and procedures

2.1. Background and target generation

Ten different random noise patterns of uniform density with amplitudes ranging from 127 to -127 were computer generated in 256×256 pixel arrays. These array sizes corresponded to an image subtending $4 \times 4°$ when viewed on our monitor. Ten 1 octave and ten 2 octave spatial-frequency bandwidth noise patterns were created by filtering the uniform noise patterns in the spatial-frequency domain with 1 and 2 octave bandwidth, radially symmetric filters, that were designed to have a Gaussian profile on a logarithmic spatial frequency axis. The filters used for making backgrounds always had a center-frequency of 4 c/deg. Backgrounds were made from these images by multiplying the filtered noise patterns by the space domain profile function P described below;

$$P = 0.0 \text{ for } r < 0.5°$$

$$P = 0.5(1 - \cos\ 2\pi((r - 0.5)/0.5)) \text{ for } 0.5° < r < 0.75°$$

$$P = 1.0 \text{ for } r < 0.75° \quad (1)$$

where r is the distance from the center in degrees of visual angle. This function removed the central region of the noise pattern and smoothed the transition boundary between the blank central area and the surrounding noise. After this function was applied to the noise, a dc level of amplitude 127 was added so the stimulus could be displayed. The amplitude 127 level corresponded to a monitor luminance of 100 cd/m^2. A typical 1 octave spatial-frequency bandwidth background pattern is shown in panel A of Fig. 1.

Target generation was similar to background generation, but will be discussed in some detail, because multiple targets were made from each of the ten unfiltered random patterns mentioned above. The following discussion describes how a set of targets with spatial-frequency bandwidths varying above and below 1 octave was generated. Other targets were generated in a similar manner.

Consider the unfiltered uniform noise pattern generated from the first random-number seed. A total of 16 full 256×256 pixel filtered images were created from this pattern using radially symmetric filters with a Gaussian profile on a log spatial frequency axis. The filters were centered at 4 c/deg, with spatial-frequency

bandwidths ranging from 0.125 to 2 octaves in steps of 0.125 octaves. Once these images were made, they were multiplied by a spatial profile function of the following form;

$$P = 1.0 \text{ for } r < 0.5°$$

$$P = 0.5(1 + \cos 2\pi((r - 0.5)/0.5)) \text{ for } 0.5° < r < 0.75°$$

$$P = 0.0 \text{ for } r > 0.75° \quad (2)$$

where r is the distance from the center in degrees of visual angle. The target boundary smoothing function is the complement of the background boundary smoothing function of Eq. (1). Again, a dc level was added so the stimulus could be displayed. A typical target stimulus is shown in panel B of Fig. 1.

This operation produced a set of 16 targets based on the same random-number seed. Ten 16 target sets were generated; one set for each random-number seed. These targets were used with 1 octave spatial-frequency bandwidth backgrounds.

2.2. Possible edge and contrast artifacts

After the targets were made, two procedures were used to remove edge and contrast artifacts when target and background were added together. First, only targets and backgrounds generated from the same seed were combined. This restriction was required because local differences in internal structure of targets and backgrounds generated from different seeds created the strong perception of a sharp edge along the boundary between target and background even when target and background bandwidths were the same. However, a common seed for both target and background was not a full solution. Increases or decreases in the spatial-frequency bandwidth of the target could change the target RMS contrast slightly, producing both a contrast artifact and a visible sharp edge for a large range of bandwidth differences. To understand how this artifact was removed one must note that there is one target in each set of 16, which we will call T_0, with the same spatial-frequency bandwidth as the background. This target always blended perfectly into that background. No contrast artifact was visible. Consequently, we scaled the central region ($r < 0.5°$) of all targets in each set of 16 to have the same RMS contrast as the central region of the T_0 target for that set. The same scaling factor was used to multiply the target transition region. After viewing all the stimuli we found that this adjustment removed any obvious contrast dependent edge artifacts for experimental conditions where target spatial-frequency bandwidth or center frequency was varied. In the case where the orientation bandwidth was varied, an obvious edge did appear for targets with a target to background orientation-bandwidth difference above 120°. The generation of this edge is inescapable due to the large orientation-bandwidth difference. However, as the data will show, targets were detected by all observers at target to background orientation-bandwidth differences below this value. A typical contrast corrected target plus background image is displayed in panel C of Fig. 1. The target has a spatial-frequency bandwidth of 1.75 octaves while the background bandwidth is 1 octave. There is no sharp circular edge evident around the target. One can, of course, detect an irregular, ill-defined boundary region around the target where one texture ends and another begins, but this boundary appears to be defined by the texture difference.

It has been pointed out by one of the reviewers that, for random stimuli, theoretically correct edge blending requires transition profiles that equate the sum of the stimulus energies for the target and background across the transition region. With a simple linear sum, such as we used, there should be an area in the transition region where the actual RMS contrast of the sum drops below the RMS contrast of the target and background inducing an artifactual edge. In our stimuli, the transition region is an annulus only 0.25° wide. This theoretical contrast reduction would be maximal over a very narrow region in the center of this annulus. The fact that we observed no obvious artifact with changes in target spatial-frequency bandwidth or center spatial-frequency may be due to the fact that the any decrease in perceived contrast was, for the most part, below our contrast discrimination threshold.

2.3. Stimulus presentation

Stimuli were displayed on a Conrac 2600 C15 monitor with a white P4 phosphor. The screen was surrounded by a large white poster board illuminated at the same average luminance as the screen. Calibration was performed with a Pritchard Spectra photometer and contrast linearity was achieved by a computer controlled video signal processor, built in-house. All experiments were conducted using a yes/no forced choice method of constant stimuli.

In most experiments, a parameter of the target, such as spatial frequency bandwidth, varied in 16 discrete steps. Each step was represented by ten different targets (one for each random number seed). During the course of a session, each of the 160 targets was selected randomly, without replacement, and added to the background generated from the same seed. The target and background were presented simultaneously for 3 s at an average luminance of 100 cd/m^2, preceded and followed by auditory tones. A 5 s interval of uniformly illuminated screen, also at 100 cd/m^2, followed the stimulus. During this 5 s interval, the

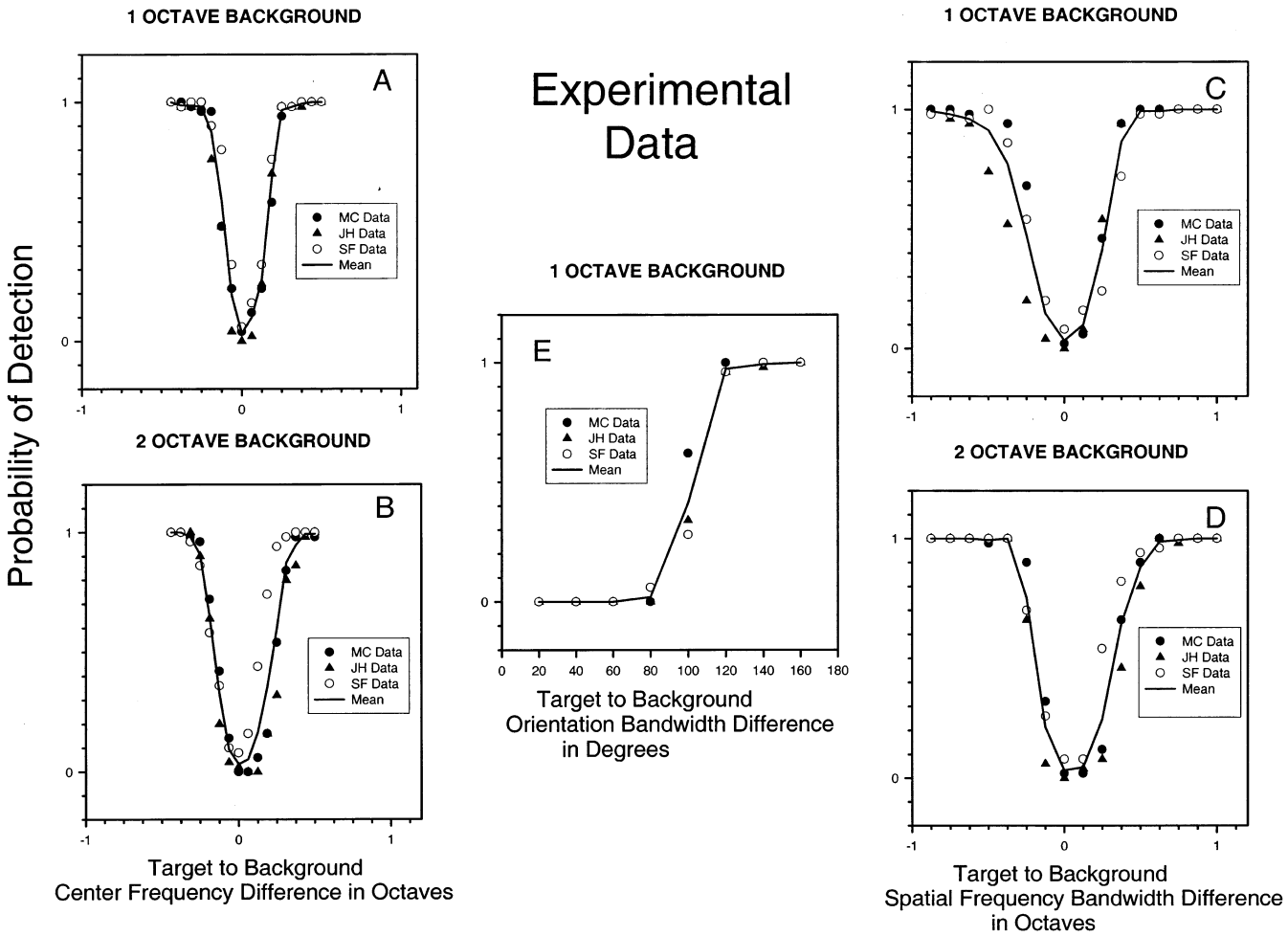

Fig. 2. Experimental data for five experiments on three observers. Solid lines connect data means. In panels A and B, the target center frequency was varied relative to the center frequency of the background. In Panels C and D, the target spatial frequency bandwidth was varied with respect to the spatial frequency bandwidth of the background. In A and C the background spatial frequency bandwidth was 1 octave. In B and D the background spatial frequency bandwidth was 2 octaves. In panel E, the target orientation bandwidth was varied relative to the background. In panel E the background spatial frequency bandwidth was 1 octave. The center spatial frequency of all backgrounds was 4 c/deg.

observer indicated by pressing a two position switch whether the target was visible or not. Observer instructions were to compare the texture in a target region in the center of the screen to the texture of the background regions located near the four corners of the monitor and respond yes if the textures were different. The position of the target region was indicated by two very small dots on the vertical centerline of the screen at 0.5° above and below the screen center. If the switch was not depressed, the same target sequence was repeated until the observer made a choice.

The results were plotted as the number of times the target was detected (a yes response) versus the target parameter value. Plots from five sessions were averaged to produce a psychometric function relating probability of detection to the target parameter.

Three experienced observers, including two of the authors, participated in the experiments.

3. Results

3.1. Changes in target center frequency

Panels A and B of Fig. 2 illustrate the probability of detection results for three observers as a function of the target to background center-frequency difference in octaves. The solid curves connect the mean detection probabilities at each measured center-frequency difference. In these experiments, the background center-frequency was 4 c/deg. The spatial-frequency bandwidths were either 1 or 2 octaves. Bandwidth labels are shown above the panels. Both target and background were created with radially symmetric filters, so they have no orientation component. The widths of the U-shaped tuning curves, at a 50% probability of detection are 0.27 octaves for the 1 octave background and 0.38 octaves for the 2 octave background. Our observers'

abilities to detect the presence of a target by changes in its center frequency appear to degrade as the background bandwidth increases.

3.2. Changes in target spatial-frequency bandwidth

Panels C and D of Fig. 2 illustrate the probability of detection results for the same three observers as a function of the target to background spatial-frequency bandwidth difference. The solid curves connect the mean detection probabilities at each measured bandwidth difference. In these experiments, the target and background center frequency was 4 c/deg. Spatial-frequency bandwidth of the background was either 1 or 2 octaves. Bandwidth labels are shown above the panels. Target and background, created with radially symmetric filters, have no orientation component. Here, the widths of the U-shaped tuning curves, at a detection probability of 50%, are 0.53 octaves for the 1 octave background and 0.52 octaves for the 2 octave background. Our observers' abilities to detect targets due to changes in their spatial frequency bandwidth appears to be unchanged as the background spatial frequency bandwidth increases. In terms of octaves, changes in center frequency are easier to detect than changes in spatial frequency bandwidth.

3.3. Changes in orientation bandwidth

Panel E of Fig. 2 illustrates the probability of detection for three observers as a function of the difference between target and background orientation bandwidth in degrees. The curve connects the means at each measured orientation bandwidth difference. The background orientation bandwidth is defined to be 180° (see Appendix A). Target and background center-frequencies were 4 c/deg and their spatial-frequency bandwidths were 1 octave. Under these conditions, target and background orientation bandwidths must differ by about 105° for the target to be detected at the 50% level.

3.4. Possible bias effects

It is possible that some of the differences among the observers evident in Fig. 2 might be due to bias or criterion effects rather than real individual differences. Due to the way the experiments were conducted, (a yes/no procedure without a rating task) we cannot evaluate the strength of these effects. However, individual bias or criterion effects are not large enough to obscure certain regularities in the data for different experimental conditions. All observers consistently show narrower tuning curves in panels A and B, of Fig. 2, than in panels C and D. All observers also show the transition from 0 to 100% detection, in panel E, over the same narrow range of orientation bandwidths.

4. Model development

4.1. Accounting for the noise target data

A model that can account for our data is outlined in Fig. 3. Input images were exactly the same 256×256 pixel target plus background images presented to the observers in the psychophysical experiments. The images were all padded to 512×512 pixels to reduce wrap around effects caused by digital filtering during the simulation. As shown in Fig. 3, each image was processed by 20 filter mechanisms, specified by five center frequencies and four orientations. The first stage of this mechanism was composed of a sine and cosine phase filter pair followed by a square root of the sum of the squares computation. The first stage sine and cosine phase filters had a 1 octave spatial-frequency bandwidth and a 30° orientation half-bandwidth. The images produced by the filter pair were squared and summed to produce a 512×512 pixel spatial map of the energy passed by the filter. The final operation of the first stage was a pixel by pixel square root of this map. The second stage was a center minus surround computation.

The center minus surround computation was performed on the square root image using the geometry indicated at the bottom of the dotted box. The black border indicates the edge of the viewing screen which covered the central 256×256 pixels of the image. The central circle, labeled, T, designates the region where the target was expected to appear, centered on pixel coordinates (128, 128). The radius of this region was 0.5°, equal to the central region of the target, but not including the transition region. Shaded regions, labeled B1–B4, are the four regions where the observer was directed to sample the background for comparison to the target. These regions were the same size as region T and centered at pixel coordinates, (64, 64), (64, 192), (192, 64) and 192, 192). If the mean image amplitude over the central area is *MT* and the means over the shaded areas are *MB1*, *MB2*, *MB3* and *MB4*, the result of the center minus surround computation is the single number:

$$FR = MT - \frac{MB1 + MB2 + MB3 + MB4}{4} \tag{3}$$

The 20 FR values were rectified and summed to produce a single model response, *R*, for the input image. The mean *R* values for data sets A–D in Fig. 2 are shown in Fig. 4. The horizontal axis is the target to background difference in octaves. When the responses are due to changes in target spatial frequency bandwidth, the axis should be read as a difference in spatial frequency bandwidth. When the responses are due to changes in target center frequency, the axis should be read as a difference in center frequency. Orientation

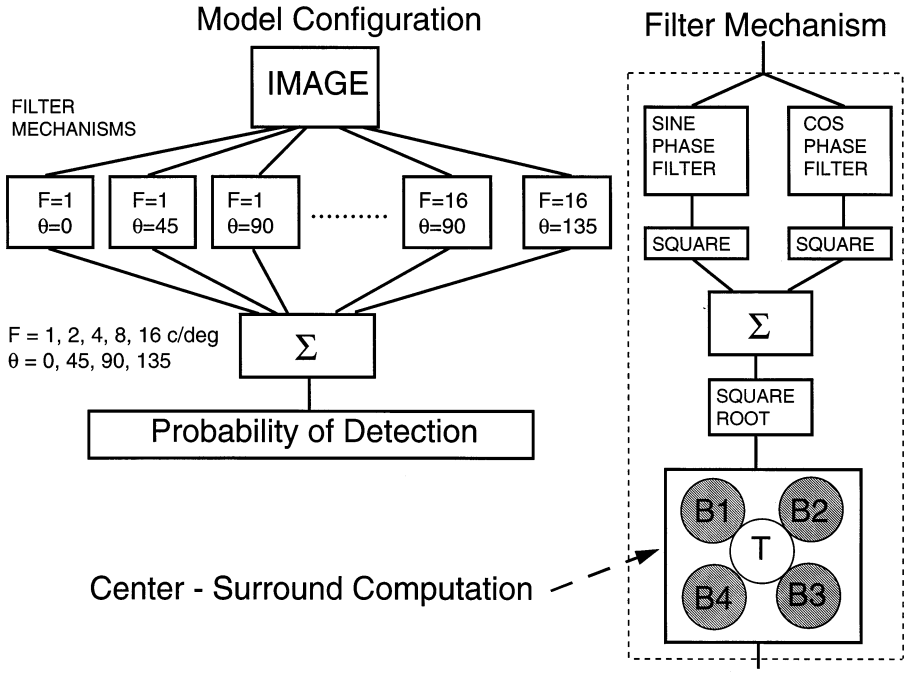

Fig. 3. A model for target detection. Each filter mechanism labeled for spatial frequency, F, and orientation θ, in the model configuration drawing on the left is made up of the components shown within the dotted rectangle on the right. The sine and cos filter components change, but the center minus surround component is the same in each filter mechanism. The model response is the sum of all the filter mechanism outputs. Further details can be found in the text.

tuning is not included in this plot since it is expressed as bandwidth in degrees. Each curve contains 16 points and each point was computed by averaging model responses, R, for the ten stimuli at each target to background difference. Responses due to changing center frequency (triangles) rise more rapidly with target to background difference than responses due to changing spatial frequency bandwidth (circles), in agreement with the data of Fig. 2. However, the model responses continue to increase as the target to background difference increases, unlike the probability of detection functions in Fig. 2 which level off after some critical change in the target to background difference.

In order to account for the probability of detection functions we introduce a transducer function that directly relates the probability of detection to model response, R.

The mean model response function represented by the filled circles in Fig. 4 was used for the derivation. In this data set, the target spatial frequency bandwidth varied by ± 1 octave, in 16 discrete steps, around a mean of 2 octaves. Note that each model response amplitude in this function is associated with a specific target to background spatial frequency bandwidth difference. Note also that the mean observer detection probabilities in Fig. 2D are associated with the same set of target to background spatial frequency bandwidth differences.

These two plots provide us with a functional relationship between model responses and observer detection probabilities. In Fig. 5, the individual observer detection probabilities are plotted as a function of the mean model response amplitude. A function of the form,

$$\mathrm{PD} = 1 - \exp - (((MR - A)/B)^C) \quad (4)$$

was fitted to the data. The parameter values of the fit were $A = 11.3$, $B = 11.5$ and $C = 2.85$. PD stands for probability of detection and MR is the mean model response for this data set. This function maps mean model response onto probability of detection. If the model has any predictive value, we should be able to use Eq. (4) to account for the detection probabilities for all stimuli shown in Fig. 2.

These predictions are the solid curves in Fig. 6. The data points and dashed curves are the data points and data means from Fig. 2. The solid curve of panel D is not a prediction because this is the data set that was used to derive the transducer function. The solid curves in panels A, B, C and E are true predictions. The model accounts quite well for several features of the tuning curves. The predictions in panels A and B reflect the narrow tuning curves for changes in target center-frequency and show an increase in the width of the tuning curve as the background spatial-frequency bandwidth is changed from 1 to 2 octaves. The prediction of the orientation tuning accurately captures the rapid transition from no detection at an orientation bandwidth

difference of 80° to 100% detection at 120°. However, the model prediction for panel C underestimates the mean data over most of the right hand rising portion of the U-shaped tuning curve, and the model predictions for panel B are consistently higher than the data mean near the bottom of the U.

4.2. Quality of fit assessment

The discrepancies noted above indicated that some quantitative assessment of the agreement between model predictions and the data was desirable. We chose to extend a technique described by Watson and Solomon (1997). Their rationale was to determine whether one could tell the difference between the model responses and observer responses if all were presented together. If the RMS differences between observers and model predictions were less than the RMS differences between observer data sets, the model was considered to be a good fit. We computed the RMS differences between pairs of observer data sets (MC-SF, MC-JH, SF-JH) and then computed an average RMS difference across observer pairs for each experimental condition represented by panels A through D in Fig. 6. We also computed the RMS differences between each observer and the model response and averaged these values across observers for each of the panels A through D of Fig. 6. The results shown in Table 1 demonstrate that the mean RMS difference between observer pairs is greater, in every case, than the mean RMS difference between observer data and model prediction. By this criterion, the model responses are indistinguishable from the psychophysical data and hence describe the data well. The test was not required for the data of panel D, since the model prediction and data mean are almost identical.

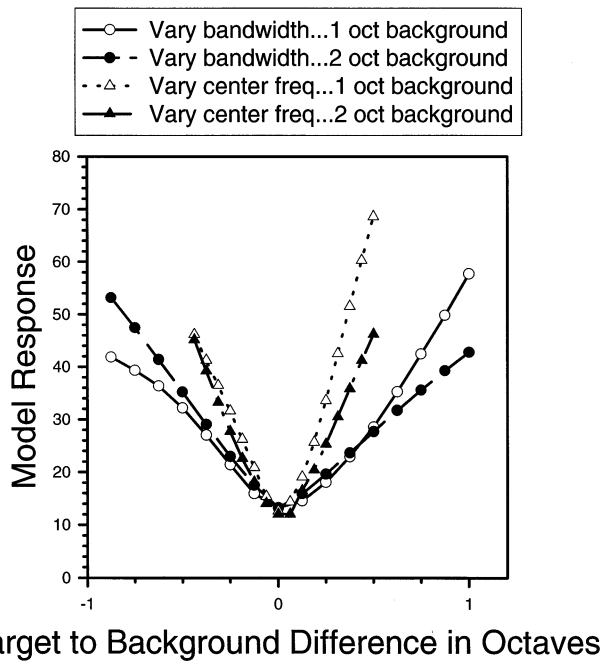

Fig. 4. Mean model responses as a function of target to background difference in octaves. The circles represent the case where target spatial frequency bandwidth varied with respect to the spatial frequency bandwidth of the background. The triangles represent the case where target center frequency varied with respect to the center frequency of the background.

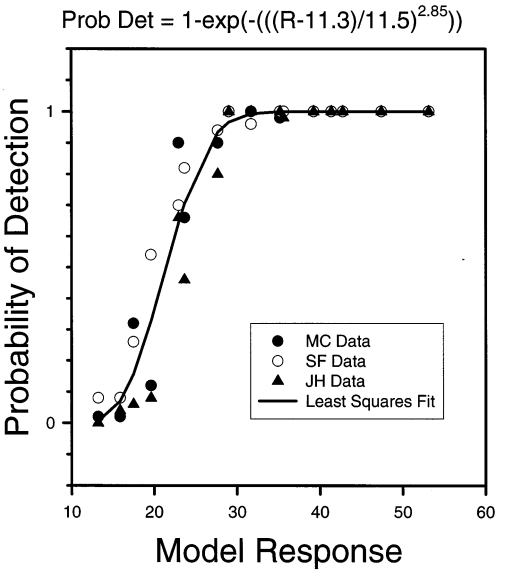

Fig. 5. Derivation of model target detection probability. Since the data points in Fig. 2D and the model responses indicated by the filled circles in Fig. 4 have common horizontal axis coordinates, we can associate a probability of detection from Fig. 2D with each model response in four. These probability of detection values for the three observers are plotted vs. model response in this figure. The equation of a least-squares fit to the data is shown above the figure. This function relates probability of detection to model output.

4.3. Model application to vehicle targets

After analyzing our noise target data, we felt the model was sufficiently promising to test its performance at predicting detection of more realistic targets. We tested the ability of the current model to predict human performance at detecting computer generated camouflaged images of military vehicles on 1 octave bandwidth noise backgrounds when the size and position of the targets were known. The vehicles were approximately 0.8° long and 0.4° wide. Varying degrees of camouflage were added to the vehicles in the following way. The pixels of the vehicle image were added to the noise pixels that they would normally obscure. The equation for this addition was: Pixel amplitude = (Vehicle pixel + ($n \times$ noise pixel))/($n + 1$). The noise parameter, n, took on values of 1, 2, 4, 8, 16, 32, 64 and 128. As n increased, the vehicle disappeared into the noise. The original vehicle images and vehicle plus noise for $n = 1$ and 2, on two typical backgrounds, are shown in Fig. 7.

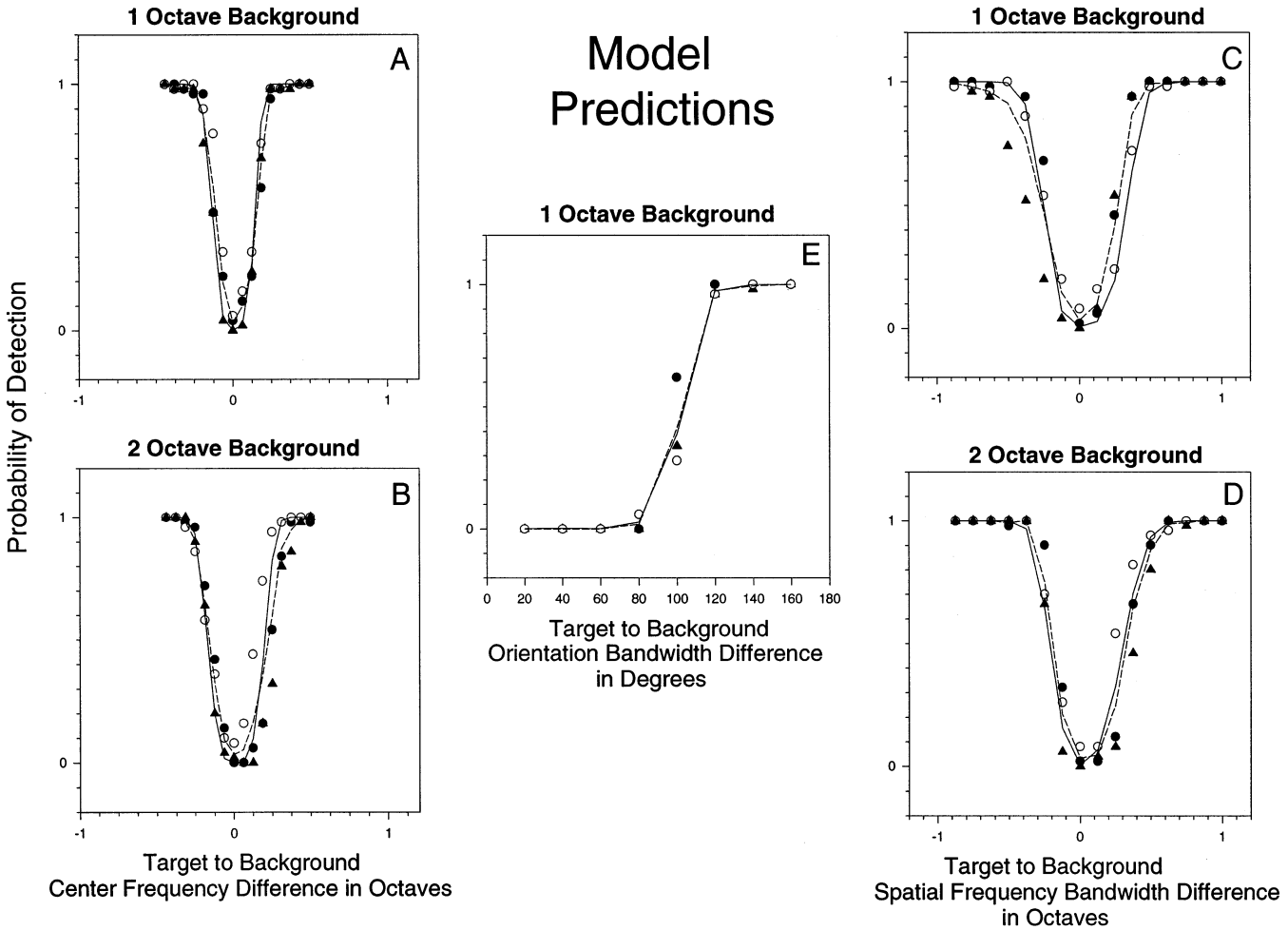

Fig. 6. Model predictions of detection probabilities. The probability of detection equation derived in Fig. 5 was applied to model responses for all the noise targets used in the experiments. Solid lines show the model predictions. The dashed lines are the data means from Fig. 2.

We ran the same type of yes/no psychophysical experiment with this stimulus as with the noise targets. There were ten noise backgrounds for each of the n values so each session consisted of 80 trials. Detection probabilities were estimated by combining the results of five sessions. Data for four observers and model predictions are shown in Fig. 8.

Model predictions underestimate human performance for n near 4, but are among observer responses for all other values of n. We applied the modified Watson and Solomon (1997) quality of fit test to the predictions and data in Fig. 8. The mean RMS difference between pairs of observer responses was smaller than the mean RMS difference between model predictions and observer responses for the truck data indicating a good fit to these data (mean RMS difference among observers = 0.0917; mean RMS difference between model and observers = 0.0837). However, the tank data failed the goodness of fit test (mean RMS difference among observers = 0.0573; mean RMS difference between model and observers = 0.0936) reflecting the fact that the model prediction underestimated all of the experimental detection probabilities at an n of 4. Note that even though the curves do not match the data well at $n = 4$, the predicted and experimental n values for 50% detection are close for both targets. The difference is $\Delta n = 0.8$ for the tank and $\Delta n = 0.3$ for the truck. Possible reasons for the discrepancy at $n = 4$ will be addressed in Section 5.

Table 1
Comparison of RMS error among observers with RMS error between model and observers

	Average mean squared error among observer pairs	Average mean squared error between model and observers
Panel A	0.01043	0.00836
Panel B	0.03955	0.01949
Panel C	0.02346	0.01662
Panel D	0.01765	0.0078

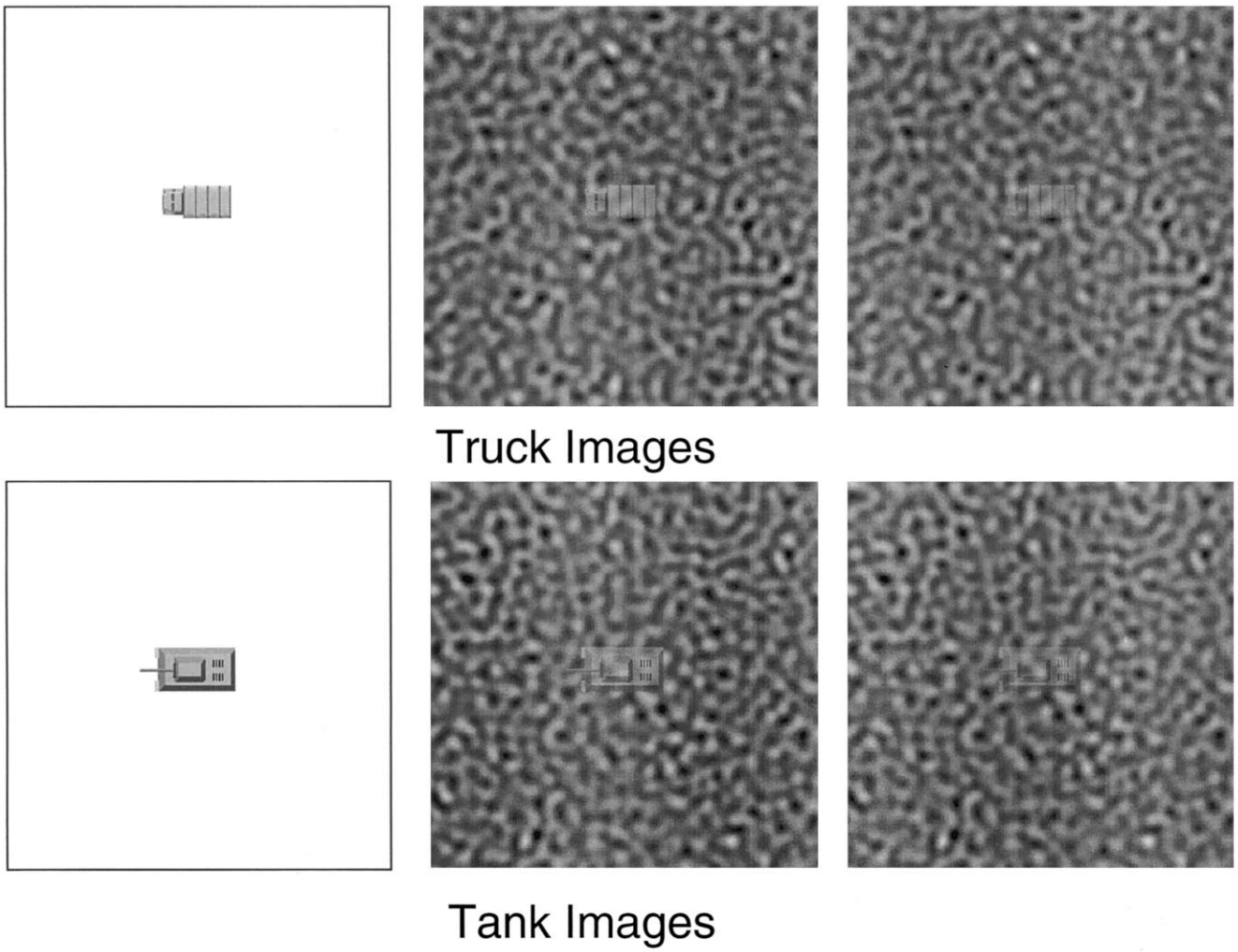

Fig. 7. Computer generated vehicle targets and camouflaged representations. Camouflage is applied by averaging each target pixel with increasing amounts of the underlying noise pixel as explained in the text. Camouflage effectiveness increases to the right.

5. Discussion

5.1. The vehicle data

The model showed a consistent trend to underestimate the human detection data for $n = 4$. This underestimation at $n = 4$ could be due to the size of the central region over which the average activity is computed. We used the same size central region for the vehicle targets as for the noise targets. This region has a diameter of 1°, while the vehicles have a width of approximately 0.4° and a length of 0.8°. Thus, the area occupied by noise alone in the central region is about equal to the area occupied by the target. Since our observers knew the size of the target they were looking for, they may have been able to adapt the size and shape of their central processing region to approximate the size of the target. A reduction of size in the central region appears likely to produce better model performance by increasing the signal to noise ratio in the target region. The hypothesis can be tested in future experiments where the observer has no prior information on target size.

5.2. The contrast sensitivity function

The reader will note that we have not used a contrast sensitivity function in our model. Many contrast discrimination or contrast masking models filter the input image with the CSF before performing any operations with the tuned parallel filters. The attenuation caused by the CSF filter, at 16 c/deg, is about a factor of 8, relative to 4 c/deg (Watson & Solomon, 1997). If a CSF filter is arbitrarily applied to our stimuli, the model responses do not agree with experimental data. This problem is due, primarily, to the suppression of components that are passed by the 16 c/deg filter. In Fig. 4, the open and filled circles show model responses, for a target with changing spatial-frequency bandwidth, when the background has a 1 and 2 octave spatial-frequency bandwidth respectively. The curves are very

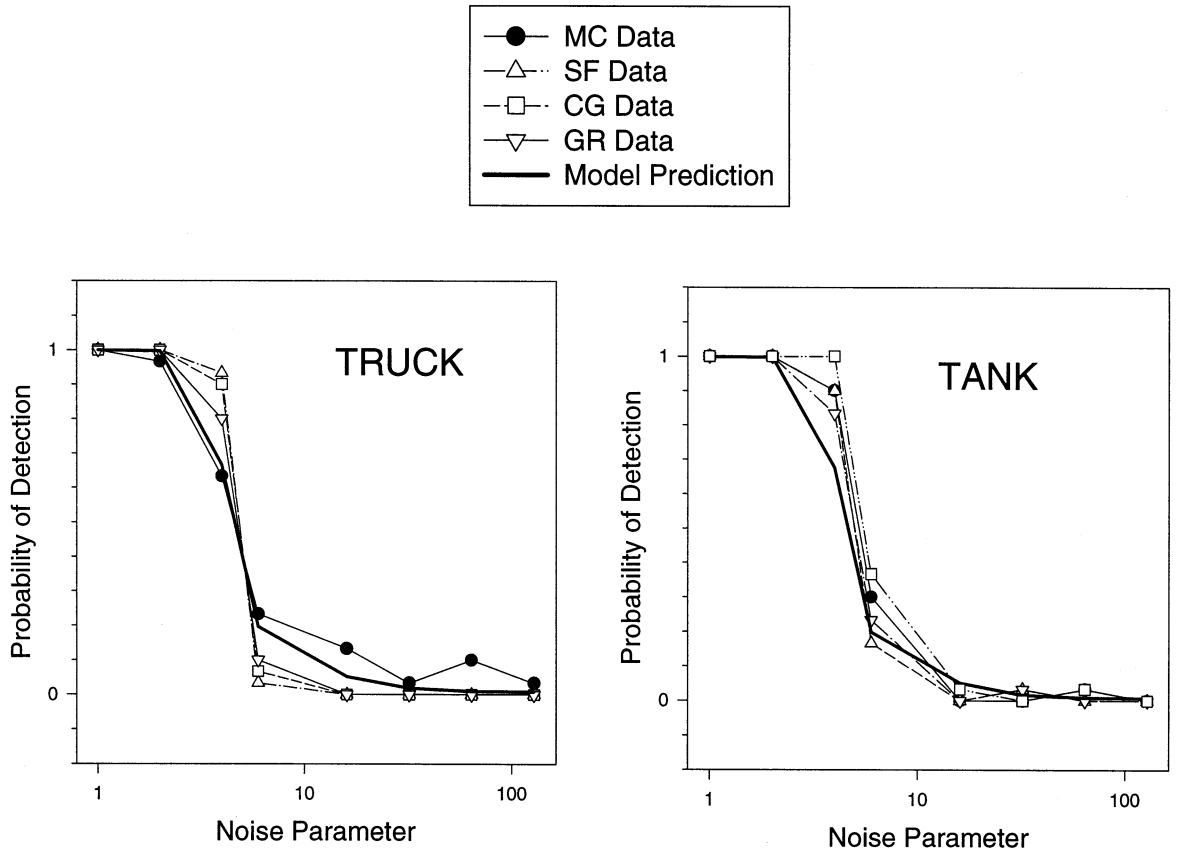

Fig. 8. Data and model predictions for camouflaged vehicles. The camouflage parameter, n, is explained in the text. Camouflage effectiveness increases with increasing n.

close over the first four points to either side of the minimum. This is why the model response curves in Fig. 6C and D are so similar. Application of a CSF filter to the stimuli changes the slopes of both the open and closed circle data in Fig. 4, requiring a new transducer function. However, it dramatically reduces the slope of the increasing bandwidth portion of the filled circle curve relative to the open circle curve. Consequently, the increasing bandwidth portion of the model tuning curve for Fig. 6D would rise much less steeply than the curve in Fig. 6C, and would not agree with the data. Similar problems would occur for any model that computed some measure of the difference between the target and the adjacent background, pooled across spatial frequency channels.

We assert that using a multiplicative CSF is incorrect for our task, since we are attempting to detect a texture difference between two equally visible suprathreshold stimuli in different regions of the screen. After all, the CSF is really the inverse of the threshold and not a multiplicative transfer function. While it is certain that the optics of the eye do cause a reduction of retinal contrast at high spatial frequencies, we propose a filter gain function, acting on suprathreshold stimuli, that increases with filter center-frequency to just counter the attenuation of the optics over the spatial frequency range we considered. Support for such a function is provided by the suprathreshold behavior known as contrast constancy. Contrast constancy has been demonstrated in contrast matching experiments (Georgeson & Sullivan, 1975) and magnitude estimation experiments (Cannon, 1985; Cannon & Fullenkamp, 1991). At our mean luminance of 100 cd/m^2, moderately suprathreshold components from 1 to 16 c/deg would have approximately equal perceived contrast for equal physical contrast and would be essentially unaffected by the high spatial-frequency attenuation due to the optics of the eye. Before we could justify this explanation, we had to determine if the components passed by a 16 c/deg filter were indeed above threshold for human observers. We used the following procedure to check the visibility of these components. We filtered several 2 octave spatial-frequency bandwidth noise images using an expanded set of 1 octave bandwidth filters (five spatial frequencies and six orientations) and added the filter responses together. This sum was an image very close in appearance to the original image, but with an RMS amplitude approximately twice the RMS amplitude of the original image. We then computed the sum of the 16 c/deg filter

responses at all orientations for the same images and displayed the sum as an image with its RMS contrast reduced by a factor of 2. A suprathreshold high spatial-frequency noise pattern was clearly visible on our display. As the target bandwidth increased to 3 octaves, the apparent contrast in the target region of the image passed by the 16 c/deg filter increased. Thus, the high spatial frequency filter components that would be passed by a 16 c/deg filter are moderately suprathreshold allowing approximate contrast constancy to be assumed for our simulation.

5.3. Comparison with other models

The center minus surround block in Fig. 3 was originally set up to perform a computation directly related to the observer instructions (compare the target region to the four corner regions) but the computation is similar to that which would be performed by a radially symmetric center minus surround filter. Extending and joining the MB regions to form an annulus surrounding the target region would not alter the results appreciably. If these computations do represent a second layer of center minus surround filters, tuned to spatial frequencies much lower than the first stage filters, the filter mechanisms illustrated in Fig. 3 are similar to the complex channels, proposed by Graham and her associates (Graham, Beck & Sutter, 1992; Graham, Sutter & Venkatesan, 1993; Sutter & Graham, 1995) and by Sutter, Sperling and Chubb (1995), to mediate second order texture segregation. An equivalent configuration called a texture grabber has also been proposed by Sperling, Solomon, Lu and Chubb (1994).

In the research of Graham and her associates the complex channels were derived from experiments that used rectangular arrays of black or white squares, luminance balanced black and white squares or Gabor patches. Observers were required to rate the segregation of a rectangular region in the center of the array where the elements differed in some property from the surrounding elements. A complex channel consisted of an oriented first layer filter, a rectification operation and a second oriented filter, tuned to a much lower spatial frequency than the first layer filter and oriented orthogonally to it. The requirement for the orthogonal orientation of the two layers may have been a result of positioning their test patches in rectangular arrays. Apparently, only simulations with one set of first layer filters and one set of second layer filters were run, but implicit in the model was the idea that both first and second layer filters could come in a variety of sizes. When the array was composed of Gabor patch stimuli, a first layer filter tuned to the Gabor center frequency was chosen since this would give the strongest response. When the array was composed of squares, a first layer filter that produced strong edge responses was chosen. The spatial frequency of the second layer filter was determined by the spacing between the element rows. Similarly, our center minus surround summation areas were determined by the size of the noise targets we were attempting to detect. Presumably, larger and smaller targets would require, respectively larger and smaller summation areas.

Another similarity between our model and the complex channel is the orientation bandwidth of the first stage filters. We ran simulations with first stage filter orientation half-bandwidths of both 15 and 30°. The 15° half-bandwidth filters produced model responses to changes in target orientation bandwidth that were much too large. With 30° half-bandwidth filters, the model predictions were quite accurate. Graham et al. (1993) required first layer filter half-bandwidths of this order to fit their data.

Orientation dependence in second layer filters would not be picked up in our noise experiments since the boundaries of our noise targets are circular. Even where the first layer filters responded strongly to oriented noise, this response more or less filled the central circular target area. After the square root of the sum of the squares computation over this area, the second layer filter would only see a radially symmetric target blob. This blob could be detected equally well by both oriented and radially symmetric second layer filters.

One difference between the complex channel of Graham et al. and our model is the connectivity between the first and second layer filters. In their model, a specific size second layer filter received inputs from only one size of first layer filter. In our model, a specific size second layer filter would receive inputs from a wide range of first layer filter sizes. The responses of these second layer filters are then summed to give an output related to detection probability. Summation or response pooling in our model is always a simple summation. Most response pooling equations in the literature are of the form of a Minkowski sum with an exponent of 2, 3 or 4. It should be noted, however, that Graham et al. (1992) investigated response pooling effects in complex channels with a variety of exponents and were able to adequately model their data with an exponent of 1, which is equivalent to our linear summation.

Complex channel models assume the existence of real second layer filters. Our experimental instructions requiring observers to scan several regions of the display appeared to create a situation where observers compared their memories of the average appearance of several foveated surround areas to the foveated image of the target area and performed some mental equivalent of a center minus surround computation. This would preclude interpreting our center minus surround computation as a real filter. However, we found that observers did not follow these instructions exactly. For most trials, observers reported that detection was im-

mediate. No corner scanning was required and none was initiated. This ability to detect the target with no purposefully initiated eye movement supports the real second stage center minus surround filter hypothesis for most of our trials. For trials when the decision was difficult, observers followed the experimental instructions by foveating both the target and the several background regions before making a decision. The presence of these two strategies implies that second layer filters may be used for target detection if the characteristics of the target and background parameters differ by some critical value. If the differences are less than the critical value a decision must be made on the memories of foveal inspections of the two regions.

It should be noted that the comparison process proposed here overcomes one of the criticisms (Watson & Solomon 1997) of current discrimination models. In most of these models a pixel by pixel differencing operation requires the observer to have a photographic memory of both images. In the model described in this paper, the observer is only required to compare average responses computed over several spatial regions. This appears to be a more realistic approach to target detection.

Appendix A

A.1. Radially symmetric filter

The envelope of the radially symmetric portion of the filter spectrum is given by

$$R(f) = \exp - [\log(f/f_0)/((n/2)0.36157)]^2 \quad \text{(A1)}$$

where f_0 is the center frequency in cycles per degree; n is the spatial frequency bandwidth in octaves; f is the radial distance from the origin of the f_x, f_y plane in cycles per degree.

The filter is symmetric about f_0 on a log f-axis. If the spatial frequency bandwidth is 1 octave ($n = 1$), $R(f) = 0.5$ when $f = 1.414 f_0$ or $f = 0.707 f_0$.

This filter function was used to create the filtered random noise used in the experiments. An oriented version, used to create the first layer filters and the oriented noise patterns is a simple modification of this spectrum as explained below.

A.2. Orientation selective filter

The spectrum of an oriented filter is obtained by multiplying the radially symmetric spectrum derived above by another function. Consider the drawing in Fig. A1.

The circles labelled f_1 and f_2 are the low and high frequency half amplitudes of the radially symmetric

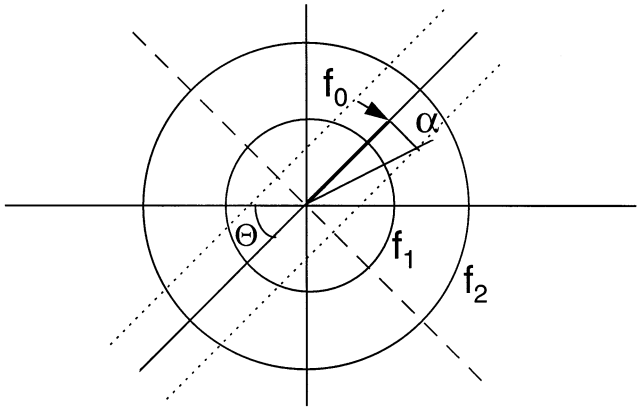

Fig. A1. The geometry of the orientated filter. See text for details.

spectrum. The center frequency is labeled f_0. We want a filter with orientation axis defined by Θ. Define a line perpendicular to the orientation axis at center frequency f_0. The length of this line is

$$y'_0 = f_0 \tan \alpha. \quad \text{(A2)}$$

The angle α is defined as the orientation half bandwidth in degrees. The term y'_0 describes the space constant of a Gaussian weighting function

$$W = \exp(-(y'^2)/(y'_0)^2) \quad \text{(A3)}$$

The dotted lines parallel to the orientation axis represent constant amplitude contours for this Gaussian function at $y' = y'_0$.

If f_x and f_y are the coordinates of some point in the spectrum, the distance, y', of this coordinate from the orientation axis is obtained from the usual formula for the distance from a rotated axis,

$$y' = -f_x \sin \Theta + f_x \cos \Theta. \quad \text{(A4)}$$

Thus, for any f_x, f_y pair we can compute an orientation weighting function W.

The oriented filter spectrum is the product of R and W.

Cosine phase first layer filters are obtained by making the real portion of the spectrum equal to RW and setting the imaginary part equal to zero. Sine phase filter spectra are obtained by setting the real part of the spectrum equal to zero and setting the imaginary part equal to RW multiplied by $+1$ over the region of the plane to the left of the dashed line in the figure and by -1 over the region of the plane to the right of the dashed line in Fig. A1. Sine and cosine phase filters are illustrated in Fig. A2.

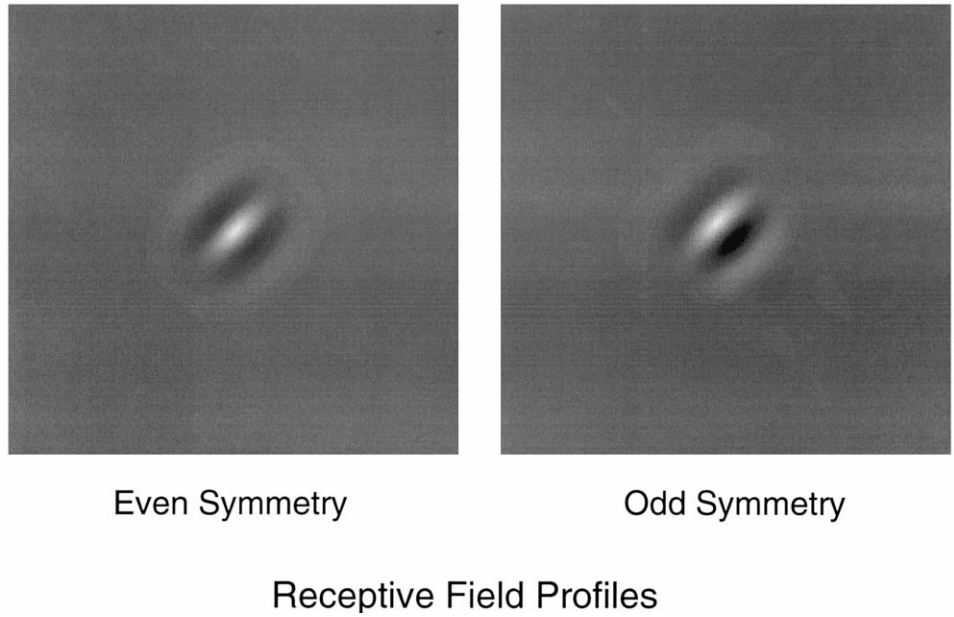

Fig. A2. Filter receptive field profiles for filters with 1 octave bandwidth and 30° orientation half bandwidth.

References

Ahumada, A. J. Jr., Watson, A. B., & Rohaly, A. M. (1995). Models of human image discrimination predict object detection in natural backgrounds. In B. Rogowitz, & J. Allebach, *Human vision, visual processing and digital display IV*, *SPIE Proc*., vol. 2411 (pp. 355–362). Bellingham, WA: SPIE.

Blackwell, H. R. (1946). Contrast thresholds of the human eye. *Journal of the Optical Society of America*, *36*, 624–643.

Cannon, M. W. (1985). Perceived contrast in the fovea and periphery. *Journal of the Optical Society of America A*, *2*, 1760–1768.

Cannon, M. W., & Fullenkamp, S. C. (1991). A transducer model for contrast perception. *Vision Research*, *31*, 983–998.

Daly, S. (1993). The visible differences predictor: an algorithm for the assessment of image fidelity. In A. B. Watson, *Digital images and human vision* (pp. 179–206). Cambridge MA: MIT Press.

Georgeson, M. A., & Sullivan, J. M. (1975). Contrast constancy: deblurring in human vision by spatial frequency channels. *Journal of Physiology*, *London*, *252*, 627–656.

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

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

Legge, G. E., & Foley, J. M. (1980). Contrast masking in human vision. *Journal of the Optical Society of America*, *70*, 1458–1471.

Lubin, J. (1993). The use of psychophysical data and models in the analysis of display system performance. In A. B. Watson, *Digital images and human vision* (pp. 163–178). Cambridge MA: MIT Press.

Lubin, J. (1995). A visual discrimination model for imaging stystem design and evaluation. In E. Peli, *Spatial vision models* (pp. 245–283). New Jersey: World Scientific.

Sperling, G., Solomon, J. A., Lu, Z., & Chubb, C. (1994). First and second-order processes in the perception of motion and texture. In J. M. Zurada, R. J. Marks II, & C. J. Robinson, *Computational intelligence: imitating life* (pp. 223–236). New York: IEEE Press.

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

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

Watson, A. B., & Solomon, J. A. (1997). Model of contrast gain contral in pattern masking. *Journal of the Optical Society of America A*, *14*, 2379–2391.