Detection in Fixed and Random Noise in Foveal and Parafoveal Vision

Explained by Template Learning

Bettina L. Beard and Albert J. Ahumada, Jr.

NASA Ames Research Center, Human Information Processing Research Branch, Moffett Field,

California 94035-1000

Foveal and parafoveal contrast detection thresholds for Gabor and checkerboard targets were measured in white noise using a two-interval forced-choice paradigm. Two white noise conditions were used: *Fixed* and *Twin*. In the *Fixed* noise condition, a single noise sample was presented in both intervals of all trials. In the *Twin* noise condition, the same noise sample was used in the two intervals of a trial, but a new sample was generated for each trial. *Fixed* noise conditions usually resulted in lower thresholds than *Twin* noise. We present template learning models that attribute this advantage of *Fixed* over *Twin* noise either to fixed memory templates reducing uncertainty by incorporating the noise, or by the learning process itself introducing more variability in the *Twin* noise condition. Quantitative predictions of the template learning process show that it contributes to the accelerating nonlinear increase in performance with signal amplitude at low signal-to-noise ratios.

-

1. Introduction

Both image discrimination[1-7] and template-matching models[8-10] may be used to predict noise masking in target detection. Image discrimination models measure, or predict, the discriminability between two input images. To predict noise effects with an image discrimination model, a single white noise sample can be added to both input images, while only one image contains the target. A visual system module giving internal image representations processes the two images. The module simulates features of human vision, such as optical blurring, luminance and contrast effects, and masking by nonlinear transduction within orientation selective channels. Differences between the representations are then summed. Representation components unaffected by the target do not contribute to the aggregated difference image since the noise samples are identical, but the nonlinear processing can allow the noise to reduce the difference in the components responding to the target. Image discrimination models can predict masking effects when a single noise sample (to be called the *Fixed* noise condition) is used to mask a target or its absence.[7,11]

Another condition that permits prediction by image discrimination models is the *Twin* noise condition used by Ahumada & Beard,[7] where a single noise sample is used for both images within a single two-interval trial, but a new random noise sample is used for each trial. Although image discrimination models predict the same average thresholds in *Twin* and *Fixed* conditions, and little variation from different white noise samples, we found a threshold elevation for *Twin* noise conditions relative to *Fixed* noise samples, particularly at higher noise RMS (root mean square) contrast levels.[7] Here we present data that extend this finding to other stimuli and to parafoveal retina.

A threshold elevation for *Twin* noise relative to *Fixed* noise suggests that the observer can not compare the two eidetic image representations during a trial. One might try to explain the degradation caused by *Twin* noise by including a model of the short-term visual storage system. A simpler, more common, approach is to assume that a sensory representation of each image is compared to an internal memory representation, or template, on each stimulus presentation. Then only the result of the comparison need be remembered. If the template is modified by the current sensory representation, the changing noise samples in the *Twin* condition could add noise to the template as it is being constructed. *Fixed* noise, on the other hand, could contribute to the template construction process.

Template matching models[8,9] correlate sensory representations of the images with one or more memory templates. However, when the template models assume a single, unchanging memory template for *Fixed* and *Twin* noise conditions,

they do not predict a detection threshold difference. In this paper we introduce template-matching models with adaptable templates. The models are too simple to predict our threshold data, but they do illustrate the possible role of template learning in the lowering of target detection thresholds in *Fixed* noise conditions relative to *Twin* noise conditions.

In contrast to the *Fixed/Twin* effect found by Ahumada & Beard[7], Watson *et al.*[12] reported no difference between Gabor target thresholds under *Fixed* and *Twin* conditions. Since a Gabor target may be difficult to localize, especially along the grating bars, we compare a Gabor target with a checkerboard target. The reasoning is that template construction and matching processes should depend upon the position alignment between the sensory representation and the memory template. The constantly changing noise in the *Twin* noise condition should increase uncertainty about the target position, while the unchanging noise pixels in the *Fixed* noise should aid in locating the target position. Increasing uncertainty should then dilute the *Fixed/Twin* effect. We include two variables that may affect positional uncertainty: (i) Gabor versus checkerboard targets and (ii) eccentricity.[13]

If template construction is influenced by the noise structure, then following practice on one *Fixed* noise sample with a different *Fixed* noise sample should result in poorer performance (i.e., negative transfer of learning), both from having a less appropriate template and from increased positional uncertainty. To test this, we train observers on one noise sample and observe transfer of training to others.

In the absence of a mask, peripheral target detection thresholds are elevated relative to foveal measurements.[14,15] Noise mask effects on peripheral target detection are unknown. Here we extend measurements to a parafoveal location. We also investigate whether the cortical-magnification-scaling factor often used to equate foveal and peripheral detection thresholds[16] equates them in the presence of a white noise background.

In summary, we

(1) Attempt to replicate the earlier Ahumada & Beard[7] finding of a *Fixed/Twin* noise threshold difference with two new stimuli,

(2) Test whether stimulus structure or eccentricity affect the *Fixed/Twin* noise threshold difference,

(3) Measure white noise effects on object detection in foveal and peripheral vision,

(4) Determine whether the cortical magnification scaling factor often used to equate foveal and peripheral detection thresholds holds in the presence of a white noise background,

(5) Present two template learning models,

(6) Give reasons to include memory templates and template learning in models of detection and discrimination.

2. Methods

2a. Apparatus

Stimulus images were presented using the green gun of a cathode-ray-tube (SONY Trinitron Color Graphics Display, model GDM-20E1) with 640 x 480 pixel resolution at 60 frames per sec. The viewing distance was either 122 cm, which gives a display resolution of 55.3 pixels/deg, or 40.6 cm, giving 18.4 pixels/deg. Images were 128 x 128 pixels (2.3 deg of visual angle for the farther distance and 6.9 deg for the closer). The surrounding screen luminance was 19 cd/m^2. To ensure that luminance was linear with digital value, a look-up table was used. A resistive mixing circuit was used to increase the accuracy of the linearization as described in Ahumada & Beard.[7]

2b. Experimental Stimuli

To study noise effects on detection performance, two gray-scale digital target images were generated. One image contained a horizontally oriented, odd-symmetric Gabor with circular Gaussian windowing. The Gabor target had a vertical spatial frequency of 3.7 c/deg and its 1/e spatial half spread (s in $e^{-(x/s)^2}$) was 0.9 deg. A Gabor stimulus was chosen because Watson *et al.*[12] found no difference between *Fixed* and *Twin* noise thresholds using a Gabor stimulus. The second digital image contained two light squares and two dark squares arranged as a checkerboard in a 1.8 x 1.8 deg square, similar in size to the Gabor.

The upper two panels of Fig. 1 illustrate the targets used in this experiment. The lower two panels show the result of a

simulation illustrating the relative spatial uncertainty of the checkerboard and Gabor targets in the presence of noise. These simulation distributions are the best cross-correlating positions of the target with itself in 200 random samples of white noise. Darker pixels represent a higher number of maximum scores. The positional uncertainty is greater for the Gabor than for the checkerboard and greater for the Gabor in the direction parallel to the edge rather than perpendicular to it. The signal-to-noise ratio was set so that the ideal observer d' would be 3.5. This value, representing the external as well as sensory noise, allows for additional decision and memory noise following the signal localization. The simulation, therefore, supports the possibility that a Gabor stimulus could produce greater spatial uncertainty than a checkerboard stimulus.
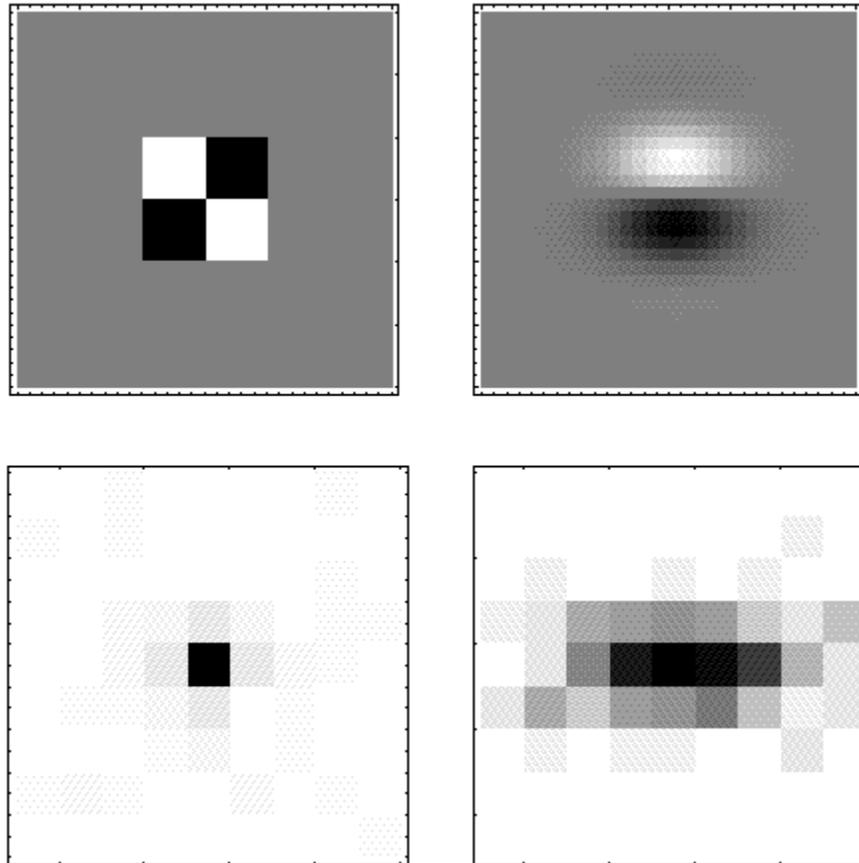

Fig. 1. Top two panels, the checkerboard and Gabor stimuli used in the experiment. Bottom two panels, the 200 best-correlating positions for each target with the (centered) target in noise, showing greater uncertainty for the Gabor target than for a checkerboard target.

Experimental images were constructed by adding a fraction of the target image to a white noise image. The white noise image pixels were independently, identically distributed uniform random variables with a mean of zero. One white noise image was constructed with uniformly distributed amplitudes. It had a peak contrast of 0.33 and associated root-mean-square (RMS) contrast of 0.19. To get a new sample, a random number from 1 to 128^2 was chosen as the starting position for the upper left corner of the noise and then subsequent values in the table were copied with wrap-around. For *Fixed* noise samples, this starting position was unchanged from trial to trial within a block. For *Twin* noise, a new starting position was determined for each two-interval trial.

Small location marks were used to increase position knowledge. These marks were five pixels wide and positioned at the same vertical location as the target center just outside of the 128 x 128 image.

2c. Noise Type

There were two noise sample randomization conditions: (i) In the *Fixed* noise condition, a single *Fixed* noise sample was used throughout a series of trial blocks. (ii) In the *Twin* noise condition, a new noise was generated for each trial and used for both intervals in a two-interval trial.

For each observer, the order of the conditions (2 stimuli x 3 eccentricities) was chosen at random. For observer CM the

order for 2 stimuli x 2 eccentricity conditions were randomized. The noise conditions (*Fixed* versus *Twin*) were always run successively, with their order independently randomized. Limited sets of *Fixed* noise samples were randomly determined for each condition. From this set, the experimenter determined the particular *Fixed* noise sample for any set of trials and how many trial block repetitions there were for any single *Fixed* noise sample.

2d. Eccentricity

Our earlier measures[7] were taken in the fovea. Because the periphery codes less phase information,[13] here we measured detection thresholds in the fovea and at 4 deg eccentricity (inferior visual field) from the same 122 cm viewing distance. Observers also viewed the display from 40.6 cm at 4 deg eccentricity to get an image scaled by a factor of three. This factor approximates cortical magnification estimates for contrast detection thresholds.[16]

2e. Procedure

A two-interval forced-choice staircase tracking procedure was used with stimulus duration of 0.5 sec and an interstimulus interval of 1.0 sec. The stimulus images appeared in the screen center and were replaced by the background luminance during the interstimulus and intertrial intervals. For one of the intervals, selected at random on each trial, the target image was absent (multiplied by zero). The observer's task was to determine the interval in which the target was presented. Auditory feedback was given if the keyboard response was incorrect. If the observer made an error, the target contrast energy was increased by 1.4 dB. If the observer was correct three trials in a row, the target energy was decreased by 1.4 dB. Conditions (i.e., target, noise type, and eccentricity) were fixed for blocks of 60 trials. Thresholds were determined for each block by maximum likelihood probit analysis that estimates the multiplier level leading to 75% correct.[17] The analysis assumed the psychometric function is a cumulative normal distribution.

2f. Observers

Five observers participated in this experiment: CM, NR, PW, SM and BLB. Observers NR and BLB had extensive experience detecting an aircraft image in a runway scene masked by *Fixed* noise in an earlier experiment.[7] Observer CM did not gather data in the 4 deg magnified condition.

3. Results

First we present the data averaged across trial blocks for each observer. These results are followed by foveal block-by-block data of three of the five observers to illustrate practice effects.
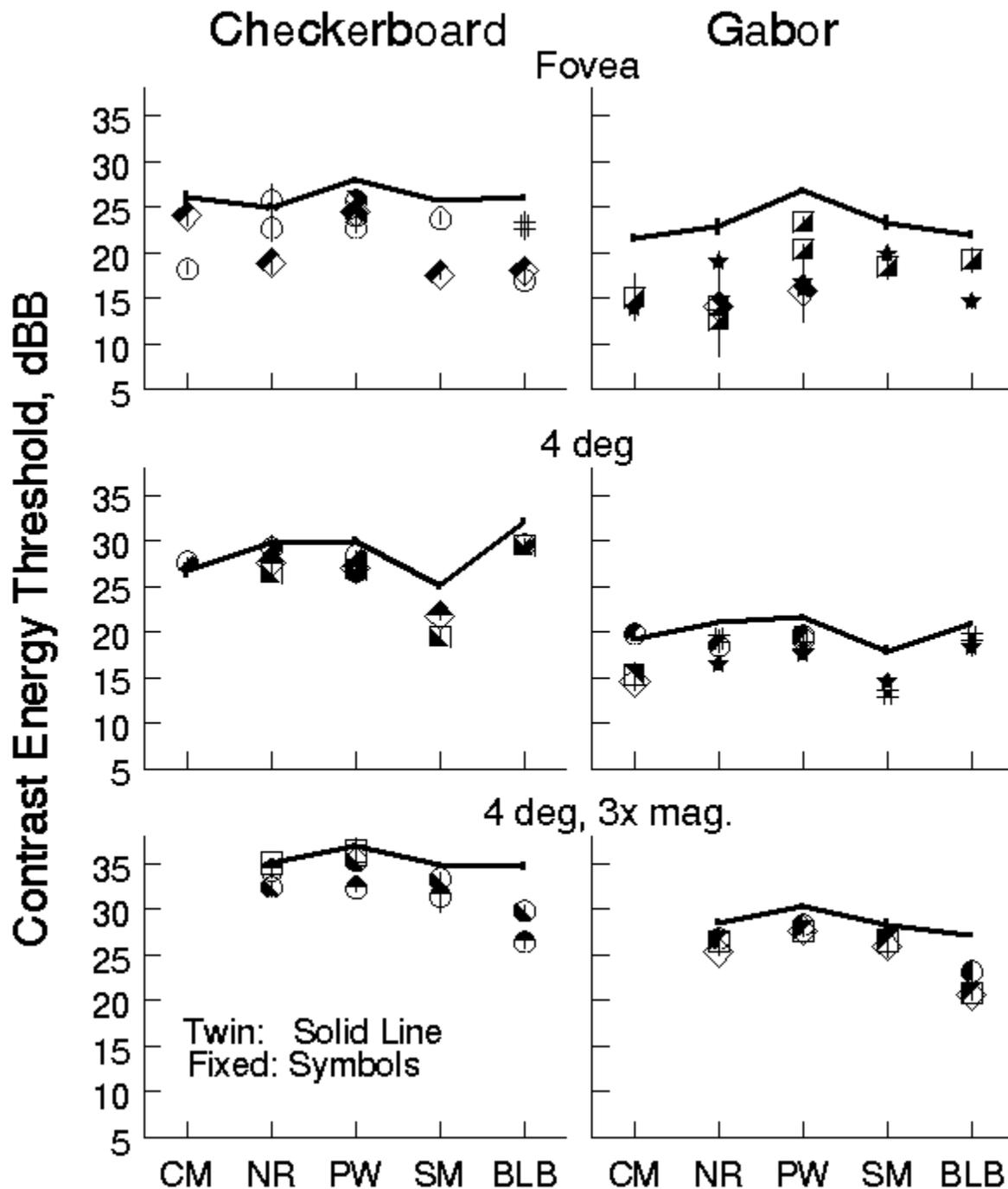
Fig. 2. Average contrast energy thresholds (in dBB) are plotted for five observers. Thresholds for the checkerboard target are shown in the left-hand panels; those for Gaobr targets are in the right-hand panels. Data collected in the fovea are in the upper panels, data from 4-deg eccentricity (foveal viewing distance) are in the center panels, and data from 4-deg eccentricity from a closer (one third) distance are in the bottom panels. Solid curves represent the twin noise data. Particular symbols represent averaged thresholds for each fixed noise sample.

In Fig. 2, average target contrast thresholds are plotted for five observers. Thresholds are reported in dBB, a decibel contrast energy scale that is defined as

$$dBB = 10 \log_{10}(CE/CE_0), (1)$$

where CE_0 is 10^{-6} deg^2 s, the best contrast threshold reported by Watson et al.[18] in their observer HB Barlow).

Contrast energy (CE) is determined by,

$$CE = AT \, S \, c_{ij}^2, \quad (2)$$

$$c_{ij} = (a_{ij} - B) / B \quad (3)$$

where a, the digital amplitude (linear with luminance) of each image pixel is converted to contrast c by B, the digital background level. A is the area of a single pixel in deg^2 and T is the stimulus duration in sec. Solid lines connect *Twin* noise averages. *Fixed* noise averages are shown with different symbols for each different *Fixed* noise sample. Foveal data are presented in the two upper panels, unmagnified parafoveal data are shown in the center panels, and the lower panels show parafoveal data for the image magnified three times. The left panels show data for the checkerboard target and the right panels show data for the Gabor target.

Fig. 2 illustrates several major results. Statistical comparisons were made using 95% confidence intervals based on the variability of the comparison between observers. The major result is the significant performance improvement for *Fixed* over *Twin*. In the fovea, the *Fixed/Twin* difference (collapsed across target type) averaged 5.5 ± 0.8 dB, while at 4 deg it was 2.6 ± 1.5 dB unmagnified, and 3.2 ± 3.1 dB magnified. Excluding observer CM, since she did not make measurements for the 4 deg magnified condition, the foveal effect is significantly larger than the peripheral effects, which do not differ from each other. None of the noise type (*Fixed/Twin*) effects interacted significantly with the target type (checkerboard vs. Gabor).

Next we investigate if there is a significant difference between the checkerboard and Gabor target thresholds. For the *Twin* noise conditions, the Gabor target led to significantly lower contrast energy thresholds than the checkerboard in the fovea (2.9 ± 1.7), at 4 deg unmagnified (8.6 ± 1.9), and at 4 deg magnified (5.5 ± 4.4). The effects of eccentricity depended on the target type. For the *Twin* conditions, moving the checkerboard target into the periphery led to a 2.5 ± 3.5 dB, non-significant performance decrement. To determine if the magnification factor used equated the 4 deg magnified and foveal thresholds, we transformed the contrast energy (dBB) scores to contrast threshold units. The graphs of Fig. 2 show contrast energy (dBB). If the 4 deg magnified graphs are shifted down by 9.5 dB ($20\log_{10}(3)$), the graph can be compared with the others in the contrast domain. The contrast domain comparison shows that the magnification condition led to a threshold that was only 0.4 ± 0.9 dB below that of the fovea. Moving the Gabor target into the periphery led to a significant 3.1 ± 2.4 dB improvement in detection, so no compensatory magnification is required. Again we transformed contrast energy scores to contrast threshold units and found that for a Gabor target, the magnification led to performance 4.7 ± 1.6 dB better than that expected from contrast threshold equality.
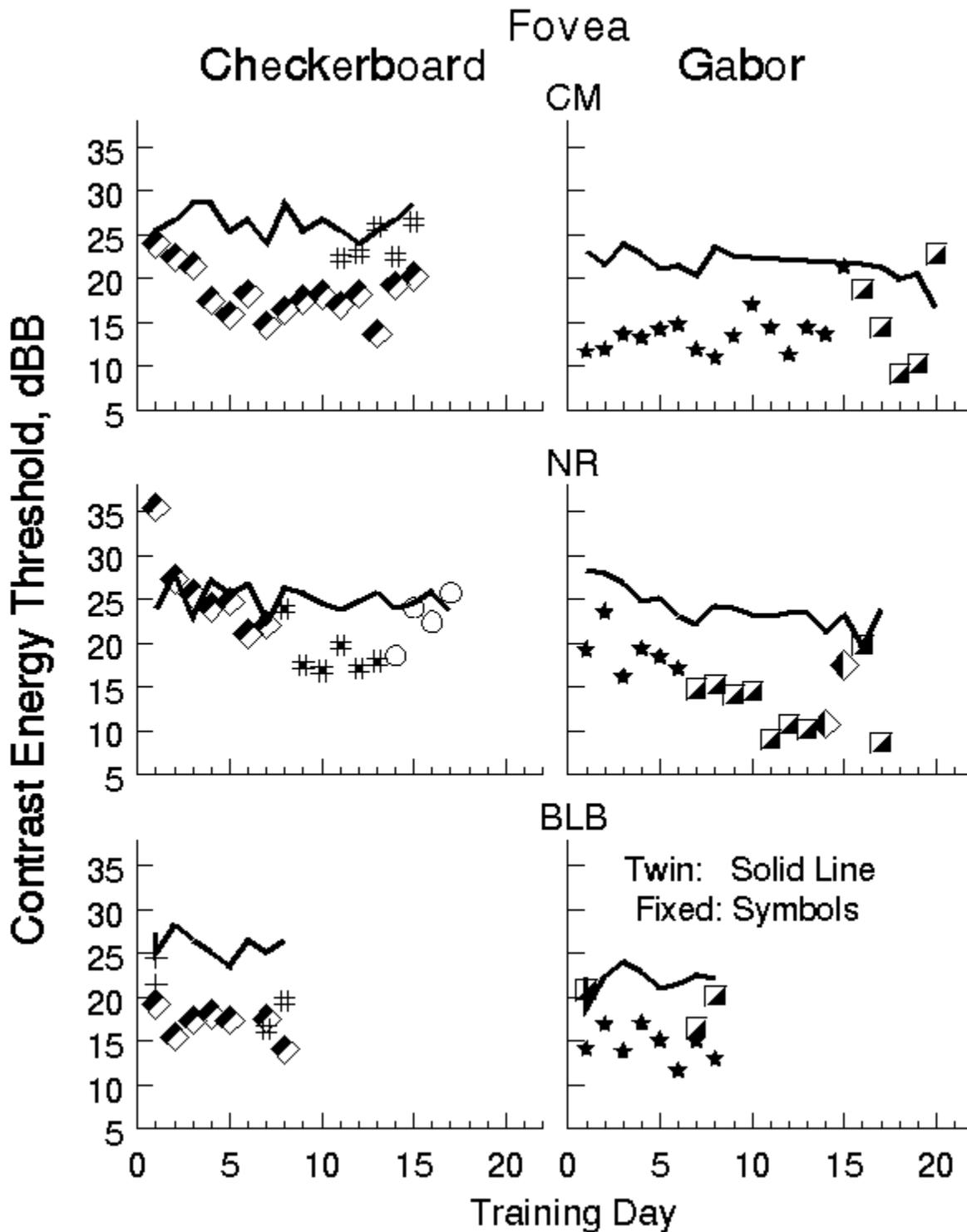
Fig. 3. Foveal contrast energy threshold (in dBB) is plotted as a function of the training day. Each data point is based on 60 two-alternative-forced-choice trials. Solid curves represent the twin noise data. The remaining symbols represent fixed noise data. Fixed noise samples have the symbols assigned in Fig. 2. Checkerboard target thresholds are shown in the left-hand panels; Gabor target thresholds in the right-hand panels. The foveal data for three of five observers are shown.

Fig. 3 illustrates the learning and transfer of training results for the foveal data of observers CM, NR and BLB. The left panels show data for the checkerboard pattern and the right panels for the Gabor stimulus. *Twin* noise data are connected with a solid line, while data for each *Fixed* noise has a unique symbol for that noise sample. The *Twin* noise data show little improvement over sessions, or training day. Regression analyses showed significant initial improvement in *Twin* noise thresholds for observer NR (Gabor) and improvement on the final day for observer CM (Gabor). With *Fixed* noise

the more naive observers CM and NR showed strong improvement over as many as five days, sometimes in just one day (NR, Checkerboard, filled #), and sometimes none at all (CM, Gabor, star). The most experienced observer (BLB) showed no significant learning. There is evidence of idiosyncratic negative transfer of training. When observer CM added the noise (#), her performance was poor, while BLB performed well with the same maskers. Also the switch from masking the Gabor by the noised marked by the star to that marked by the lower-right-filled square appeared to be disruptive for CM and possibly BLB, but not for NR. Similar, but less significant effects are seen in the parafoveal learning sequences (not shown), consistent with the smaller size of the difference between *Twin* and *Fixed* conditions in the parafovea.

4. Discussion

Ahumada & Beard[7] found lower contrast detection thresholds for an aircraft target in a single *Fixed* noise as compared with *Twin* random noise. Here we extend this finding of a *Fixed*/*Twin* threshold difference to multiple samples of *Fixed* noise, to Gabor and checkerboard stimuli, and to parafoveal detection. One explanation is that an internal sensory representation of each input image is compared with templates stored in memory and that observers modify the templates as the experiment progresses. We hypothesize that the *Fixed*/*Twin* effect is the result of template learning. A *Fixed* noise stimulus would be incorporated in the template and thus help to reduce target position uncertainty. *Twin* noise samples, which change from trial-to-trial, would not provide this position information and might contribute noise to the templates.

Because Watson *et al*.[12] did not find a *Fixed* noise advantage over *Twin* noise for a Gabor target we initially hypothesized that an inherent position uncertainty of the Gabor stimulus may slow learning the *Fixed* noise components, thus removing any potential advantage of the *Fixed* noise display. Our simulation results (see Fig. 1) lent support to this hypothesis. However, our experimental results did show a *Fixed*/*Twin* noise effect. It is likely that differences in experimental parameters between Watson *et al*. and our current study can explain this discrepancy. For example, Watson *et al*. used noise pixel sizes that were considerably smaller than those used here which would make learning the local *Fixed* noise pixels a more difficult task.

We found that, overall, performance was better for a Gabor target than for a checkerboard. Image discrimination models[1-7] predict this effect as a result of lower sensitivity to the diagonal frequency components of the checkerboard and imperfect summation over different channels. Neither of these features are yet available in template matching models that have been used to predict target detectability in noise.[8-10]

Since accuracy of positional localization in peripheral vision is poor,[13] we included conditions with the checkerboard and Gabor targets presented 4 deg parafoveally. We hypothesized that increased positional uncertainty with eccentricity would decrease the *Fixed*/*Twin* difference. In one condition, the target size was the same as in the fovea (unmagnified). In the other, the target was magnified by a factor of 3 to account for the progressive reduction in cortical area with increasing eccentricity.[16] Confirming our hypothesis, the *Fixed*/*Twin* difference was larger in the fovea.

In the unmagnified 4 deg eccentricity condition, detection thresholds in noise increased with eccentricity as they have been shown to do in the presence of a pattern mask,[14] whereas, for the Gabor stimulus in noise, thresholds were significantly reduced with eccentricity. We can only speculate that the Gabor, whose form is based only on low frequency variations, shows more release from masking by high frequency components.[19]

In the magnified eccentricity condition, the checkerboard stimuli again behaved in an expected fashion. The contrast (not contrast energy) thresholds in the periphery were similar to those in the fovea, supporting the use of the scaling factor obtained for detection without noise.[16] This scaling factor was not required for the Gabor stimulus because performance improved without stimulus magnification.

We hypothesize that the *Fixed*/*Twin* effect is the result of template learning. The initial state of the memory templates should depend on the experimenter's description of the experiment, the initial stimulus presentation, prior experience, and other factors and is not addressed here. Our data does provide information about memory template updating. In some cases *Fixed* noise thresholds were substantially below *Twin* noise thresholds on the first trial block suggesting that template updating can be very rapid. Rapid template adjustment may explain the fast perceptual learning described elsewhere.[20,21] Some *Fixed* noise conditions also show gradual improvement over the course of five to six practice blocks, which may reflect template refinement.

Data showing positive and negative transfer of training effects may be seen in Fig. 3. The data of observer NR show low thresholds for the second noise sample which appear to be a continuation of the learning curve or a plausible extension of it from the first noise sample. The transfer of improvement suggests some of the learned relevant features of the target

plus noise sample are valid in both situations. In other words, the best-matched template to the new sensory representation is similar to the just-refined template, and therefore requires little further updating. In other cases, there seems to be negative transfer of improvement,[22-24] where thresholds are higher for a new noise sample. Negative transfer suggests that the original relevant task features are also applied to the new noise sample when these features are no longer valid or useful. A final characteristic that can be taken from these data is that low thresholds observed for almost all *Fixed* noise samples suggests that the observer is able to hold multiple templates for the same task.

4a. Template Learning Model with Uncertainty

Fig. 4 presents a schematic of a simple image detection model with template learning and positional uncertainty. A model without position uncertainty is presented in the Appendix. Gray-shaded boxes represent functions that have yet to be included in the working mathematical model. An input image plus added noise enters the visual system. An internal sensory representation is formed of the stimulus, which includes internal noise sources. This noisy sensory representation is correlated with memory templates, a decision is made as to which stimulus was present, feedback is given, and the templates are updated.
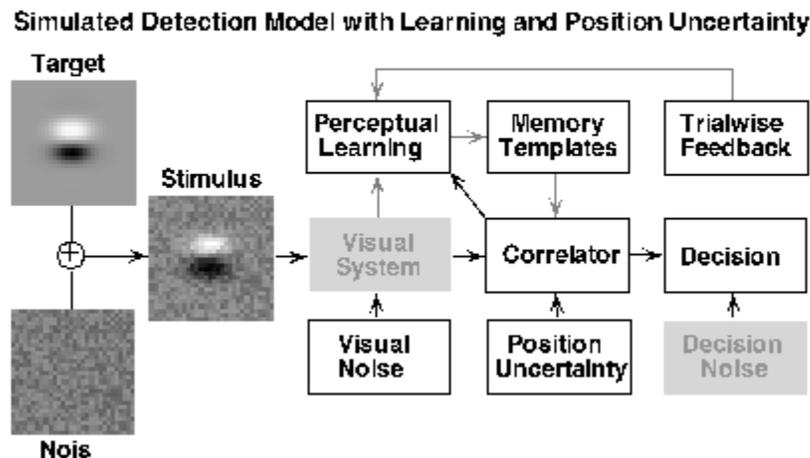

Fig. 4. Schematic of an image detection model with template learning and positional uncertainty. An input image plus added noise enters the visual system, where an internal sensory representation is formed of the stimulus. This noisy sensory representation is correlated with memory templates of the targets over a range of positions, and a decision is made as to which target was present. Based on the trialwise feedback, the template is updated. The visual system module and the decision noise module are in dashed boxes because they are not included in our current models.

4a.1 Response generation rule

In our simplified modeling environment, the two 4 x 4 pixel targets were a simple 2 x 2 checkerboard and (stretching the term) a Gabor stimulus consisting of a 2 x 4 light bar above a 2 x 4 dark bar. The targets were centered in a 6 x 6 array of white noise. On each simulated two-interval trial there are two input images, one contains the target stimulus plus noise and other contains noise alone. The internal sensory representation (I) is composed of the input image plus internal sensory noise (see Fig. 5). For each interval, the internal sensory representation is cross-correlated with the 4 x 4 memory template of the target at all 9 possible positions. The largest cross-correlation is compared with the largest cross-correlation for the no-target template and the difference is saved for comparison with the same difference from the other interval. The larger of these differences is used to select the interval with the signal.
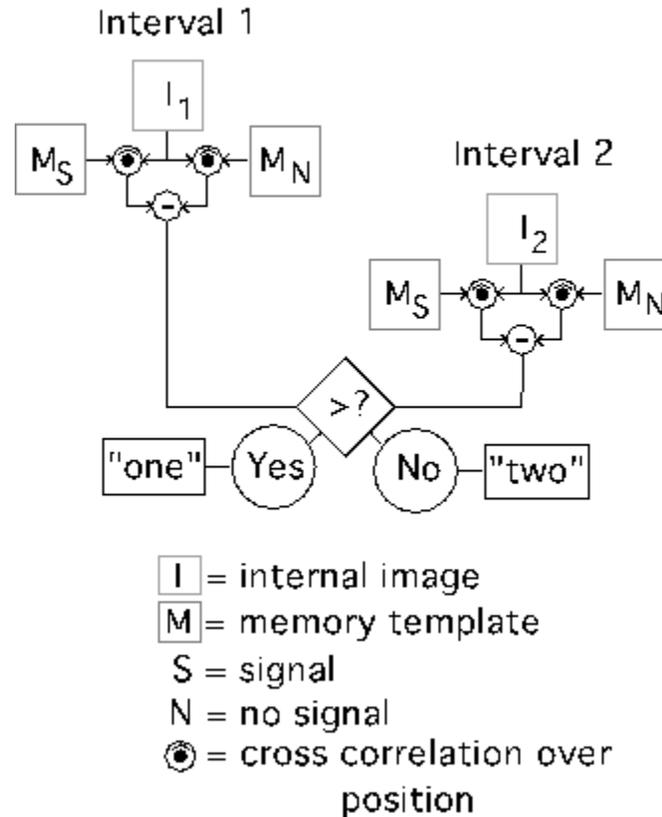
Response Generation Rule

Fig. 5. Template learning model generates a response indicating which input image contains the target. For each image the model forms an internal sensory representation (I) composed of the image plus internal sensory noise. This representation is cross correlated with memory templates for the target (M_s) and no target (M_N). The careted dot symbol indicates the maximum of correlations over a range of positions. The difference of the two maximum correlations represents the signal-presence likelihood for each interval, and a comparison of these likelihoods determines the response.

4a.2 Learning rule

The memory templates are modified on the basis of feedback indicating the interval containing the signal (see Fig. 6). Each template is replaced by a weighted average of the old template and the internal sensory representation of the appropriate image positioned by the cross-correlation. The relative weighting of the average is determined by l, the learning rate parameter, which is near one if learning is rapid and near zero if learning is slow. The initial templates in our model were ideal, the asymptotic templates learned with arbitrarily small learning rates. For *Fixed* conditions the initial templates included the external noise sample, while for *Twin* conditions the initial templates were the target and a zero image. For *Fixed* noise, the fixed external component of the template noise helps lock-in on the signal position. For *Twin* noise, however, the external noise is changing on each trail and effectively adds random noise to the templates whenever the target and no-target templates correlate best at different positions.

Learning Rule
(correct feedback)

λ = learning rate $(0 \le \lambda \le 1)$

I_S = internal stimulus image
I_N = internal noise image
M_S = stimulus template
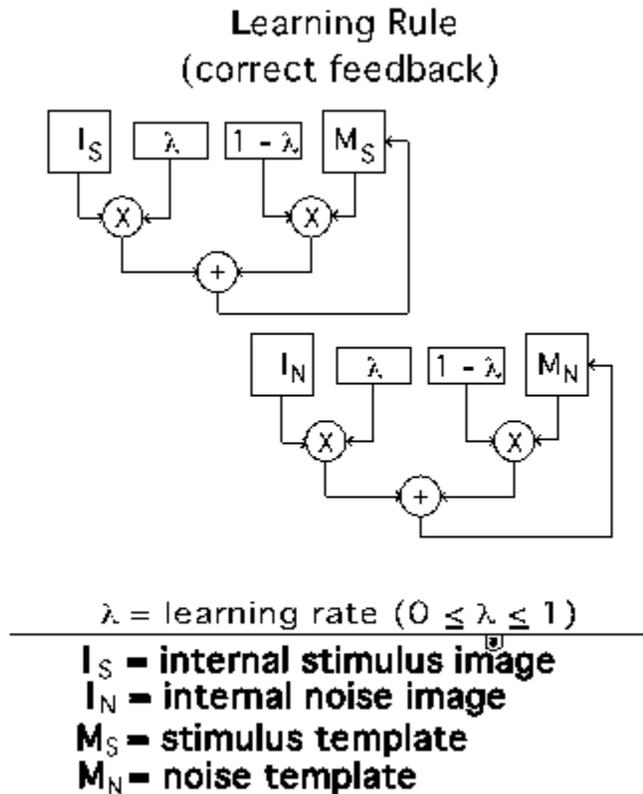M_N = noise template

Fig. 6. Learning rule. Correct feedback assigns the internal images to the appropriate memory template, which is assumed to be translated to the best-correlating position. Each template is then replaced by a weighted average of the old template, and the associated internal sensory representation. If l, the learning rate parameter (a number between 0 and 1), is large, the average contains mostly the current internal image; if it is small, the template is only slightly changed.

To see the role of learning rate and the ratio of internal to external noise on the *Fixed/Twin* effect we estimated the 79% correct threshold level for the two conditions. Each threshold was estimated from six repetitions of 400 trials at 4 signal levels. Fig. 7 shows the difference (*Twin-Fixed)* for the two target patterns, checkerboard and "Gabor". The abscissa is the ratio of the internal noise standard deviation to the external noise standard deviation. The parameter is the learning rate. The horizontal lines indicate the average psychophysical values and the associated confidence interval from our data. The largest *Fixed/Twin* noise difference occurred for the smaller internal noise level since the external noise carries the effect. Of the conditions we simulated, the best matching was for l = 0.0 and 0.1 and internal noise level = 1. The fair fit of the model with a learning rate of zero does not mean that learning is not needed to predict the result, since the initial templates were set to reflect learning of the *Fixed* noise sample. In a separate simulation, we also estimated the *Fixed/Twin* threshold difference with the learning rate set to zero and the initial templates set to the target and zero noise images for both conditions. In this case, the sign of the difference actually reversed slightly.
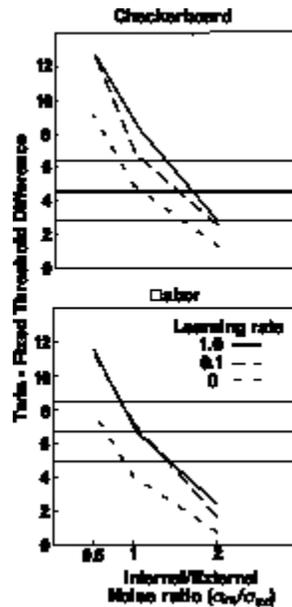
Fig. 7. Model simulations of the fixed/twin effect. Proportion correct scores from six 400-trial repetitions at four signal levels were converted by regression in the d' domain to 79% correct threshold estimates for each of 36 conditions [fixed-twin, checkerboard, and Gabor target, three internal-to-external noise ratios (0.5,1.1), and three learning rates (0,0.1,1)]. The graphs show the amount that the twin threshold exceeded the fixed threshold (in dB) as a function of the other variables. The horizontal lines show the average foveal difference for our observers and the 95% confidence interval based on variability over observers. Parameters: l = 0.0, 0.0, and internal-to-external noise ratio equals 1 best fit to the target results.

Template matching models[8,9] which correlate sensory representations of the images with one or more memory templates, can predict target detectability in *Random* noise masking conditions, where a different noise sample is added to each input image. If the same template is used in *Fixed* and *Twin* noise conditions, they do not, however predict a *Fixed*/*Twin* noise masking difference. Image discrimination models cannot predict either a *Fixed*/*Twin* difference or *Random* noise effects (since the difference calculation would include not only the target, but also any changes in the noise). Our template learning model presented in the Discussion section demonstrates that combining template learning with positional uncertainty can explain the *Fixed*/*Twin* noise threshold difference. The model presented in the Appendix shows how template learning alone could explain the *Fixed*/*Twin* effect, and how template learning contributes to the accelerating nonlinearity associated with uncertainty and transducer functions.[25,26]

5. Acknowledgments

-

6. References

[1] A. B. Watson, "Efficiency of a model human image code," J. Opt. Soc. Am. A **4**, 2401-2417 (1987).

2 S. Daly, "The visible difference predictor: An algorithm for the assessment of image fidelity," in *Digital Images and Human Vision*, A. B. Watson, ed., MIT Press, Cambridge, MA. (1993).

3 J. Lubin, "The use of psychophysical data and models in the analysis of display system performance," in *Digital Images and Human Vision*, A. B. Watson, ed., MIT Press, Cambridge, MA (1993).

4 P. C. Teo, D. J. Heeger, "Perceptual image distortion," in *Human Vision, Visual Processing, and Digital Display*, B. Rogowitz and J. Allebach, eds., Proc. SPIE **2179**, 127-141 (1994).

5 H. R. Wilson, "Quantitative models for pattern detection and discrimination," in *Vision Models for Target Detection and Recognition*, E. Peli, World Scientific Publishing, New Jersey, (1995).

6 A. B. Watson, J. A. Solomon, "A model of visual contrast gain control and pattern masking," J. Opt. Soc. Am. A **14**, 2379-2391 (1997).

7 A. J. Ahumada, B. L. Beard, "Image discrimination models predict detection in fixed but not random noise," J. Opt. Soc. Am. A **14**, 2471-2476 (1997).

8 H. Barrett, "Evaluation of image quality through linear discriminate models," SID Proceedings 871-873 (1992).

9 A. Burgess, "Statistically defined backgrounds: Performance of a modified non-pre whitening observer," J. Opt. Soc. Am. A **11**, 1237-1242 (1994).

[illegible small text]

-

7. Appendix: A Template Learning Model without Positional Uncertainty

This appendix describes another template learning model for two-alternative forced-choice (2AFC) detection experiments. The main difference between this model and the model presented in the Discussion is that here no position uncertainty is assumed in the model. This simplification allows closed form expressions for the performance of the model and clarifies the role of the different parameters. If we simply remove position uncertainty from the model in the Discussion section, there is no *Fixed/Twin* difference because the same external noise sample is accumulated equally in each template and they cancel in the response calculation. Here we assume that the observer remembers only the image from the second interval and updates the template that feedback associates with the second interval. The performance formulas show an accelerating nonlinearity of performance with signal level that is generated by the poor quality of the templates at low signal levels.

A.1 Response generation rules

For each stimulus presentation, the observer is assumed to have an internal sensory representation vector, I, which is the sum of three vectors, (i) an internal representation of the signal (if present), (ii) an internal representation of the external noise, and (iii) internal noise,

$I = S + N_{ext} + N_{int}$. (4)

The observer is assumed to have internal sensory representations of both stimuli. M_N represents the target-absent sensory representation and M_{SN}, the target-present sensory representation. For a 2AFC experiment, the observer is assumed to respond "interval 1" if

$(M_{SN}\text{-}M_N) * I_1 > (M_{SN}\text{-}M_N) * I_2$, (5)

where the dot symbol "$*$" indicates inner product This expression can be rewritten as

$(M_{SN}\text{-}M_N) * (I_1\text{-}I_2) > 0$. (6)

A.2 Template learning rule

The template adjustment process replaces the template with an average of itself and the internal sensory representation of the stimulus, using a learning rate parameter, l,

$M <= (1 - l)M + l\, I$. (7)

Ideally, l should be a function of many things, such as the similarity of M and I, the feedback or reinforcement on the trial, and the experience of the observer.[27] Template learning processes must be able to proceed without feedback and likely depend on the stimuli in both intervals. It seems reasonable, however that the strongest influence would be the most recent stimulus (i.e., interval 2) and that feedback does play a role. For simplicity and to accentuate the process that differentiates the *Fixed* and *Twin* noise conditions, we update the template according to the rule (assuming correct feedback),

$M_N <= (1 - l)M_N + l\, I_2$, if I_2 is a target-absent interval, (8)

$M_{SN} <= (1 - l)M_{SN} + l\, I_2$, if I_2 is a target-present interval. (9)

A.3 Asymptotic behavior of the model

The performance of the model is controlled by the quality of the templates. The updated template is a weighted average of the stimulus plus the internal noise on all the trials on which it was adjusted. If we let M_0 be the initial value of one of the templates, and Ii be the image on the ith updated trial, then the template M_n at the end of the nth trial is given by

$$M_n = (1 - l)^n M_0 + l (1 - l)^{(n-1)} I_1 + ... + I_n l. \quad (10)$$

Let F be the fixed part of the stimulus on the ith update of this template and R_i be the random part of the stimulus, then as n becomes large, the contribution from M_0 becomes negligible and the template approaches

$$F + S R_{n-i} l (1 - l)^i, \quad (11)$$

where the summation index i ranges from 0 to n.

Since the R_i are independent, with mean 0 and covariance matrix S_R, the asymptotic template distribution has a mean of F and a co-variance matrix S_M of

$$S_M = S_R l^2 / (1-(1- l)^2)$$

$$= S_R l / (2- l).$$

$$= S_R / n_l. \quad (12)$$

The quantity $n_l = (2 - l) / l$ can be regarded as the effective number of noise components averaged for a given l. For $l = 1$, it is 1, for $l = 0.5$, it is 3.

A.4 Asymptotic detection performance of the model

Let the difference of the templates be $D_M = M_{SN} - M_S$, and the difference of the stimulus representation be $D_S = I_1 - I_2$, the response rule of Equation 6 becomes, respond "interval 1" if

$$D_M * D_S > 0. \quad (13)$$

Because the randomness in D_M comes from past trials, D_M and D_S are independent random vectors. The mean of D_M is S. The mean of D_S is S when the signal is interval 1 and -S when it is in interval 2. The mean of $D_M.D_S$ is E = S.S, when the signal is in the first interval and -E when it is in the second. The performance of the model can be characterized by

$$d = (E - (-E))/SD[DM.DS] = 2 E/SD[DM.DS], \quad (14)$$

where SD[] indicates the standard deviation. If the individual components of the representations are independent and have the same variance, the variance of $D_M.D_S$ can be expressed as

$$SD[D_M.D_S]^2 = E (s_M^2 + s_S^2) + n s_M^2 s_S^2 \quad (15)$$

where the ss are the standard deviations of the individual components and n is their number.

For the *Twin* and *Fixed* noise cases, the signal difference pixel variance is given by

$$s_S^2 = 2 s_{int}^2. \quad (16)$$

For the *Fixed* noise case, the template difference pixel variance is given by

$$s_M^2 = 2 (s_{int}^2) / n_l, \quad (17)$$

while for the *Twin* noise case, it is

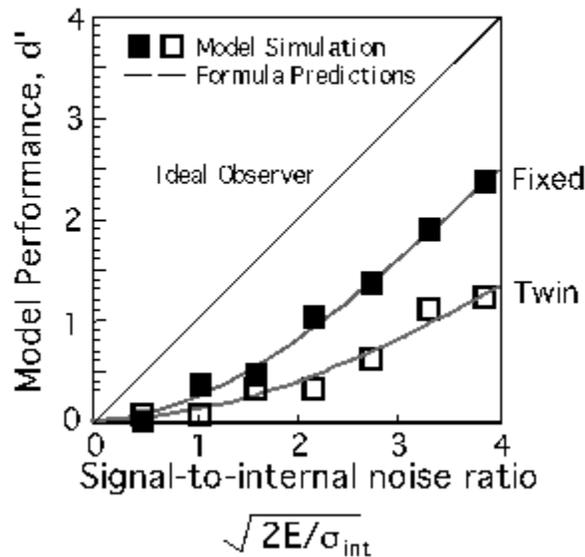$$s_M^2 = 2 (s_{int}^2 + s_{ext}^2) / n_l. \text{ (18)}$$

Fig. 8. Mathematical formula predictions and simulation results for the template learning model with no positional uncertainty. Performance in twin and fixed noise conditions is plotted as a function of the signal-to-noise level. The ideal observer with internal noise would perform along the main diagonal. The simulation proportions of correct responses were transformed to d' by means of the cumulative Gaussian distribution, which seems to fit well. Note that the model predicts an accelerating nonlinearity in the absence of uncertainty or a transducer function.

Fig. 8 plots the performance of the *Twin* and *Fixed* models as a function of d' for the ideal observer, limited only by the internal noise,

$$d_{Ideal} = Sqrt[2 E / s_{int}^2]. \text{ (19)}$$

As the learning parameter l approaches zero, n_l becomes infinite, and the performance for both conditions becomes ideal (although it would take infinite time to reach this level). The higher curve below the unity slope line shows predicted model performance in the *Fixed* noise condition with the parameters n = 16 and l = 0.5. The lower curve shows the *Twin* noise prediction for those same parameters and $s_{ext} = 2 s_{int}$.

Fig. 8 illustrates another feature of the template learning model. In addition to predicting a *Fixed/Twin* threshold difference, it also quantitatively predicts an accelerating nonlinearity of performance with signal level generated by the poor quality of the templates at low signal levels.[28] Tanner inferred this effect when he failed to find a close correspondence between observer behavior and a stimulus uncertainty model as signal strength increased.[28] Template learning now joins signal uncertainty [25] and nonlinear [26] transducer as a quantified process for explaining non-linear performance for low signal-to-noise ratios.