

# Classification images for detection, contrast discrimination, and identification tasks with a common ideal observer

**Craig K. Abbey**

Department of Psychology, University of California, Santa Barbara, CA, USA, & Department of Biomedical Engineering, University of California, Davis, CA, USA



**Miguel P. Eckstein**

Department of Psychology, University of California, Santa Barbara, CA, USA



We consider three simple forced-choice visual tasks—detection, contrast discrimination, and identification—in Gaussian white noise. The three tasks are designed so that the difference signal in all three cases is the same difference-of-Gaussians (DOG) profile. The distribution of the image noise implies that the ideal observer uses the same DOG filter to perform all three tasks. But do human observers also use the same visual strategy to perform these tasks? We use classification image analysis to evaluate the visual strategies of human observers. We find significantly different subject classification images across the three tasks. The domain of greatest variability appears to be low spatial frequencies [ $<5$  cycles per degree (cpd)]. In this range, we find frequency enhancement in the detection task, and frequency suppression and reversal in the contrast discrimination task. In the identification task, subject classification images agree reasonably well with the ideal observer filter. We evaluate the effect of nonlinear transducers and intrinsic spatial uncertainty to explain divergence from the ideal observer found in detection and contrast discrimination tasks.

Keywords: classification image, contrast discrimination, detection, identification, spatial vision, visual mechanisms

## Introduction

Much of the research in spatial pattern vision consists of determining the mechanisms or features that observers use to perform visual tasks. From these mechanisms, and often with insights from physiological data, fundamental workings of the visual system are elucidated and then implemented in models of higher order visual function such as object recognition (Brady & Kersten, 2003; Tjan & Legge, 1998), search (Beutter, Eckstein, & Stone, 2003; Navalpakkam & Itti, 2005), and reading (Pelli, Farell, & Moore, 2003). One example of this process has been the discovery of spatial-frequency-selective channels in the human visual system (Campbell & Robson, 1968; Graham & Nachmias, 1971; Mostafavi & Sakrison, 1976; Sachs, Nachmias, & Robson, 1971; Stromeyer & Klein, 1974; among others). Investigators utilized the results of masking and adaptation studies along with cell recordings to discover the existence of spatial-frequency-selective filters (i.e., channels) that mediate observer performance in visual tasks.

The Bayesian ideal observer is an important tool in the study of visual mechanisms because it defines an optimal mechanism for performing a given task. To utilize the ideal observer, there must be some random component to performing the task. This stochastic requirement can be fulfilled by postulating sources of internal noise arising

within the observer and/or by adding external noise to visual stimuli. Typically, performance of human observers is measured relative to the performance of the ideal observer and reported as statistical efficiency (Burgess, Wagner, Jennings, & Barlow, 1981; Pelli, 1981; Tanner & Birdsall, 1958). Low efficiency indicates suboptimal mechanisms for a given task and/or a large source of internal noise within the observer.

However, efficiency is not generally revealing of mechanism. Put in another way, there may be many visual mechanisms that will yield the same level of suboptimal statistical efficiency. Burgess and Colborne (1988) and others have been able to determine the level of internal noise through the use of two-pass agreement studies, but this still does not resolve the problem of different mechanisms with equivalent performance levels. Solomon and Pelli (1994) have analyzed visual mechanisms in letter identification using threshold elevations in band-pass noise to map out the profile of a mediating filter. However, it is still not entirely clear how an observer applies such a filter on a stimulus-by-stimulus basis to actually identify letters. Furthermore, results from simpler detection and discrimination tasks that demonstrate significant off-frequency looking effects (Abbey & Eckstein, 2000; Burgess, Li, & Abbey, 1997; Solomon, 2000) indicate that some assumptions necessary for the Solomon and Pelli approach may not generalize to all visual tasks.

Given the limitations of these alternatives, the classification image approach developed by Ahumada et al. (Ahumada, 2002; Ahumada & Lovell, 1971; Ahumada, Marken, & Sandusky, 1975; Beard & Ahumada, 1998) appears to be a useful way to advance our understanding of basic visual mechanisms. The fundamental idea of the classification image approach is to use stimulus noise fields along with the corresponding observer decisions to obtain an image of the spatial weighting used by a subject to perform the task. This connection between the classification image and the subject's internal mechanism is particularly clear if the subject uses a linear template to perform the task. In this case, the classification image is directly related to the weights incorporated in the template. However, it is still possible to generate a linear classification image from a nonlinear model of visual processing. This allows the testing of nonlinear models to see if they predict the linear classification images obtained from subjects.

In this paper, we apply classification image analysis to investigate how visual processing changes in the context of different visual tasks. Extensive previous work in this area has focused on differences between detection and discrimination using combinations of signals and pedestal functions (for example, Foley, 1994; Foley & Legge, 1981; Legge & Foley, 1980; Thomas & Olzak, 1996; Watson & Solomon, 1997; Wilson, McFarlane, & Phillips, 1983). Various contrast pedestals have been shown to facilitate or inhibit task performance over detection in the absence of a pedestal. One common finding in many situations is that contrast thresholds will drop somewhat in going from detection to contrast discrimination with a low-contrast pedestal before rising again as the pedestal contrast moves into the suprathreshold regime. This threshold profile is sometimes referred to as the dipper function (Chen & Foley, 2004). Explanations of the dipper function often incorporate nonlinearities such as nonlinear signal transduction or spatial uncertainty (Foley, 1994; Legge, Kersten, & Burgess, 1987; Pelli, 1985).

We consider visual detection, contrast discrimination, and identification tasks masked by Gaussian white luminance noise using the two-alternative forced-choice (2AFC) paradigm. Our tasks are designed so that the difference signal—the difference between the mean target and mean alternative—has the same difference-of-Gaussians (DOG) profile in all three tasks, and hence the defining characteristics of each task are the pedestal to which the signal is added and the contrast of the DOG signal. The white noise distribution of the image noise fields ensures that the ideal observer uses this DOG filter—and therefore a consistent mechanism—to perform all three tasks.

But does this consistency hold for human observers? We investigate the consistency of features used by human observers in these different tasks from analysis of observer efficiency with respect to the ideal observer and by

classification image analysis. Because the optimal feature for performing the task is held constant across tasks, to the extent that observers can optimize their visual strategy in a given task, they will tend to converge on the same visual feature. Thus, residual differences between classification images reflect task- and subject-specific visual processing. As we shall see below, human observers are not using a consistent mechanism in the data we have collected. As possible explanations for this inconsistency, we also explore a number of global nonlinearities to see if they can reconcile the differences we find across tasks. The approach here is to collect a classification image from a nonlinear model of visual processing and compare it to the classification image obtained from human observers. We examine the classification images derived from early and late nonlinear transducer models as well as a spatial uncertainty model.

## Methods

### Experimental stimuli

In each trial of a 2AFC experiment, an observer is shown two images and asked to identify the image representing the target. Figure 1 shows the mean (noiseless) target and alternative profiles for the experiments reported here, which can be described as follows. For the detection task, we chose a mean target luminance field with a DOG signal profile. In this case, the mean alternative profile is a uniform flat field, and hence the difference signal is simply the DOG profile. For the contrast discrimination task, the observer had to discriminate a high-contrast DOG from one with slightly lower contrast. This also yields a difference signal that has the same DOG profile. For the identification task, the goal was to classify two Gaussian luminance profiles. The parameters of the two Gaussians (amplitudes and spatial standard deviations) were set so that the difference between them assumed the same DOG profile as the detection and contrast discrimination tasks. We refer to this as an identification task because it is equivalent to identifying which of the two Gaussian profiles is present at a given location.

In all cases, both target and alternative were rotationally symmetric (to the scale of display pixels), and hence well described by a radial profile. An attractive feature of a DOG difference signal is that it assumes a band-pass profile in spatial-frequency domain. By adjusting the parameters of the DOG, it is possible to tune the signal to spatial frequencies of interest. The frequency spectrum of the difference signal is plotted in Figure 1D. Under the experimental conditions used for obtaining observer data (described below), the peak spectral intensity was at

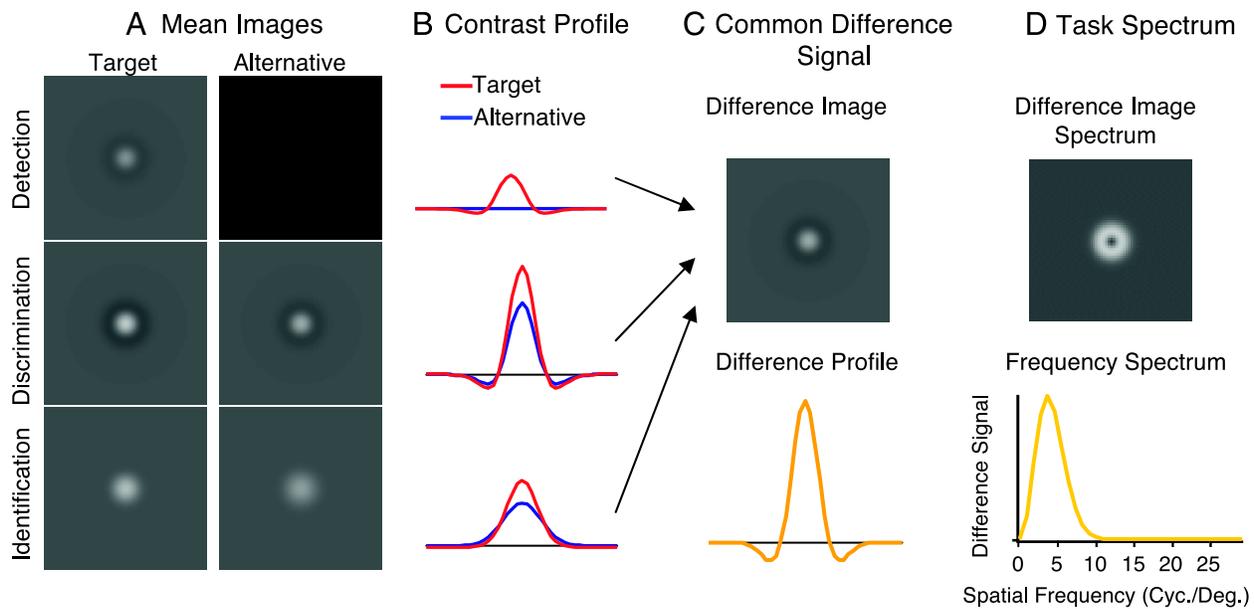


Figure 1. A graphical depiction of the three tasks considered in this work. The left side of the figure (A) shows the mean (noiseless) target and alternative images for each task (at enhanced contrast for visualization). To the right of these (B) are contrast profiles through the center of the target (red) and alternative (blue) images. All three target and alternatives share a common difference signal (C), which is shown both as an image and a central profile. The spatial-frequency spectrum of the difference image (D) is seen to possess a ring of frequency content that peaks at approximately 4 cpd of visual angle.

approximately 4 cycles per degree (cpd) of visual angle, with a bandwidth (full-width at half-max) of approximately 1.8 octaves. This peak spectral intensity was chosen to be roughly in the area of peak contrast sensitivity of the human visual system (Hood & Finkelstein, 1986; van Nes, Koenderink, Nas, & Bouman, 1967). Example target and alternative stimuli are shown in Figure 2.

For the purpose of explanation, we will consider an image to be a column vector with the number of elements equal to the number of pixels in the image. We will refer to the noisy signal-present (target) image in the  $j$ th trial by  $\mathbf{g}_j^+$ , and the signal-absent (alternative) image by  $\mathbf{g}_j^-$ . The

trial index,  $j$ , runs from 1 to the number of trials,  $N_T$ . These images are defined by

$$\begin{aligned} \mathbf{g}_j^+ &= \mathbf{b} + \mathbf{n}_j^+ + \mathbf{s} \\ \mathbf{g}_j^- &= \mathbf{b} + \mathbf{n}_j^- \end{aligned} \quad (1)$$

where  $\mathbf{b}$  is the task-dependent mean background intensity—including any common signal pedestal,  $\mathbf{s}$  is the difference signal, and  $\mathbf{n}_j^+$  and  $\mathbf{n}_j^-$  are noise fields associated with each alternative. The profile of the difference signal is held constant across tasks although the contrast of the signal was adjusted from the results of pilot studies to achieve targeted levels of task performance. We utilize the method of constant stimuli, so the signal,  $\mathbf{s}$ , is unchanging throughout an experiment. Random number generators are used to create uncorrelated (white) Gaussian luminance noise with a pixel standard deviation of  $\sigma_n$ . The noise fields in signal-present and signal-absent alternatives are independent of each other and independent across experimental trials as well.

### Forced-choice experiments

In each trial of a forced-choice experiment, an observer gets a score of 1 for correctly identifying the target image ( $\mathbf{g}_j^+$ ), and 0 for incorrectly identifying the alternative image ( $\mathbf{g}_j^-$ ) as the target. We will refer to the trial score by the variable  $o_j$  (indicating the outcome of trial  $j$ ). The proportion of correct responses ( $P_C$ ) is defined as the expected value of  $o_j$ ,  $P_C = E(o_j)$ , where  $E$  indicates the

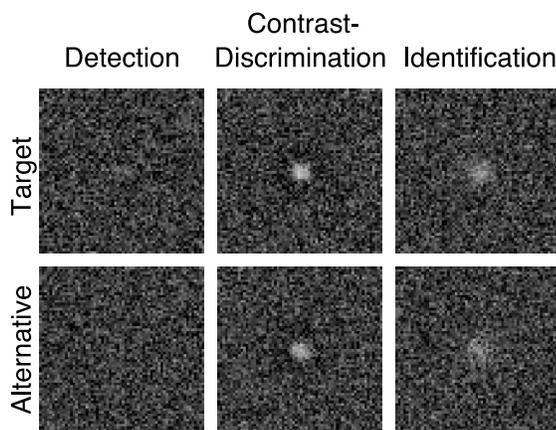


Figure 2. Example stimuli used in the experiments. Note that the target and alternative contrasts are higher than actual experiment settings for clarity of presentation.

mathematical expectation of a random variable. In 2AFC experiments, we estimate this quantity with the sample average across trials

$$\bar{P}_C = \frac{1}{N_T} \sum_{j=1}^{N_T} o_j. \quad (2)$$

Proportion correct can be converted to the detectability index,  $d'$ , by the formula (Green & Swets, 1966),

$$d' = \sqrt{2} \Phi^{-1}(\bar{P}_C). \quad (3)$$

The observer efficiency with respect to the ideal observer in a given task is determined by the squared ratio of the detectability index of the observer to the detectability index of the ideal observer (described below) in the same task, usually multiplied by 100 to give a percentage (Tanner & Birdsall, 1958),

$$\text{Efficiency} = 100\% \times \left( \frac{d'_{\text{observer}}}{d'_{\text{ideal observer}}} \right)^2. \quad (4)$$

## Linear models of visual processing

For the purposes of modeling the decision process, we can define the observer score in terms of a scalar-valued internal response function,  $\lambda$ , as

$$o_j = \text{Step}[\lambda(\mathbf{g}_j^+) - \lambda(\mathbf{g}_j^-)], \quad (5)$$

where the Step function gives a value of 1 for arguments greater than 0, and a value of 0 for all others. Equation 5 implies that if the internal response to the signal-present image is larger than that of the signal-absent image, then the observer makes a correct decision. If the internal response of the signal-absent image is greater, an incorrect decision is made.

For a linear observer, the internal response function is presumed to be a (noisy) linear function of the image defined as

$$\lambda(\mathbf{g}) = \mathbf{w}^t \mathbf{g} + \varepsilon, \quad (6)$$

where  $\mathbf{w}$  is a vector of spatial weights and  $\varepsilon$  is a stochastic internal noise component. The internal noise component is assumed to be a Gaussian-distributed random variable that is independent of  $\mathbf{g}$  with a mean of 0 and standard deviation,  $\sigma_\varepsilon$ . The weighting vector,  $\mathbf{w}$ , governs how the spatial distribution of intensity in the image influences the

observer's response. As such, it encodes the visual strategy used by the observer. Under this linear model, the detectability index for the images defined in Equation 1 is

$$d' = \frac{\mathbf{w}^t \mathbf{s}}{\sqrt{\sigma_n^2 \|\mathbf{w}\|^2 + \sigma_\varepsilon^2}}. \quad (7)$$

### Ideal observer

We will make extensive use of the ideal observer as a benchmark for comparison with human observers. For two-class classification tasks in Gaussian white noise, it is well known that cross-correlation with the difference signal is equivalent to the ideal observer decision strategy (for example, Geisler, 2003). In terms of Equation 6, the ideal observer is realized by setting  $\mathbf{w} = \mathbf{s}$ , and  $\sigma_\varepsilon = 0$ .

The difference signal is identical in the tasks described above. Therefore, the ideal observer uses the same linear filter to perform all three tasks used in this work. Such strategies are ideal for study by classification image analysis because the resulting classification images are directly related to the linear filter weights used by the observer (Abbey & Eckstein, 2002a). In all three tasks, there is no incentive derived from the stimuli for the observer to change visual strategies or adopt a nonlinear strategy.

## Classification image analysis

The basic idea behind classification image estimation in 2AFC experiments is that the presence of noise in the images influences the probability of making a correct decision by propagating through the visual pathways to the features used to formulate a decision variable. Hence, by averaging the noise fields of trials associated with correct and incorrect decisions, it is possible to obtain an image revealing components of the observer's decision function and visual strategy (Beard & Ahumada, 1998). A fundamental assumption of the methodology is that observers do not change their visual features substantially in the presence of noisy stimuli.

Estimation of a classification image in 2AFC tasks has been described using a weighted difference in noise fields (Abbey & Eckstein, 2002a),

$$\Delta \mathbf{q}_j = \frac{N_T}{(N_T - 1) \sigma_n^2} (o_j - \bar{P}_C) \Delta \mathbf{n}_j, \quad (8)$$

where  $\Delta \mathbf{n}_j = \mathbf{n}_j^+ - \mathbf{n}_j^-$ , and  $N_T$  is the number of trials. For experiments with  $P_C$  values near 0.85, the  $o_j - \bar{P}_C$  term will assign a relatively small positive value to trials in which a correct response is given and a relatively large

negative value to trials in which an incorrect response is given. Under the assumption of the linear model in Equation 6, this weighting scheme can be shown to be nearly optimal (Abbey & Eckstein, 2002b). Assuming the linear observer response function given in Equation 6, the expectation of  $\Delta\mathbf{q}_j$  is directly related to the template (Abbey & Eckstein, 2002a) by

$$E(\Delta\mathbf{q}_j) = \frac{1}{\sqrt{\pi(\sigma_n^2 \mathbf{w}^t \mathbf{w} + \sigma_\varepsilon^2)}} e^{-(d'/2)^2 \mathbf{w}}, \quad (9)$$

where  $d'$  is the detectability index defined in Equation 7. To estimate the classification image, we use the sample average across 2AFC trials instead of the expectation in Equation 9 to obtain

$$\Delta\bar{\mathbf{q}} = \frac{1}{N_T} \sum_{j=1}^{N_T} \Delta\mathbf{q}_j. \quad (10)$$

We will refer to  $\Delta\bar{\mathbf{q}}$  as the estimated classification image.

### Radial averages

While the estimation procedure in Equation 10 is suitable for estimating the entire classification image, it is often desirable to restrict attention to regions or features in the classification image and to employ averaging across elements of the classification image as a way to reduce measurement noise that arises from a finite number of trials. These constraints can be particularly valuable for conducting statistical hypothesis testing (Abbey & Eckstein, 2002; Solomon, 2002) because they reduce degrees of freedom, which results in more powerful hypothesis tests provided the averaging does not cancel important features of the data.

As we shall see below, in this work we find it convenient to look at radial averages of classification images as depicted in Figure 3. Radial averaging involves no substantial loss of information if the observer template is radially symmetric and centered on the target. The rationale for making this assumption is that the difference signal and pedestal profiles are all radially symmetric, and the noise is isotropic. Hence, there is no preferred orientation in the stimuli. Also, in previous experiments similar to these (Abbey & Eckstein, 2002a; Abbey, Eckstein, & Bochud, 1999; Solomon, 2002), we have not observed classification images with strong orientation features.

The radial averages reported below are computed by calculating each pixel's distance from the center of the image and rounding to the nearest integer, which can be thought of as a radial binning operation. All pixels with a common bin value are averaged together. The first bin, at

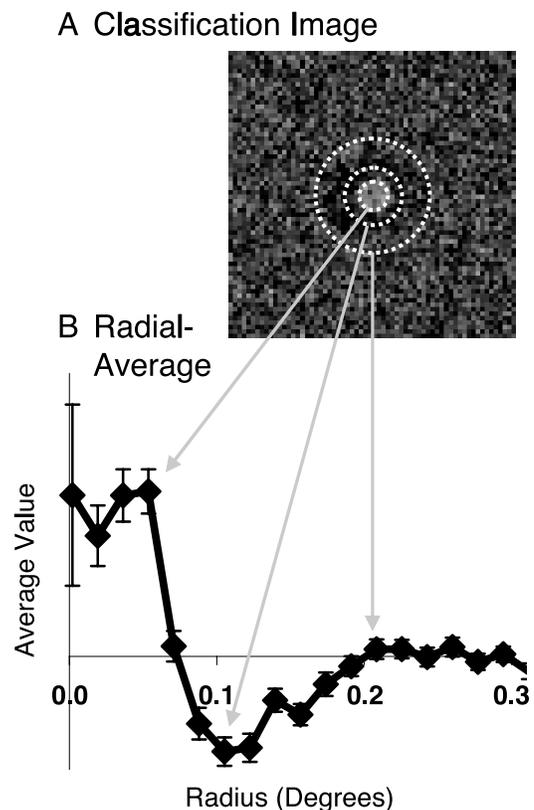


Figure 3. Radial averaging of classification images. An estimated classification image (A) has been averaged to form a plot of the average value as a function of distance from the center of the stimulus (B). The dotted lines over the image show the circular path over which the averaging takes place for points in a given distance from the center of the image. Note that the error bars ( $\pm 1$  SE) are determined from Equation 12 not from the variance of pixel values over the circular path.

the origin, will contain only one pixel, but as the bins get more distant from the origin, the radial averages encompass greater numbers of image pixels. This same procedure can be used in the spatial-frequency domain after appropriate shifting of the image data to center the origin (DC) of the discrete Fourier transform (DFT). In this case, only the real component of the complex Fourier coefficient is important because the conjugate symmetry of the DFT ensures that the radial average of the imaginary component must be zero. We have examined imaginary components of our subject's DFT classification images (data not shown), but they do not appear to contain any structure other than estimation error. The imaginary component of the DFT is indicative of odd-symmetry about the origin, which is not present in rotationally symmetric functions. Hence, we neither expect nor observe significant imaginary components in the classification images.

Let us define  $\Delta y_j$  as the radial average of  $\Delta\mathbf{q}_j$ , and let  $\Delta u_j$  be the radial average of the DFT of  $\Delta\mathbf{q}_j$ . The spatial and

spatial-frequency radial averages can be computed analogously to Equation 10 with

$$\Delta\bar{\mathbf{y}} = \frac{1}{N_T} \sum_{j=1}^{N_T} \Delta\mathbf{y}_j \quad \text{and} \quad \Delta\bar{\mathbf{u}} = \frac{1}{N_T} \sum_{j=1}^{N_T} \Delta\mathbf{u}_j. \quad (11)$$

Note that these quantities involve both an average over the stimuli in an experiment as well as the pixels in each radial bin. We can also characterize estimation errors in the radial averages by computing multivariate sample covariance matrices for the estimates according to

$$S_{\Delta\mathbf{y}} = \frac{1}{N_T - 1} \sum_{j=1}^{N_T} (\Delta\mathbf{y}_j - \Delta\bar{\mathbf{y}})(\Delta\mathbf{y}_j - \Delta\bar{\mathbf{y}})', \quad (12)$$

with the corresponding formula for  $S_{\Delta\mathbf{u}}$ . The standard errors for each radial bin in Figure 3 are computed from the square root of the diagonal elements of these matrices after division by  $N_T$ . Because of the increased number of pixels included in larger radial bins, the standard errors decrease further from the center of the image. Various Hotelling  $T^2$  hypothesis tests for classification images based on these sample statistics are described by Abbey and Eckstein (2002a). An alternative maximum-likelihood approach is described by Solomon (2002).

### Scaling the ideal observer classification image

Equation 9 provides a way to scale the ideal observer model to human observer data. Let us assume for the moment that human observers use the same filter as the ideal observer ( $\mathbf{w} = \mathbf{s}$ , defined in Equation 1) but have a significant internal noise component ( $\sigma_\varepsilon > 0$ ). We can scale this model by finding the internal noise variance such that the performance of the model is equal to the performance of each human observer. For a linear model with the Gaussian assumptions used here,

$$P_C = \Phi(d'/\sqrt{2}), \quad (13)$$

and hence we can solve this equation for  $\sigma_\varepsilon$  using the human observer  $P_C$  values and the definition of  $d'$  in Equation 7 for the ideal observer. Once we have obtained  $\sigma_\varepsilon$  in this manner, the magnitude of the expected template can be computed from Equation 9. The scaled ideal observer serves as a convenient reference for human observer data. If the human observer were simply equivalent to the ideal observer corrupted by internal noise, the resulting classification image should closely match this reference.

### Psychophysical studies

A total of three observers participated in the psychophysical studies for each of the three tasks used in this

work. Data from a fourth observer, who completed only the identification task, are not included here. With the exception of one of this article's authors (CA), subjects were naive to the goals of the research and monetarily compensated for their time. All subjects have participated in numerous psychophysical experiments involving noisy images similar to those reported here, and hence they can be considered experienced observers.

The monochrome monitor (Image Systems, Minnetonka, MN) used to display stimuli in the experiments had a linear luminance function that ranged from  $<0.01$  to  $79.6 \text{ cd/m}^2$ . The monitor control board (Dome Imaging Systems, Waltham, MA) used photometer measurements to calibrate monitor driving voltages leaving access to 256 gray levels (GL) on the linear scale. The mean background luminance of the stimuli was  $31.3 \text{ cd/m}^2$  (100 GL). Viewing distance was approximately 1 meter. Display pixels were 0.3 mm (0.017 deg visual angle) on each side.

For all experiments, the two alternatives of the forced-choice trial are displayed sequentially in what is often referred to as a two-interval forced-choice experiment. Each trial is initiated by the subject, at which point a blank field at mean background luminance is displayed for 1000 ms. The blank field is followed by the first noisy stimulus (selected at random to be either target or alternative) for 500 ms. As part of the display procedure, four "hash marks" were added to the stimulus as location cues. The cues began outside the spatial extent of the signal (0.2 deg from the signal center) and were used to reduce any inherent signal location uncertainty in the tasks. After the first stimulus, a blanking field of white noise was displayed for 1000 ms followed by 500 ms of the second stimuli (also with location cues). The second stimulus was terminated by another blanking field of white noise with an overlay that queried the observer for a response. Observers had an unlimited time to respond once the response image was displayed, but typically they responded in less than 500 ms.

After an initial training series consisting of 210 forced-choice trials in each task, we collected data for psychometric functions for each observer in each task. The purpose of this psychometric function was twofold; it gave the observers additional training in the task (five contrast levels for each psychometric function with 200 trials per contrast level), and it allowed us to find the signal contrast corresponding to approximately 85% correct ( $d' = 1.5$ ) across observers in each task. The resulting target and alternative contrasts in each experiment are given in Table 1.

The contrast of the image noise (defined as noise standard deviation divided by mean background luminance) was set to 15% yielding a noise spectral density of  $6.7 \times 10^{-6} \text{ deg}^2$ , across all experimental conditions. At this level, the noise in the images is well above threshold (see Figure 2). The influence of the noise is important because classification image analysis only works for tasks that are at least partially limited by some form of external

Task	Target, contrast	Alternative, contrast
Detection	DOG, 7.96%	None, 0.0%
Discrimination	DOG, 69.93%	DOG, 60.00%
Identification	Gaussian, 19.26% ( $\sigma = 3.3$ min)	Gaussian, 12.32% ( $\sigma = 4.1$ min)

Table 1. Parameters of the mean target and alternative. For each task, the profile type (two-dimensional, rotationally symmetric DOG or Gaussian) is listed along with the contrast of this profile with respect to mean background luminance.

noise in the image. In addition, the noise encompasses a fairly large number of luminance quantization levels in the display (for example, 29 quantization levels between the  $\pm\sigma$  levels of the noise). Burgess (1985) has shown that the errors introduced by quantizing a signal to discrete intensity levels are well modeled psychophysically as an additional noise source with variance equal to the size of the quantization step squared and divided by 12. For the noise contrast used here, quantization to discrete GL boosts the overall noise variance by less than 0.04% and can be neglected.

The psychophysical experiments used to obtain the classification images of the next section were generated from 2000 trials for each subject in each task. These trials were broken down into 20 sessions of 100 trials, collected over two or three consecutive days. Subjects started each subsequent day of an experiment with a set of 50 independent trials to refamiliarize them with the task, and they completed all 2000 trials for each task before moving on to the next.

## Results

### Observer performance

Figure 4 plots two measures of observer performance in the three tasks. Figure 4A gives the observers' percentage of correct responses in all three tasks. Performance across observers and tasks is relatively close to the nominal level of 85% correct. There is fairly consistent performance across the observers with two exceptions. The first is subject CA in the detection task, who exhibits somewhat higher performance (87.8%). The second case is subject CH in the identification task, whose performance is lower than the others (80.3%). Aside from these two exceptions, deviations in performance are not significantly different from the nominal level of 85% correct. The lack of significant deviation from the targeted performance levels and analysis of the individual session results (data not shown) indicate that observer performance was relatively stable across experiments and show very little effect of interval bias (Klein, 2001).

Figure 4B shows observer performance in terms of efficiency with respect to the ideal observer. In contrast to the proportion correct in Figure 4A, efficiency is not equivalent across the three tasks. Detection efficiency ranges from approximately 40% to 50%, whereas contrast discrimination ranks lowest for all observers with efficiencies near 25%. The identification task resulted in a divergence in observer performance. Two subjects (DV and CA) achieved their highest efficiency for this task with values near 55%. The third subject (CH) achieved a statistical efficiency of approximately 35%, roughly 5 percentage points less than his efficiency in the detection task.

### Classification images

Classification images for all observers in each task are shown in Figure 5. The classification images are shown in the spatial domain on the left side of the figure. On the right side, the spatial-frequency domain is represented by showing an image of the real component of the DFT of the classification image.

The spatial classification images all show a bright facilitatory region at the center of the image, indicating that increased intensity in this area makes the observer more likely to respond as target present. A careful inspection of the spatial classification images reveals that many of them appear to have inhibitory surrounding regions of varying strength across both tasks and observers. The spatial-frequency classification images often show activation in what appears to be a ring, indicating that only selected spatial frequencies are contributing to the observer template. A more detailed investigation of these effects can be made by considering radial averages across orientations as described above in Figure 3.

### Radial averages of classification images

Figure 6 shows radial averages of spatial and spatial-frequency classification images for each observer in the detection, discrimination, and identification tasks, respectively. For display purposes, a total of 19 radial bins are used extending from 0 to 0.3 deg in the spatial domain and 0 to 15 cpd in the frequency domain. Standard errors on these plots are taken from the diagonal elements of the radial average error-covariance matrix given in Equations 11 and 12 and have to be regarded with some care because the observed radial averages are not necessarily independent. Also plotted is the radial profile of the difference signal for the tasks, which is the linear template of the ideal observer. This template is normalized using Equation 8, with internal noise variance set to match the nominal 85% correct performance level as shown in Equation 13. This model is intended to serve as a point

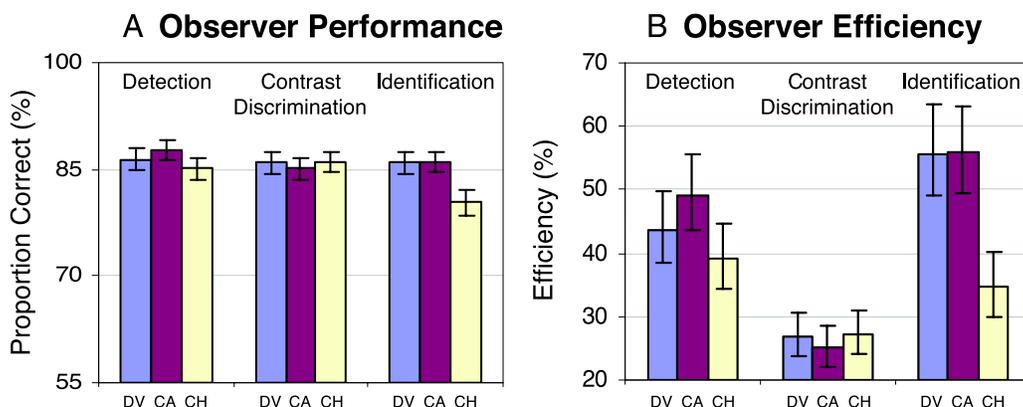


Figure 4. Observer performance data. Error bars all represent a 95% confidence interval derived from bootstrap resampling. Panel A shows the proportion of correct responses, expressed as percentages, for the observers in each experiment. Generally, performance is close to the targeted 85% correct ( $d' = 1.5$ ) level. Panel B plots the efficiency with respect to the ideal observer. While proportion correct is fairly constant, efficiency changes by over a factor of 2.

of reference, and as can be seen in the figure, it does not always agree with human observer data.

Table 2A and B gives Hotelling  $T^2$   $p$  values testing the ideal observer matched to each human observer’s performance in each task. We have performed these tests in both the spatial and spatial-frequency domain to capture a greater number of possible features for inferring the significance of differences. We use a subset consisting of the 12 bins closest to the origin but excluding the origin itself. In the spatial domain, the region used for inference includes bins centered on radial distances from 0.017 to 0.189° visual angle. In the frequency domain, this re-

gion consists of bins centered on 0.9–10.0 cpd. These bins were chosen to capture the area where the most differences would be expected to occur, while excluding the origin because of its relatively high standard error. As can be seen in the table, in most cases the ideal observer with internal noise can be rejected as a model of the observed classification images. Also in Table 2C and D are  $p$  values for differences between the classification images in different tasks for each observer. For all subjects, we find that the classification image for the discrimination task is significantly different than the detection of identification tasks. For comparisons between detection

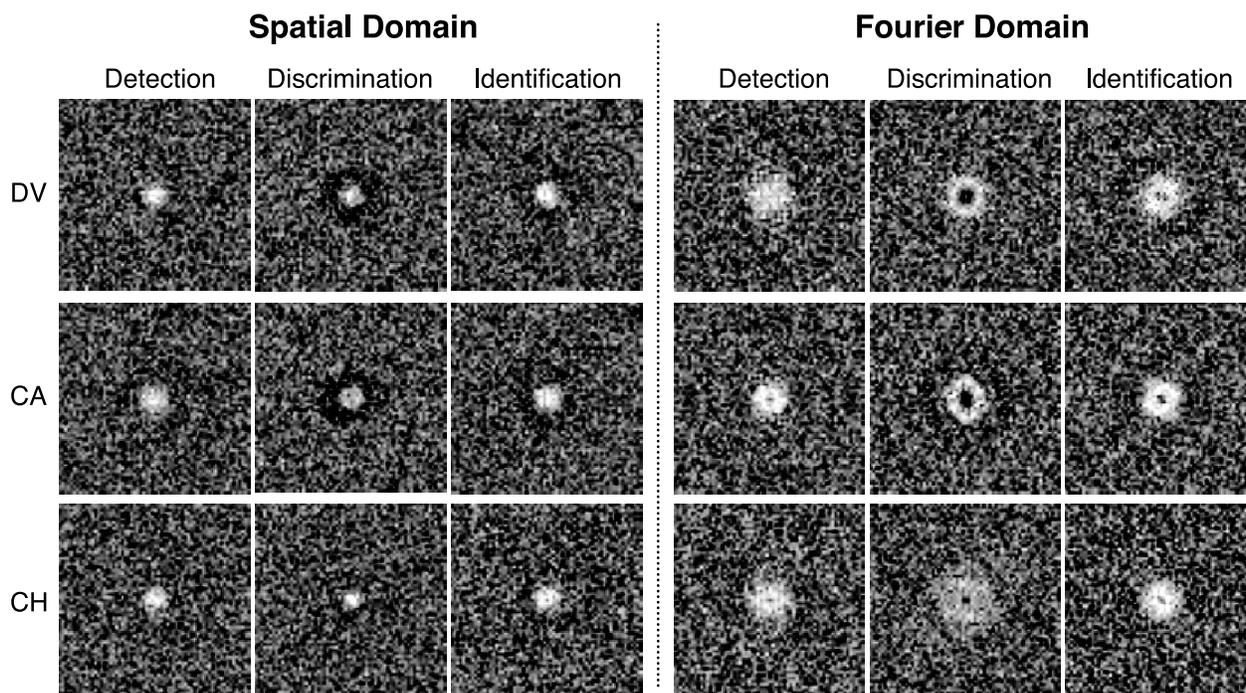
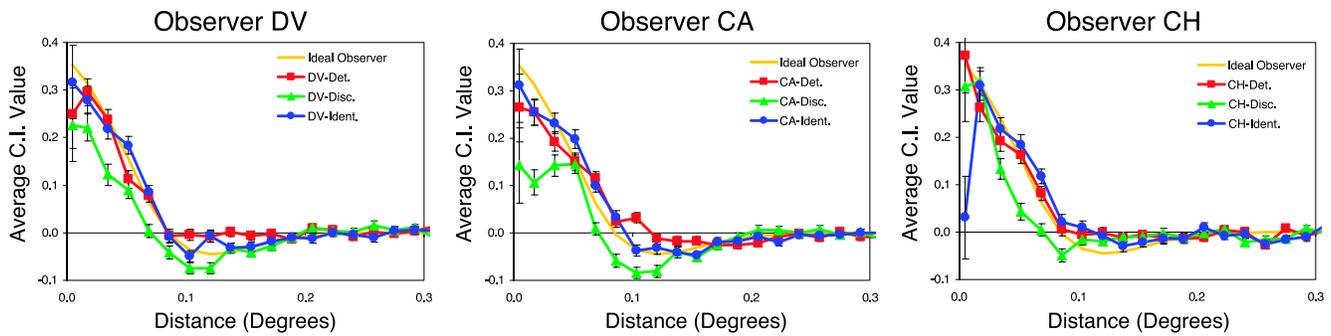


Figure 5. Observer classification images found in the three tasks. Each row corresponds to one of the observers. The column corresponds to the task in either the spatial domain (left side) or the Fourier domain (right side) using the real component of the FFT. Spatial and frequency templates for the ideal observer are found in Figure 1.

### A Spatial Domain Radial Averages



### B Spatial-Frequency Domain Radial Averages

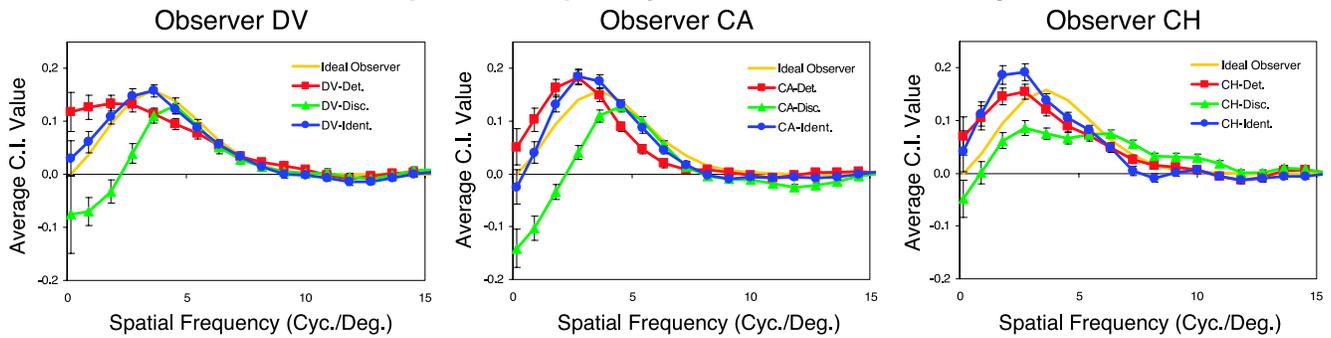


Figure 6. Radial averages of classification images. Spatial and spatial-frequency domain plots of the radially averaged classification images are shown along with classification image of the ideal observer normalized to 85% correct. The top panel (A) shows the spatial domain plots for each observer in all three tasks while the bottom panel shows the same data in the spatial-frequency domain.

	Detection	Discrimination	Identification
(A) Departure from ideal observer: Spatial domain <i>p</i> values			
DV	<.0001	<.0001	.0261
CA	<.0001	<.0001	.0032
CH	.0005	<.0001	<.0001
(B) Departure from ideal observer: Frequency domain <i>p</i> values			
DV	.0003	<.0001	.6972
CA	<.0001	<.0001	.0003
CH	.0002	<.0001	<.0001
	Det. vs. Disc.	Det. vs. Ident.	Disc. vs. Ident.
(C) Differences between tasks: Spatial domain <i>p</i> values			
DV	<.0001	.0053	<.0001
CA	<.0001	.0004	<.0001
CH	<.0001	.6999	<.0001
(D) Differences between tasks: Frequency domain <i>p</i> values			
DV	<.0001	.1080	<.0001
CA	<.0001	.0187	<.0001
CH	<.0001	.2469	<.0001

Table 2. The *p* values for agreement with ideal observer model and for intrasubject agreement across tasks. In all cases, hypothesis tests were computed for 11 radial bins immediately surrounding the central bin. Panels A and B show the Hotelling  $\chi^2$  *p* values for testing the null hypothesis that the observed classification images are derived from the ideal observer matched to human performance in the task. Panels C and D show the significance of differences between tasks for each observer.

and identification, the plots are qualitatively more similar, and the significance of differences is mixed depending on subject and test (spatial or spatial frequency).

## Discussion

The results presented above provide some insights for understanding context effects in the ways observers perform detection, discrimination, and identification tasks masked by noise. We begin by examining how well the underlying linear filter model fits the subject data. We then discuss how the classification images vary across the different tasks considered here and how they vary across observers in a given task. Finally, we turn to modeling of nonlinear effects that may help explain some of the divergence from the ideal observer in the detection and contrast discrimination tasks.

### The linear model

The linear decision variable given in Equation 6 (also known as a linear cross-correlation or linear-amplifier model) is a strong assumption about how observers perform visual tasks in noise. We use three independent

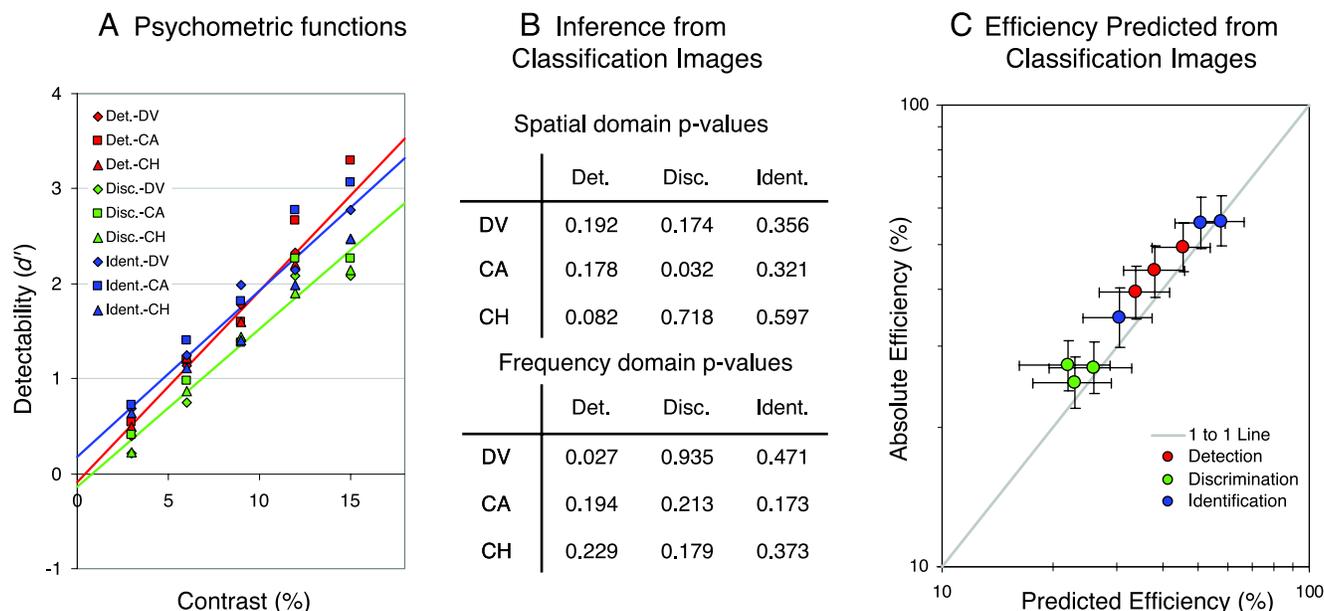


Figure 7. Three evaluations of a linear decision variable. Psychometric functions (A) plotting detectability as a function of signal contrast are well fit with by lines with a  $y$ -intercept near zero. Tables of  $p$  values (B) for the Hotelling two-sample  $T^2$  test of a linear template in the spatial and spatial-frequency domains cannot generally reject the null hypothesis of a linear observer. A comparison of absolute and predicted efficiency from classification images (C) shows that absolute efficiency is within a 95% confidence interval of the predicted efficiency using Equation 15.

measurements to assess this assumption in our data. The findings are summarized in Figure 7. To better put the results of this section in perspective, it is important to interpret the notion of a linear response. The linearity assumption in Equation 6 only requires local linearity valid for a task performed on a given set of stimuli, as described by Ahumada (1987, 2002). This linearity assumption does not require linear responses for the entire functioning visual system. It is likely that some forms of nonlinearity will be well modeled as locally linear.

A second question with regard to the linearity assumption used in this work is its stability over experimental parameters such as target performance level and noise spectral density. One consideration for this issue is the fact that a stable linear decision variable maintains constant efficiency with respect to the ideal observer. While it is difficult to pin down the range of validity of the classification images we observe without extensive further investigation, we would expect stability over the range of experimental settings that reproduce the efficiency results of Figure 4.

### Psychometric functions

It is well known that a linear decision variable produces a linear psychometric function when plotted as detectability ( $d'$ ) versus signal contrast (Eckstein, Ahumada, & Watson, 1997). This fact has been used to test for a linear decision variable by fitting a line to psychometric data, and then—assuming it fits reasonably well—looking for a  $y$ -intercept that is different from zero. Figure 7A shows

psychometric data for the experiments reported in this work with line fits to average observer performance (solid lines). The intercepts are all fairly close to zero (none are significantly different), which is in agreement with the hypothesis of a linear decision variable.

### Classification image tests for nonlinearity

As has been described in previous publications (Abbey & Eckstein, 2002a; Ahumada, 2002; Barth, Beard, & Ahumada, 1999), classification images themselves can provide a means to check the linearity assumption. For classification images derived from the 2AFC experimental paradigm, the linearity assumption can be checked by comparing a classification image derived from only from the signal-present images (i.e., using  $\mathbf{n}^+$  instead of  $\Delta\mathbf{n}$ ) to that derived from only the signal-absent images (Abbey & Eckstein, 2002a). Under the null hypothesis of a linear observer, these should not be significantly different. Figure 7B gives tables of  $p$  values for testing this hypothesis. The analysis shows relatively little evidence for nonlinearity given  $p$  values above .01 and the number of comparisons. A Bonferroni correction for multiple comparisons in each table would require a  $p$  value less than .0012 for a group rejection rate of .01, which is not found in either the spatial or spatial-frequency data.

### Efficiency of observer classification images

Recently, Murray, Bennett, and Sekuler (2005) described a method for estimating observer efficiency directly from

an observer's classification image under the assumption of a linear decision variable. Comparing efficiency computed from the classification image to efficiency computed from performance measures as in Equation 4 serves as another way to check the validity of the linear model. We have found, over a factor of 2, range of efficiency values in the different tasks investigated here for which we have collected classification images. The results of Murray et al. allow us to ask if the classification images we have collected capture these efficiency differences.

Because Murray et al. (2005) derived their formula for "yes–no" tasks, a small modification to the formula is necessary to apply the approach directly to the classification image estimated from 2AFC data by Equation 10 above. Alternatively, one can analyze the 2AFC experiment as a "yes–no" task with a compound image composed of the images in both alternatives.

At the heart of Murray et al.'s (2005) approach is the square of the cross-correlation between the observed classification image and the ideal observer template scaled to have a Euclidean norm of 1. This quantity is normalized to the square of the standard error of the estimated classification image. For 2AFC experiments, Abbey and Eckstein (2002b) have shown that this is well approximated by

$$\sigma_{\text{C.I.}}^2 = \frac{2\bar{P}_C(1-\bar{P}_C)}{N_T\sigma_n^2}, \quad (14)$$

where  $\bar{P}_C$  is the observer's proportion correct estimated in Equation 2, and  $\sigma_n$  is the standard deviation of the luminance noise in the stimuli. Following the derivation of Murray et al. (2005), we find that the efficiency of a 2AFC classification image computed from Equation 10 is estimated by

$$\text{Eff}_{\text{C.I.}} = \left( \frac{(\mathbf{s}_n^t \Delta \bar{\mathbf{q}})^2}{\sigma_{\text{C.I.}}^2 - 1} \right) \frac{\bar{P}_C(1-\bar{P}_C)}{\pi N_T} e^{\frac{1}{2}(d')^2}, \quad (15)$$

where  $\mathbf{s}_n$  is the ideal observer's classification image scaled to have a norm of 1,  $\Delta \bar{\mathbf{q}}$  is the observer's classification image from Equation 10, and  $d'$  is the observer's detectability index estimated from  $\bar{P}_C$  using Equation 3.

Figure 7C plots both methods of estimating efficiency analogously to the results of Murray et al. (2005). "Absolute" efficiency, derived from the ratio of  $d'$  values as in Equation 4, is plotted along the y-axis, and the predicted efficiency computed from the classification image is used for the x-axis. Consistent with Murray et al., we find that efficiency estimates derived from the classification images are slightly less than those obtained from the ratio of  $d'$  values, by a factor of 0.11 (they report 0.13).

It is clear from the plot that the classification images are capturing statistical efficiency of the human observers and

generally fall within the confidence interval of the estimation procedure. The different efficiency levels across tasks and observers are clearly seen in the predicted efficiency. For example, the relative drop in efficiency of subject CH in the identification task can be seen in the efficiencies calculated by both methods.

In general, Figure 7 indicates that the classification images we have measured are reflective of the factors influencing performance for our three tasks.

## Comparisons of classification images across tasks and observers

The radial averages of the classification images plotted in Figure 6 show fairly consistent differences across tasks for each observer. The plots show that the most substantial differences occur at spatial frequencies below approximately 5 cpd. There also appears to be a transition in the classification images from overemphasis of low spatial frequencies in the case of detection to suppression and sign reversal in contrast discrimination.

### Detection

Radial averages of the detection classification images in Figure 6 show relatively good agreement with the ideal observer to a radius of about 0.08 deg (4.8 arcmin) in the spatial domain. Above this point, all three observers show a reduced inhibitory response from that given by the ideal observer. The observer showing the greatest inhibitory response (CA) shows that response at a greater distance from the center than the ideal observer (0.18 vs. 0.12 deg for the ideal observer). The relatively small inhibitory response of the spatial classification images translates into a relatively large template response at low spatial frequencies. In all cases, the observer templates show a relatively greater response than the ideal observer at spatial frequencies below 3–4 cpd. The observer templates also show a slightly reduced response in spatial frequencies from 3 to 4 cpd to approximately 6–8 cpd giving the appearance of a band-pass observer classification image that is tuned to a lower spatial frequency. As seen in Table 2, in all cases we can reject the ideal observer template with high statistical significance.

### Contrast discrimination

The radial averages of the contrast discrimination classification images show a striking difference from the detection task. This task also shows the greatest divergence between observers. Two of the three observers (DV and CA) show an enhanced inhibitory response in the spatial domain plots with peak inhibitory responses at 0.10–0.12 deg. The third observer (CH) shows less

inhibitory response, with a peak value at 0.086 deg. The spatial-frequency plots for observers DV and CA peak at approximately 5 cpd, and then drop well below zero to negative values consistent with an overemphasis on the inhibitory surround. The third observer shows a fairly broad peak (2.7–6.4 cpd) and crosses zero only in the dc component. As in the detection task, we can reject the ideal observer template in all cases with high statistical significance.

### Identification

The identification task produced much better agreement with the ideal observer than either the detection or contrast discrimination tasks. All observers show a facilitatory central region with a mild inhibitory surround similar to the ideal observer. In the spatial-frequency domain, all observers show a peak at or slightly below the 4-cpd peak in the ideal observer response. Observer DV cannot be rejected from the ideal observer using an error rate ( $\alpha$ ) of 0.01. The remaining two observers show low-frequency enhancement and a downward shift of the peak frequency response.

### Observer differences

We have also regrouped the frequency domain classification images—where observer differences were the most pronounced—to better show the differences between observers in Figure 8. Observer differences are most striking in the contrast discrimination task, where two of the three observers (CA and DV) show relatively similar peaking at approximately 4.5 cpd, and a severe drop at lower frequencies to the point of negative classification image weights below 2 cpd. The differences between these two observers are on the borderline of statistical significance using a one-sample paired Hotelling  $T^2$  test ( $df = 11$ ,  $p > .048$ ) (Abbey & Eckstein, 2002a). These two

observers are both significantly different from observer CH ( $p < .0001$ ), who utilizes a much broader range of spatial frequencies, extending above 10 cpd, and shows little to none of the negative weighting at the lowest spatial frequencies. This finding is similar to recent works by Meese, Hess, and Williams (2005) and Solomon (2002), which suggest individual differences in spatial summation for suprathreshold tasks. Also noteworthy here is that the three observers had nearly identical levels of performance in this task (CA: 85.1%, DV: 85.9%, and CH: 86.1%). So we find that classification image analysis is able to detect differences in the way that observers perform tasks that are not observable from performance measures alone.

The detection and identification tasks show less substantial differences between observers. In the detection task, we see that the observers differentiate by the sharpness in the peak (bandwidth) of their frequency profile. Observer DV shows the least peaked profile, and observer CA the most. All plots appear to peak at 2–3 cpd, which is below the peak frequency of the ideal observer. The identification task resulted in relatively good agreement with the ideal observer for observers DV and CA, as might be expected from their relatively high statistical efficiency seen in Figure 4.

### Modeling differences in observer classification images across tasks

It is clear from Figures 6 and 8 that a simple ideal observer filter with internal noise is not an adequate model for human observers across these three tasks. No single linear filter will explain the measured classification images across tasks for any subject in this study. In this section, we consider three global effects as possibilities to explain task-dependent differences in subject classification images we have observed. We consider effects of

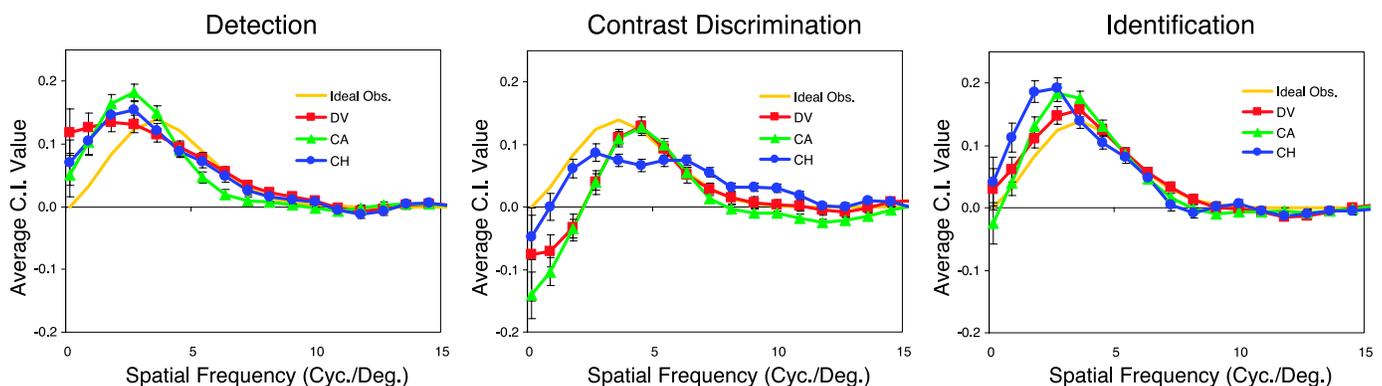


Figure 8. Fourier domain classification images grouped by task. Each plot gives the Fourier domain radial average for all three observers in each task allowing visualization of observer differences. The ideal observer, plotted in gold and normalized to the nominal 85% correct level, is plotted for reference.

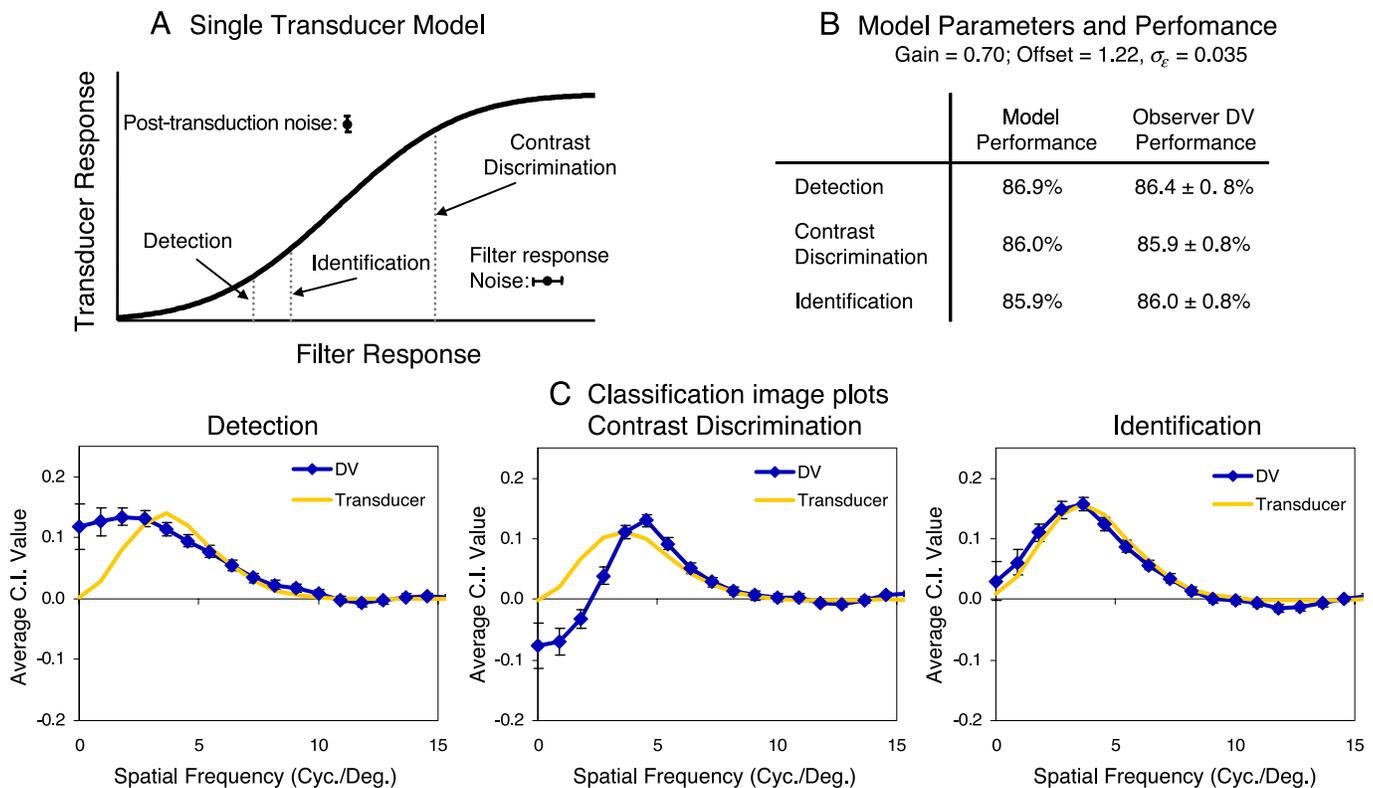


Figure 9. A single-filter transducer model. The sigmoidal curve (A) shows the transducer response as a function of the filter response. The dotted lines indicate the mean filter response to signal-absent stimuli (the mean response to signal-present data is slightly higher). Noise in the stimulus propagates through the filter resulting in response variability on the scale ( $\pm 1\sigma$ ) indicated by the horizontal error bars in the lower right corner. Posttransduction internal noise is added with magnitude indicated by the vertical error bars in the upper left corner. Parameters of the model and performance comparisons with observer DV (B) show good agreement in all tasks. However, comparisons of the Fourier domain radial averages of the classification images (C) in the detection and contrast discrimination tasks.

a postfilter transducer function, an early nonlinear transduction of image intensity, and spatial uncertainty on classification images. Each of these effects incorporates some form of nonlinearity into the decision variable defined in Equation 6, although the nonlinearity is generally weak and still permits estimation of a meaningful classification image. As we shall see below, none of the models investigated will account for all of our observed data. However, modeling efforts are ongoing, and the purpose here is to see if they share qualitative features of the human observer data.

### Nonlinear transducer

We begin by modeling a nonlinear transducer (for example, Foley & Chen 1999; Heeger, Simoncelli, & Movshon, 1996; Klein & Stromeyer, 1980; Nachmias & Sansbury, 1974; Watson & Solomon, 1997) to explain the differences in classification images across tasks. Because we are focusing on the transducer as the mechanism leading to different classification images, the DOG difference signal is used as the spatial weighting in all three tasks (i.e.,  $\mathbf{w} = \mathbf{s}$ ). The functional implementation of this transducer

is to take the cross-correlation component of Equation 6 and apply a sigmoidal nonlinearity to it before adding the Gaussian internal noise component. For a sigmoidal response function, we choose the cumulative normal distribution,  $\Phi$ . This yields the nonlinear response function,

$$\lambda(\mathbf{g}) = \Phi[(\gamma(\mathbf{w}'\mathbf{g} - \beta))] + \varepsilon, \quad (16)$$

where  $\gamma$  and  $\beta$  are the gain and offset applied to the filter response before transduction, and  $\varepsilon$  is posttransduction internal noise. A plot of the transducer function and values of gain, offset, and standard deviation of the internal noise component are given in Figures 9A and B.

Figure 9B also gives the performance of this model in 200,000 Monte Carlo trials in each task, in comparison with observer DV. We see that the model very closely matches the performance of observer DV with differences well below the standard error in  $P_C$  determined from 2000 psychophysical trials in each task. This good agreement is not overly surprising given that we have three observed performance levels to match and three free parameters in the model (gain, offset, and internal noise standard deviation). Nonetheless, Figure 9B shows that a sigmoidal

transducer can explain the observed performance data with high accuracy.

In contrast to the performance data, the single-filter transducer model does not capture the salient differences in the classification images across tasks, as shown in [Figure 9C](#). Here we plot the radial averages of subject DV's Fourier domain classification images along with the predictions of the nonlinear transducer model. The model generally reproduces the magnitude of the classification image and appears to be a good model for the identification task where it cannot be statistically rejected (neither can the linear model). However, the transducer model retains the profile of the difference signal in the detection and discrimination tasks whereas the subject data do not.

This model illustrates a well-known result from the reverse correlation literature (Nykamp & Ringach, 2002). Namely, estimates of linear receptive fields are robust to a postfilter nonlinearity. A linear filter followed by a (monotone) nonlinear transducer will produce a classification image (or receptive field) that still resembles the linear filter. The nonlinearity only modifies the magnitude of the classification image, not the shape. Hence, because we observe three different observer templates for our three tasks, we can conclude that there is no single-filter transducer model that will adequately explain our human observer data across the three tasks considered.

### Early nonlinearity

The single-filter transducer model described above can be thought of as adding a late nonlinearity that occurs after cross-correlation. Here we consider the effects of an early nonlinearity that occurs before cross-correlation. In particular, we investigate the possibility that an ideal observer strategy following an early nonlinearity might explain the differences in human classification images across tasks. The early visual system is often modeled as incorporating a compressive (decelerating) nonlinearity to encode luminance or contrast (see, for example, Hood & Finkelstein, 1986). We will investigate cross-correlation after such a nonlinear transformation. We will use the cumulative normal function again as our transducer function, but the offset will be set so that this function is essentially compressive over the intensity range of the experiments.

We begin by positing a (noisy) nonlinear transducer function,  $T$ , that acts on each point in an image,  $\mathbf{g}$ . The posttransduction image,  $\mathbf{y}$ , is defined by

$$y_m = T(g_m) + \nu_m, \quad (17)$$

where  $g_m$  is the intensity of the image at location (pixel)  $m$ ,  $y_m$  is the posttransduction intensity at  $m$ , and  $\nu_m$  is posttransduction (internal) noise at  $m$ . We will assume that the posttransduction noise is white with uniform standard deviation  $\sigma_\nu$ . We consider the ideal observer strategy after

the image has undergone an early nonlinearity followed by noise as described in [Equation 17](#). [Appendix](#) describes how a first order Taylor series yields approximately optimal image weights,

$$w_m = \frac{s_m T'(b_m + 0.5s_m)^2}{T'(b_m + 0.5s_m)^2 \sigma_n^2 + \sigma_\nu^2}, \quad (18)$$

where  $T'(b_m + 0.5s_m)$  is the derivative of the transducer function evaluated at the average of the target ( $b_m + s_m$ ) and alternative ( $b_m$ ) defined in [Equation 1](#).

[Figure 10](#) examines this model in the context of the detection, discrimination, and identification tasks used in this work. [Figure 10A](#) shows the transducer function, which is decelerating over the range of contrast values used here. In [Figure 10B](#), the performance of the model is well matched (subject DV) in the contrast discrimination and identification tasks, but it outperforms the subject in the detection task. Unlike the late nonlinearity model above, we did not find reasonable parameter settings that would fit performance in all three tasks. The classification images derived from this model using [Equation 18](#) only provided a good quantitative fit to subject DV's spatial-frequency classification image data in the identification task. However, it is worth noting some qualitative similarity in the contrast discrimination task, particularly at low spatial frequencies where the model's classification image shows a peak below 5 cpd and a steep drop with sign change at lower frequencies similar to the human observer data. In the detection task, the model remains similar to the ideal observer (without nonlinearities) and does not capture features of the human observer data at low spatial frequencies.

[Equation 19](#) can also be used to evaluate the notion of a common transducer function. It is most convenient to divide both sides by  $s_m$  to obtain a weight-to-signal ratio,

$$\frac{w_m}{s_m} = \frac{T'(b_m + 0.5s_m)^2}{T'(b_m + 0.5s_m)^2 \sigma_n^2 + \sigma_\nu^2}, \quad (19)$$

which shows an explicit dependence of the weights on the transducer slope. If we have estimated a set of weights, we can use [Equation 19](#) to investigate how those weights might fit in the context of an optimal posttransduction response. For example, if  $w_m$  is the opposite sign of  $s_m$ , we can rule out the template as optimal because there is no (real-valued) transducer function that can make the right side of [Equation 19](#) negative. Also, if two different points,  $m$  and  $m'$ , in an image happen to satisfy  $b_m + 0.5s_m = b_{m'} + 0.5s_{m'}$ , then we must have  $w_m / s_m = w_{m'} / s_{m'}$ , or else the weights are not optimal. This latter point can be applied to different tasks as well.

Imagine that we have two tasks, one with background  $\mathbf{b}$  and signal  $\mathbf{s}$  (as above), and a second task with background

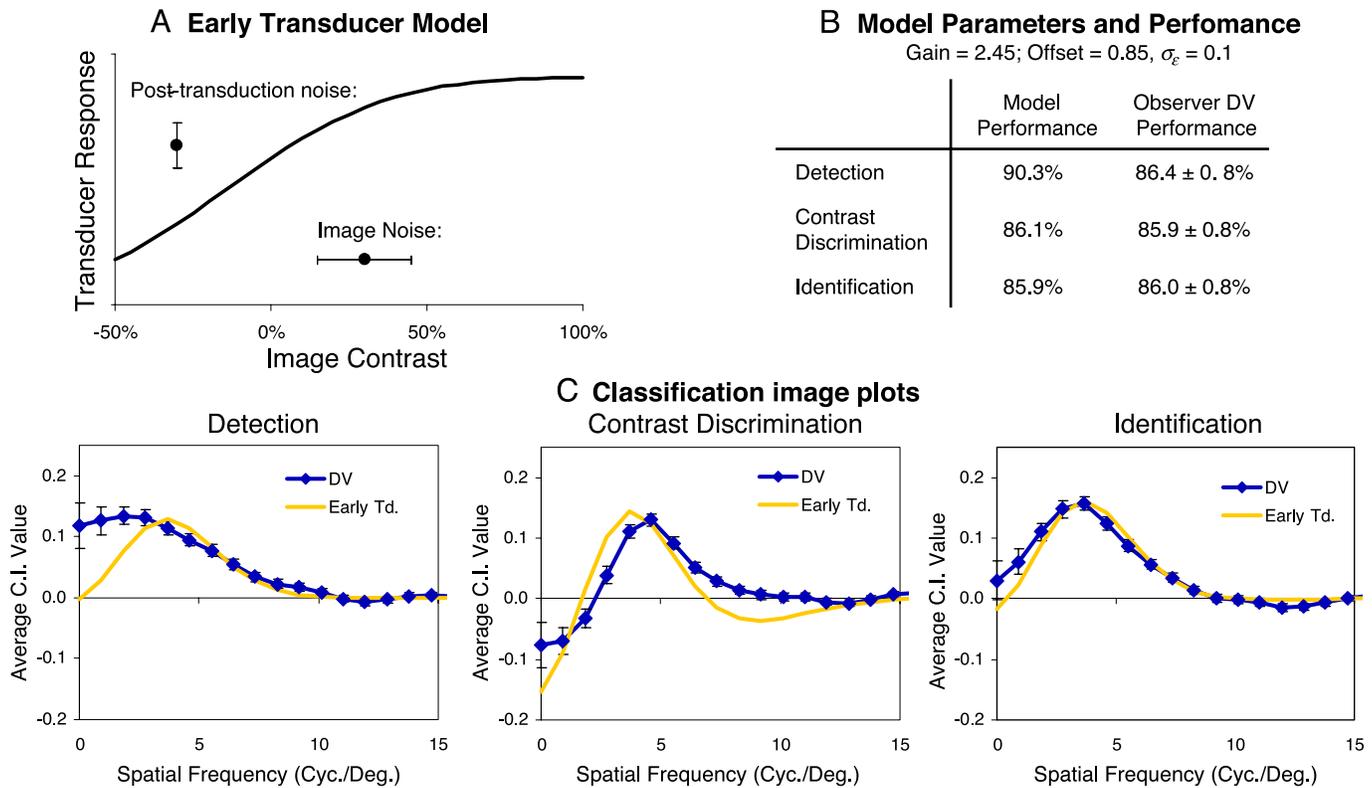


Figure 10. Early transducer model. The sigmoidal curve (A) shows the transducer response as a function of the image contrast (relative to mean monitor luminance of 31.3 cd/m<sup>2</sup>). The magnitude of noise in the image ( $\pm 1\sigma$ ) is indicated by point with the horizontal error bars, and the magnitude of posttransduction internal noise is indicated by the point with vertical error bars. Parameters of the model and performance comparisons with observer DV (B) show good agreement in the contrast discrimination and identification tasks. Comparisons of the Fourier domain radial averages of the classification images (C) show poor agreement in the detection task, notably better agreement in the contrast discrimination task, and reasonably good agreement in the identification task.

**u** and signal **x**. We also have two sets of weights, **w** for the first task and **z** for the second task, and we would like to know if they are optimal under the same transducer function. If there is some contrast level in the images that is common to both, then  $b_m + 0.5s_m = u_{m'} + 0.5x_{m'}$ , for some  $m$  and  $m'$ . If the same transducer applies to both tasks, then it must have the same slope at these points and hence we must have  $w_m / s_m = z_{m'} / x_{m'}$ . Failure to achieve this effectively rules out the possibility of optimal weights with any common transducer. Figure A1 in the Appendix gives an example of what the weight-to-signal ratios look like for a common early transducer.

Figure 11 plots the weight-to-signal ratios for subject DV as a function of average contrast for the detection, discrimination, and identification tasks. Because estimated classification images are used as the weights, Equation 9 indicates that there will be an additional scaling factor for these weights. Nonetheless, each task should fall on the same curve up to a scale factor. Figure 11 shows this is not the case. The ratio values for the detection task appear to be decreasing for the discrimination task. The detection task starts with values very close to zero and increases with increasing average contrast, and the identification task shows a mild peak near 5% contrast. This would appear to rule out the possibility of explaining the differences in our

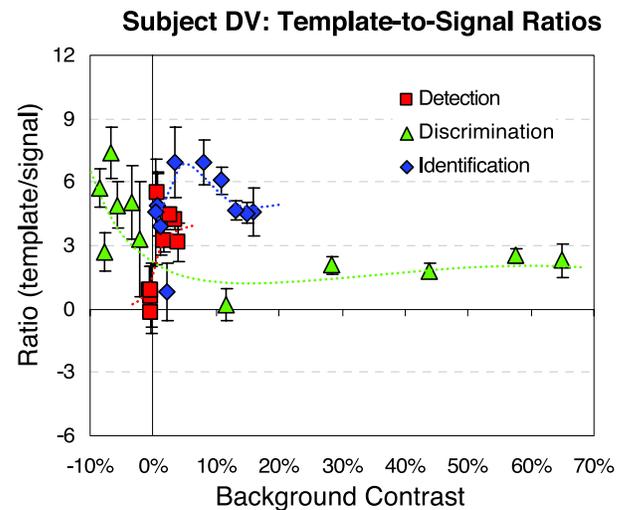


Figure 11. The ratio of the estimated classification image to the difference signal is plotted as a function of the average contrast at each point. Under the assumption of an (approximately) optimal decision maker after an early nonlinearity, all three tasks should fall on the same curve up to a scale value.

observed classification images with any single early non-linearity. Perhaps a more intuitive way to understand this result is to consider the spatial classification image plots in Figure 8. A notable feature of these plots is the differing amount of inhibitory surround across tasks. Equation 18 implies that large template values—positive or negative—require larger values of  $T'$  (i.e., a steeper transducer function). The strong inhibitory region in the discrimination task implies a relatively steep transducer slope for image intensities that are slightly less than the average intensity of the background. However, we would then also expect relatively strong inhibitory regions in the detection task as well, which are not observed in the data.

**Spatial uncertainty**

The final model we test incorporates nonlinear effects of spatial uncertainty on the classification images in the three tasks. Spatial uncertainty, as used here, can be thought of as another late nonlinearity because it occurs after linear processing (for example, Burgess & Ghandeharian, 1984a, 1984b; Nachmias, 2002; Pelli, 1985). Figure 12A shows a simple one-dimensional illustration of how spatial uncertainty is implemented. The observer’s decision variable is taken to be the max of several noisy linear filter outputs

that correspond to the same DOG template positioned on various locations near the target center. The range of locations reflects the degree of spatial uncertainty in the observer. The observer decision variable is then defined by

$$\lambda(\mathbf{g}) = \max_u(\mathbf{w}_u^t \mathbf{g} + \varepsilon_u), \tag{20}$$

where the index  $u$  identifies a spatially displaced DOG template,  $\mathbf{w}_u$  ( $\|\mathbf{w}_u\| = 1$ ). We presume that each filter has its own independent internal noise component, so the internal noise also requires the subscript  $u$  in this model. This model has two free parameters, the range of uncertainty, which determines the number of filters, and the standard deviation of the internal noise in each filter.

Figure 12B gives the model parameters and performance in 200,000 Monte Carlo trials. A spatial DOG template was positioned at every pixel less than 0.08 deg from the target center, for a total of 69 filter responses. This level of uncertainty covers the entire central excitatory region of the DOG target. Independent Gaussian noise was added to each response. The internal noise standard deviation was adjusted so that performance matched observer DV in the detection task where we expect uncertainty effects of be most pronounced. We see in Figure 12B that performance was indeed very close in the detection task. However, in the contrast

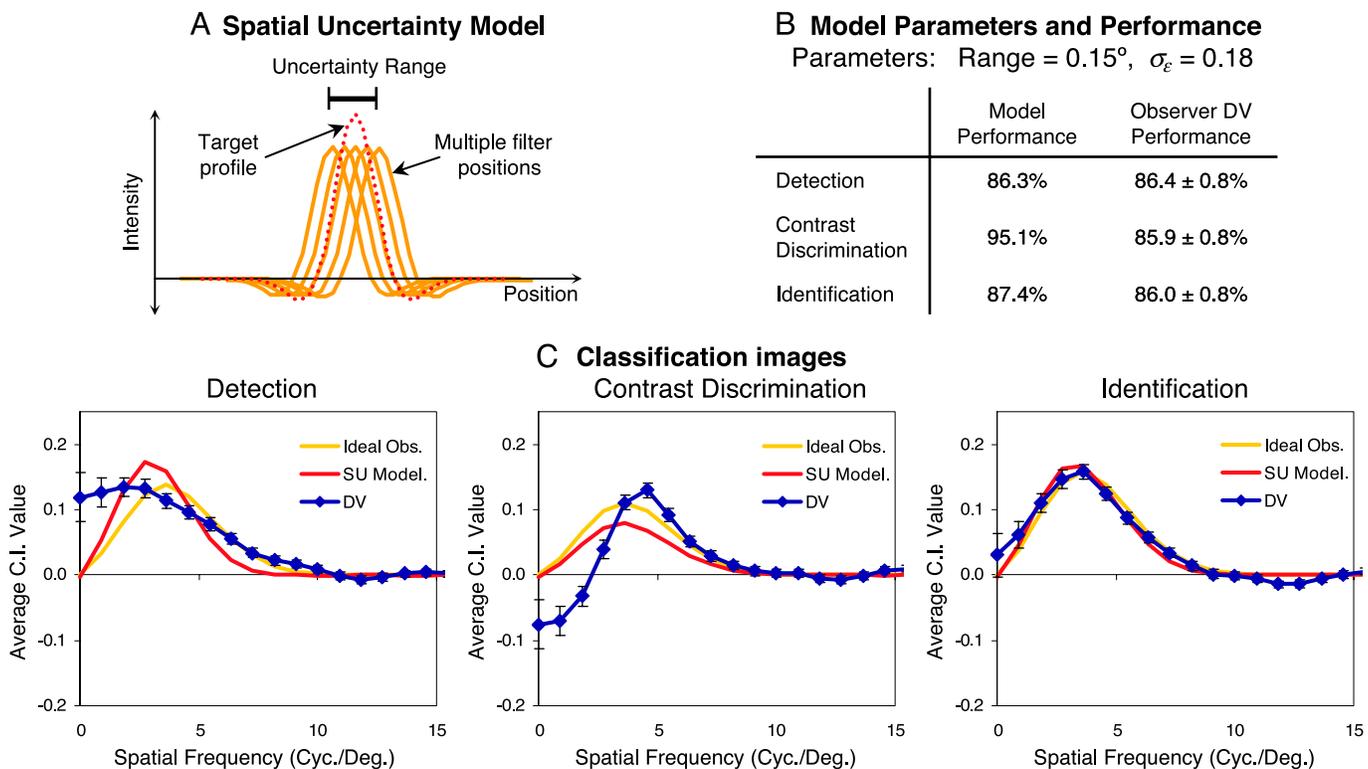


Figure 12. A nonlinear spatial uncertainty model. Multiple DOG filters (A) at locations around the target center (in red) generate a set of noisy filter responses. From these, the observer is presumed to form a decision variable by a max operation. Parameters of the model and performance comparisons with observer DV (B) show good agreement in the detection and identification tasks. Comparisons of the Fourier domain radial averages of the classification images (C) show low-frequency enhancement in the detection task with some qualitative similarities to the human observer data.

discrimination task, where the presence of a strong pedestal reduces the effect of uncertainty, the model substantially outperforms human observers. This is true to a lesser extent in the identification task where the Gaussian pedestal is weaker as well.

Figure 12C shows radial plots of spatial-frequency classification images for observer DV, the ideal observer fitted for equal performance, and the spatial uncertainty model assessed from 200,000 Monte Carlo trials. In the identification task, the model is almost indistinguishable from the ideal observer. In the contrast discrimination task, the classification image obtained from the uncertainty model resembles the ideal observer closely but at a smaller magnitude, which would be expected from Equation 9, given the high performance of the model (and hence larger value of  $d'$ ). The most pronounced differences between the ideal observer and the spatial uncertainty model occur, as expected, in the detection task. While the uncertainty model clearly does not match the classification image of subject DV, it does display one qualitative similarity, namely, the shift of emphasis to lower spatial frequencies. It is also worth noting that the uncertainty model is a better qualitative match to observers CA and CH, whose spatial-frequency classification images are more strongly peaked in the detection task, as can be seen by comparison with Figure 8.

## Summary and conclusions

We report the results of 2AFC experiments for detection, contrast discrimination, and identification tasks in Gaussian white noise for which the mean difference between target and alternative has a common DOG profile. The common difference signal leads to a common linear mechanism for the ideal observer implying the same classification image (up to scale factors) in all three tasks. The classification images we report here reveal that human observers use images in different ways to perform these tasks. We find significantly different classification images across the three tasks for our group of subjects as well as significant departures from the ideal observer strategy in the detection and contrast discrimination tasks. These differences occur towards the low end of the spatial-frequency domain, below 5 cpd. In the detection task, we find enhancement of these low spatial frequencies, whereas in the contrast discrimination task, low frequencies are suppressed and then subject to a change in sign. In the identification task, subjects' classification image profiles generally agree reasonably well with the ideal observers', although we do find some low-frequency enhancement for some observers.

We have attempted to model a number of simple nonlinear visual components that may be able to explain qualitative characteristics of our classification image data. We have looked at early (prefilter) and late (postfilter) nonlinearities as well as intrinsic target location uncertainty.

We find that a late nonlinearity in the form a sigmoidal transducer that accurately captures the performance of human observers in these tasks has no effect on the form of the classification image. This finding is consistent with known properties of reverse correlation. As a result, the different classification images we have obtained from human observers cannot be explained as a transducer effect alone. An early compressive nonlinearity on the luminance followed by first order optimal spatial weights has little effect on the detection and identification tasks, but it results in a classification image for the contrast discrimination task that shows low-frequency suppression and sign reversals similar to human observer data. However, a closer analysis of our data shows that no single late nonlinearity followed by an optimal cross-correlation filter can explain our data either. We find that the effects of nonlinear spatial uncertainty can act as a mechanism for low-frequency enhancement of the classification images for the detection task. However, none of the models investigated fully explain the observed data, and investigation into combinations of effects and other physiologically based models are ongoing.

One finding that highlights some of the additional challenges that comes in trying to model observer classification images is the magnitude of intersubject variability. Looking within a single task, we find that in some cases the different observers have significantly different classification images despite nearly identical levels of performance in the task. This supports recent findings by Meese et al. (2005), Solomon (2002), and Thomas and Knoblauch (2005), with regard to observer variability in spatial or temporal summation. While intersubject variability will require more involved models to be effective, it also is an example of how classification image analysis can reveal the workings of the visual system on a subject-by-subject basis in ways that are not observable in more traditional comparisons of absolute performance.

## Appendix

In this Appendix, we will derive the classification image (Equation 17) of an ideal observer that is constrained by an early noisy nonlinearity (Equation 16). The result is approximate because we utilize a first order Taylor series expansion for a critical step, but we will examine the validity of that expansion as well.

Given the outputs of the early nonlinearity defined in Equation 16,  $y_m$ , we model the formation of an internal response as a weighted sum over the posttransduction image,

$$\lambda = \sum_m q_m y_m. \quad (\text{A1})$$

If we assume that the transducer function is smooth, and has relatively low curvature over the intensity range of the noise and the difference signal, then we can use a

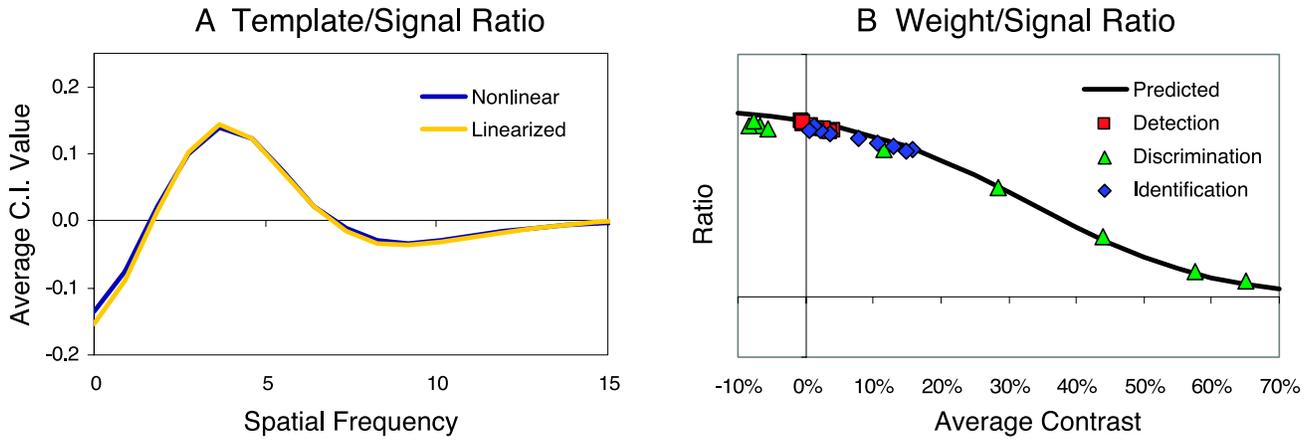


Figure A1. Validation of "linearized" template for the early transducer model. The contrast discrimination classification images (A) for the nonlinear and linearized models show good agreement. The nonlinear model (Equation A1) is evaluated through Monte Carlo (200,000 samples), and the linearized template is taken from Equation A6. The contrast discrimination task is shown because it had the largest difference with an RMS difference from 1 to 10 cpd of 0.0054. RMS differences for the detection and identification tasks were 0.0012 and 0.0014, respectively. The weight/signal ratios (B) also show good agreement between theoretical predictions from Equation 19, based on the linearized model, and the Monte Carlo templates based on the nonlinear model. Ratio data are plotted for difference signal contrasts greater than 0.5%. Note that this plot differs from Figure 11 in that scaling effects of the classification image procedure are corrected here, and hence all three tasks fall on the same line.

first order Taylor series expansion to cast the decision variable in Equation A2 in terms of the image,  $\mathbf{g}$ . We will consider Taylor series about each point  $m$ , expanded about the average intensity of the images at each point,

$$y_m \cong T(b_m + 0.5s_m) + T'(b_m + 0.5s_m) [g_m - (b_m + 0.5s_m)] + \nu_m, \quad (\text{A2})$$

where  $b_m + 0.5s_m$  is the average intensity at location  $m$ , and  $T'$  is the first derivative of  $T$ . Substituting Equation A2 into Equation A1 will give an approximate expression for the internal response. But Equation 5, that is, Score = Step $[\lambda(\mathbf{g}^+) - \lambda(\mathbf{g}^-)]$ , shows how decisions in a 2AFC task are taken as a difference in responses, and thus constant terms that will be in both signal-present and signal-absent images are irrelevant. Therefore, an equivalent linear weighting of image intensities is given by

$$\lambda \cong \sum_m q_m [T'(b_m + 0.5s_m)g_m + \nu_m]. \quad (\text{A3})$$

Equation 6 specifies a linear weighting  $\mathbf{w}$  and a scalar internal noise variable  $\varepsilon$ . Equation A3 defines these as

$$w_m \cong q_m T'(b_m + 0.5s_m), \text{ and } \varepsilon = \sum_m q_m \nu_m. \quad (\text{A4})$$

We now turn to deriving (approximately) optimal post-transduction weights,  $q_m$ .

Recall that signal-present and signal-absent images were defined as  $\mathbf{g}^+ = \mathbf{b} + \mathbf{s} + \mathbf{n}^+$  and  $\mathbf{g}^- = \mathbf{b} + \mathbf{n}^-$ , respectively. We can use these definitions in Equation A2 to obtain first order propagation through the transducer to  $y^+$  and  $y^-$ ,

$$\begin{aligned} y_m^+ &\cong T(b_m + 0.5s_m) \\ &\quad + T'(b_m + 0.5s_m)(n_m^+ + 0.5s_m) + \nu_m^+ \\ y_m^- &\cong T(b_m + 0.5s_m) \\ &\quad + T'(b_m + 0.5s_m)(n_m^- - 0.5s_m) + \nu_m^-. \end{aligned} \quad (\text{A5})$$

It is well-known (Fukunaga, 1990) that the optimal detector is given by the difference in mean values normalized by the variance at each location, and so elements of the optimal linear weights are given by

$$q_m = \frac{s_m T'(b_m + 0.5s_m)}{T'(b_m + 0.5s_m)^2 \sigma_n^2 + \sigma_\nu^2}, \quad (\text{A6})$$

where  $\sigma_n$  is the standard deviation of the image noise. The resulting linear image weights are given using Equation A4 as

$$w_m = \frac{s_m T'(b_m + 0.5s_m)^2}{T'(b_m + 0.5s_m)^2 \sigma_n^2 + \sigma_\nu^2}, \quad (\text{A7})$$

which is the approximate form of the classification image given in Equation 17.

As a test of the Taylor series approximation, we compare the classification image derived from the nonlinear response variable in Equation A1 (with weights defined in Equation A5) to the predicted classification image given in Equation A6. Results from two comparisons are found in Figure A1. Here we see a comparison between Fourier domain radial averages for the linearized classification image in Equation A6 and the classification image derived from the nonlinear model in Equation A1 evaluated by Monte Carlo studies with 200,000 forced-choice trials. Figure A1 shows that the resulting classification images are quite similar. The figure also shows that the predicted values for the weight/signal ratios given in Equation 19 are very close to those determined from the classification image of the early nonlinearity model.

## Acknowledgments

The authors would like to thank Cedric Heath and Darko Vodopich for their diligent efforts in the psychophysical studies. We thank John Foley, Steve Shimozaki, Bosco Tjan, and Gordon Legge for helpful discussions and comments on the manuscript.

This research was supported by National Institutes of Health Grants R01-53455 and R01-015925.

Commercial relationships: none.

Corresponding author: Craig K. Abbey.

Email: [abbey@psych.ucsb.edu](mailto:abbey@psych.ucsb.edu).

Address: Department of Psychology, University of California, Santa Barbara, CA 93106.

## References

- Abbey, C. K., & Eckstein, M. P. (2000). Estimates of human-observer templates for a simple detection tasks in correlated noise [Abstract]. *Proceedings of SPIE*, 3981, 70–77.
- Abbey, C. K., & Eckstein, M. P. (2002a). Classification image analysis: Estimation and statistical inference for two-alternative forced-choice experiments. *Journal of Vision*, 2(1), 66–78, <http://journalofvision.org/2/1/5/>, doi:10.1167/2.1.5. [PubMed] [Article]
- Abbey, C. K., & Eckstein, M. P. (2002b). Optimal estimates of human-observer templates in two-alternative forced-choice experiments. *IEEE Transactions on Medical Imaging*, 21(5), 429–440. [PubMed]
- Abbey, C. K., & Eckstein, M. P., Bochud, F. O. (1999). Estimation of human-observer templates in two-alternative forced-choice experiments [Abstract]. *Proceedings of SPIE*, 3663, 284–295.
- Ahumada, A. J., Jr. (1987). Putting the visual system noise back in the picture. *Journal of the Optical Society of America A*, 4(12), 2372–2378. [PubMed]
- Ahumada, A. J., Jr. (2002). Classification image weights and internal noise level estimation. *Journal of Vision*, 2(1), 121–131, <http://journalofvision.org/2/1/8/>, doi:10.1167/2.1.8. [PubMed] [Article]
- Ahumada, A. J., & Lovell, J. (1971). Stimulus features in signal detection. *Journal of Acoustical Society of America*, 49, 1751–1756.
- Ahumada, A. J., Marken, R., & Sandusky, A. (1975). Time and frequency analyses of auditory signal detection. *Journal of Acoustical Society of America*, 57, 385–390.
- Barth, E., Beard, B. L., & Ahumada, A. J., Jr. (1999). Nonlinear features in Vernier acuity [Abstract]. *Proceedings of SPIE*, 3644, 88–96.
- Beard, B. L., & Ahumada, A. J., Jr. (1998). Technique to extract relevant image features for visual tasks [Abstract]. *Proceedings of SPIE*, 3299, 79–85.
- Beutter, B. R., Eckstein, M. P., & Stone, L. S. (2003). Saccadic and perceptual performance in visual search tasks: I. Contrast detection and discrimination. *Journal of the Optical Society of America A*, 20(7), 1341–1355. [PubMed]
- Brady, M. J., & Kersten, D. (2003). Bootstrapped learning of novel objects. *Journal of Vision*, 3(6), 413–422, <http://journalofvision.org/3/6/2/>, doi:10.1167/3.6.2. [PubMed] [Article]
- Burgess, A. (1985). Effect of quantization noise on visual signal detection in noisy images. *Journal of the Optical Society of America A*, 2(9), 1424–1428. [PubMed]
- Burgess, A. E., & Colborne, B. (1988). Visual signal detection: IV. Observer inconsistency. *Journal of the Optical Society of America A*, 5, 617–627. [PubMed]
- Burgess, A., & Ghandeharian, H. (1984a). Visual signal detection: I. Ability to use phase information. *Journal of the Optical Society of America A*, 1, 900–905. [PubMed]
- Burgess, A. E., & Ghandeharian, H. (1984b). Visual signal detection: II. Signal-location identification. *Journal of the Optical Society of America A*, 1, 906–910. [PubMed]
- Burgess, A. E., Li, X., & Abbey, C. K. (1997). Visual signal detectability with two noise components: Anomalous masking effects. *Journal of the Optical Society of America A*, 14, 2420–2442. [PubMed]
- Burgess, A. E., Wagner, R. F., Jennings, R. J., & Barlow, H. B. (1981). Efficiency of human visual signal discrimination. *Science*, 214, 93–94. [PubMed]
- Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology (London)*, 197, 551–566. [PubMed]

- Chen, C. C., & Foley, J. M. (2004). Pattern detection: Interactions between oriented and concentric patterns. *Vision Research*, 44(9), 915–924. [PubMed]
- Eckstein, M. P., Ahumada, A. J., Jr., & Watson, A. B. (1997). Visual signal detection in structured backgrounds: II. Effects of contrast gain control, background variations, and white noise. *Journal of the Optical Society of America A*, 14(9), 2406–2419. [PubMed]
- Foley, J. M. (1994). Human luminance pattern–vision mechanisms: Masking experiments require a new model. *Journal of the Optical Society of America A*, 11, 1710–1719. [PubMed]
- Foley, J. M., & Chen, C. C. (1999). Pattern detection in the presence of maskers that differ in spatial phase and temporal offset: Threshold measurements and a model. *Vision Research*, 39(23), 3855–3872. [PubMed]
- Foley, J. M., & Legge, G. E. (1981). Contrast detection and near-threshold discrimination in human vision. *Vision Research*, 21, 1041–1053. [PubMed]
- Fukunaga, K. (1990). *Introduction to statistical pattern recognition* (2nd ed.). San Diego: Academic Press.
- Geisler, W. S. (2003). Ideal observer analysis. In L. Chalupa & J. Werner (Eds.), *The visual neurosciences*. Boston: MIT press.
- Graham, N., & Nachmias, J. (1971). Detection of grating patterns containing two spatial frequencies: A comparison of single-channel and multiple-channels models. *Vision Research*, 11, 251–259. [PubMed]
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Heeger, D. J., Simoncelli, E. P., & Movshon, J. A. (1996). Computational models of cortical visual processing. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 623–627. [PubMed] [Article]
- Hood, D. C., & Finkelstein, M. A. (1986). Sensitivity to light. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance*, Vol. 1. New York: Wiley.
- Klein, S. A. (2001). Measuring, estimating, and understanding the psychometric function: A commentary. *Perception & Psychophysics*, 63(8), 1421–1455. [PubMed]
- Klein, S. A., & Stromeyer, C. F., III. (1980). On inhibition between spatial frequency channels: Adaptation to complex gratings. *Vision Research*, 20(5), 459–466. [PubMed]
- Legge, G. E., & Foley, J. M. (1980). Contrast masking in human vision. *Journal of the Optical Society of America A*, 70, 1458–1471. [PubMed]
- Legge, G. E., Kersten, D., & Burgess, A. E. (1987). Contrast discrimination in noise. *Journal of the Optical Society of America A*, 4(2), 391–404. [PubMed]
- Meese, T. S., Hess, R. F., & Williams, C. B. (2005). Size matters, but not for everyone: Individual differences for contrast discrimination. *Journal of Vision*, 5(11), 928–947, <http://journalofvision.org/5/11/2/>, doi:10.1167/5.11.2. [PubMed] [Article]
- Mostafavi, H., & Sakrison, D. (1976). Structure and properties of a single channel in the human visual system. *Vision Research*, 16, 957–968. [PubMed]
- Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2005). Classification images predict absolute efficiency. *Journal of Vision*, 5(2), 139–149, <http://journalofvision.org/5/2/5/>, doi:10.1167/5.2.5. [PubMed] [Article]
- Nachmias, J. (2002). Contrast discrimination with and without spatial uncertainty. *Vision Research*, 42(1), 41–48. [PubMed]
- Nachmias, J., & Sansbury, R. V. (1974). Letter: Grating contrast: Discrimination may be better than detection. *Vision Research*, 14, 1039–1042. [PubMed]
- Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research*, 45(2), 205–231. [PubMed]
- Nykamp, D. Q., & Ringach, D. L. (2002). Full identification of a linear–nonlinear system via cross-correlation analysis. *Journal of Vision*, 2(1), 1–11, <http://journalofvision.org/2/1/1/>, doi:10.1167/2.1.1. [PubMed] [Article]
- Pelli, D. G. (1981). *Effects of visual noise*. PhD dissertation, Cambridge University, Cambridge.
- Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A*, 2(9), 1508–1532. [PubMed]
- Pelli, D. G., Farell, B., & Moore, D. C. (2003). The remarkable inefficiency of word recognition. *Nature*, 423(6941), 752–756. [PubMed]
- Sachs, M. B., Nachmias, J., & Robson, J. G. (1971). Spatial-frequency channels in human vision. *Journal of the Optical Society of America A*, 61, 1176–1186. [PubMed]
- Solomon, J. A. (2000). Channel selection with non-white-noise masks. *Journal of the Optical Society of America A*, 17(6), 986–993. [PubMed]
- Solomon, J. A. (2002). Noise reveals visual mechanisms of detection and discrimination. *Journal of Vision*, 2(1), 105–120, <http://journalofvision.org/2/1/7/>, doi:10.1167/2.1.7. [PubMed] [Article]
- Solomon, J. A., & Pelli, D. G. (1994). The visual filter mediating letter identification. *Nature*, 369, 395–397. [PubMed]
- Stromeyer, C. F., III., & Klein, S. (1974). Spatial frequency channels in human vision as asymmetric (edge) mechanisms. *Vision Research*, 14, 1409–1420. [PubMed]

- Tanner, W. P., & Birdsall, T. G. (1958). Definitions of  $d'$  and  $\eta$  as psychophysical measures. *Journal of the Acoustical Society of America*, *30*, 922–928.
- Thomas, J. P., & Knoblauch, K. (2005). Frequency and phase contributions to the detection of temporal luminance modulation. *Journal of the Optical Society of America A*, *22*(10), 2257–2261. [[PubMed](#)]
- Thomas, J. P., & Olzak, L. A. (1996). Uncertainty experiments support the roles of second-order mechanisms in spatial frequency and orientation discriminations. *Journal of the Optical Society of America A*, *13*(4), 689–696. [[PubMed](#)]
- Tjan, B. S., & Legge, G. E. (1998). The viewpoint complexity of an object-recognition task. *Vision Research*, *38*, 2335–2350. [[PubMed](#)]
- van Nes, F. L., Koenderink, J. J., Nas, H., & Bouman, M. A. (1967). Spatiotemporal modulation transfer in the human eye. *Journal of the Optical Society of America A*, *57*(9), 1082–1088. [[PubMed](#)]
- Watson, A. B., & Solomon, J. A. (1997). Model of visual contrast gain control and pattern masking. *Journal of the Optical Society of America A*, *14*, 2379–2391. [[PubMed](#)]
- Wilson, H. R., McFarlane, D. K., & Phillips, G. C. (1983). Spatial frequency tuning of orientation selective units estimated by oblique masking. *Vision Research*, *23*(9), 873–882. [[PubMed](#)]