Classification images for simple detection and discrimination tasks in correlated noise

Craig K. Abbey1,2,* and Miguel P. Eckstein1

1Department of Psychology, University of California, Santa Barbara, Santa Barbara, California 93106, USA
2Department of Biomedical Engineering, University of California, Davis, Davis, California 95616, USA
*Corresponding author: abbey@psych.ucsb.edu

Received June 4, 2007; revised August 7, 2007; accepted August 17, 2007;
posted September 5, 2007 (Doc. ID 83574); published October 8, 2007

We use the classification image technique to investigate the effect of white noise and various correlated Gaussian noise textures (low-pass, high-pass, and band-pass) on observer performance in detection and discrimination tasks. For these tasks, performance is generally enhanced by an observer’s ability to “prewhiten” correlated noise as part of the formation of a decision variable. We find that observer efficiency in these tasks is well represented by the measured classification images and that human observers show strong evidence of adaptation to different correlated noise textures. This adaptation is captured in the frequency weighting of the classification images. © 2007 Optical Society of America

1. INTRODUCTION

When viewed as realizations of a random process, many classes of images exhibit strong spatial correlations. Efforts to characterize “natural” images, which make up the majority of our visual experience, suggest that the covariance in intensity between points separated by different distances follows an approximate 1/f² power spectrum in the spatial frequency domain [1–7]. Spatial correlations are also an important component of medical images [8,9], where correlations can arise from sources within the imaging system such as quantum mottle, image processing, or from anatomical variability in the patients being imaged [10]. For example, the power spectrum of noise in computed tomography images (absent any apodization) roughly follows a ramp in spatial frequency [11–13]. Burgess [14] has recently shown that breast anatomy is well characterized by a 1/f² power spectrum in x-ray mammography, which exerts a strong influence on detection performance. Given the ubiquitous nature of spatial correlations, it is natural to ask whether the human visual system is able to incorporate knowledge of spatial correlation into visual processing, and if so, how. This question motivates the research reported here.

The basic question of whether humans can modify visual processing in response to spatial correlations—often referred to as “prewhitening”—has been largely settled over the last two decades. Myers et al. [15] investigated detection in high-pass noise and found that their observer performance data was best fitted by a model that does not use knowledge of spatial correlations. However, Fiete et al. [16] and Rolland and Barrett [17] found that for various low-pass noise textures, adaptive models were needed to predict observer performance. Burgess and co-workers [18–20] investigated detection performance in noise textures with increasing power at low spatial frequencies. With reasonable assumptions about internal noise, he proved that human observers modify their detection strategy in response to the changing structure of correlation in the images [20].

The general result of many such studies has been that human observers are capable of some form of prewhitening for low-pass noise textures. This finding is further supported by results from Webster and Miyahara [21] that show clear changes in contrast sensitivity functions after adaptation to images with different correlation structures ranging from white to low-pass power-law fields. Presumably, if the vast majority of our visual experience takes place in the presence of low-pass correlations, we will have developed some ability to incorporate them into our visual processing. The contribution of this work is a better understanding of how such adaptation functions.

Performance studies of the sort described above have been effective at demonstrating the capability of human observers to adapt to correlated noise. However, it has been much more challenging to specify precisely how observers adapt. This is due to the high dimensionality of images that leave open a large number of possible degrees of freedom in how a decision variable is formulated for a visual task. Performance studies that explore a handful of experimental settings (often between 5 and 20) have a limited ability to resolve flexible adaptation precisely. Nonetheless, models based on visual channels have been used to explain partial adaptation to correlated noise in this setting [22–24] These “channelized” models typically specify a set of frequency selective channels and then allow some form of prewhitening to occur on the channel responses rather than on the image that generated them. For example, the channelized Hotelling observer [22–25] proceeds in this fashion with an optimal decorrelation of channel responses. However, this approach has been used with a variety of channel profiles (rectangular frequency bands, difference of Gaussians, Laguerre–Gauss basis functions, and Gabors) and a large range in the number of
channels considered (3 to more than 40). In addition, implementations have sometimes included channel-dependent internal noise [26], which changes the combination rule. As a result, the channelized Hotelling observer model is less a single model than a family of models, and it is not entirely clear to what extent channels explain human-observer adaptation and to what extent they simply fit empirical data.

An alternative to analyses of observer performance became available with the introduction of classification images to the field of vision. Originally used by Ahumada and colleagues to study audition in the 1970s [27,28], the approach has found widespread application in vision since its first application to Vernier acuity tasks by Ahumada and Beard [29,30]. Essentially, the methodology examines components of stimuli that correlate with the observed response and as such can be thought of as reverse correlation applied to psychophysical data [31]. By utilizing the noisy stimuli as well as the response, classification image analysis capitalizes on a much richer observation than traditional methods that use only the response. Originally developed for yes/no tasks in white noise, the approach has been generalized to the commonly used two-alternative forced-choice (2AFC) experimental paradigm and correlated Gaussian noise textures used here [32–34]. Under the assumption of a locally linear observer, the classification image can be interpreted as an estimate of the spatial weights used by the observer to formulate a decision variable. Simple (i.e., signal-known-exactly) detection and discrimination tasks with small focal targets appear to agree with this linearity assumption [35], and thus the classification image approach seems ideally suited to studying adaptation to spatial correlation with these signals as targets.

The research described here expands on our previous work in this area [36]. We use the classification image methodology to evaluate simple detection and discrimination tasks masked by noise with various textures. We evaluate white noise, low-pass noise, and high-pass noise (WN, LN, and HN, respectively, in the figures) in both tasks. In the discrimination task we evaluate an additional weighting of the subjects. We also evaluate the linear classification images for these tasks provide a more detailed view of how that adaptation is reflected in the spatial weighting of the subjects. We also evaluate the linearity assumption inherent in our interpretation of the classification images.

2. THEORY
A total of seven detection and discrimination experiments form the basis for the results reported in this work. In this section we describe the stimuli and the estimation procedure for obtaining a classification image, and we also develop methods for analyzing subject classification images in correlated noise. These include a method to use the classification image to predict subject efficiency, which extends the approach derived by Murray et al. [37] to correlated noise, and an approach to analyzing the spatial frequency content of classification images.

A. Stimulus Properties
Throughout this work we will denote quantities in the image domain as a column vector of pixel intensities. The stimuli for all experiments are generated by adding a target profile, \( p_T \), or nontarget alternative profile, \( p_A \), to noise fields, \( n \), with different correlation structures. An image, \( g \), can be thought of as arising from the target or alternative classes according to the generating equations,

\[
\text{Target: } g^+ = p_T + n^+, \quad \text{Alternative: } g^- = p_A + n^-,
\]

where the superscript (+ or −) indicates the class. The noise is constrained to be zero-mean, and hence \( p_T \) and \( p_A \) represent the ensemble mean of each class. The noise is statistically equivalent in the two classes, so \( n^+ \) and \( n^- \) are independent realizations of the same process. Noise is also independent across different trials in psychophysical experiments. When needed, we will add an additional subscript, \( j \), to \( g \) and \( n \) indicating the trial.

This work explores two different tasks. In the detection task (Dt in the figures), \( p_T \) is a Gaussian profile with a spatial \( \sigma \) of 3.0 pixel units added to a uniform luminance background. At this stage we will use pixel units to describe length in the stimuli. These will be converted to degrees of visual angle after we describe our monitor and psychophysical experiments below. The mean alternative profile, \( p_A \), is simply the uniform background. The second task is the discrimination task between two Gaussian profiles, each atop a uniform background. The profiles of the Gaussians are different; one is narrower (\( \sigma = 3.2 \) vs \( 4.0 \) pixels) and more intense (contrast is greater by a factor of 1.56) than the other. To distinguish this from other discrimination tasks (contrast, discrimination, size discrimination, etc.) we refer to it as an identification task (Id in the figures) since the observer is asked to choose the alternative with the narrower more intense profile. Images of these mean profiles can be seen in Fig. 1A.

The difference signal, defined as \( s = p_T - p_A \), for each task is plotted in the spatial and spatial-frequency domains in Fig. 1B. The difference signal for the detection task is simply the Gaussian target; differencing just removes the background in this case. This results in a Gaussian frequency profile as well. For the identification task, the difference signal is a difference of Gaussians (DOG) with a positive region near the center, surrounded by a negative region. The frequency profile of this difference is a relatively broad band that peaks near 0.06 cycle/pixel.

Noise fields are generated by filtering white noise to achieve a desired correlation structure. Four correlation structures are considered: WN, LN, HN, and BN. Experiments cover all four correlation structures with both tasks except for band-pass noise, which is paired only
with the identification task, for a total of seven experiments. Example noise fields for each of these correlation structures can be seen in Fig. 1C. All four correlation structures are approximately isotropic (up to sampling effects), and therefore the correlation between any two pixels is dependent only on the distance between them. By definition, correlation is zero for any distance greater than zero for white noise. The low-pass noise has strong positive correlations that are long in comparison with the difference signals. High-pass noise consists of relatively short correlation lengths, with anticorrelation between adjacent pixels. Band-pass noise consists of relatively weak correlations that oscillate with distance.

The correlation structures are all defined by their discrete noise power spectrum (NPS) as shown in Fig. 1D. Let \( \Sigma \) be the covariance matrix associated with a given noise texture; then this covariance matrix is diagonalized by a 2D finite Fourier transform (FFT) with diagonal elements contained in the diagonal elements of the matrix \( \Lambda \),

\[
\Sigma = F_{2D}^H \Lambda F_{2D},
\]

where \( F_{2D} \) is the 2D FFT in matrix form, and the subscript \( H \) indicates the Hermitian (complex conjugate transpose) of the matrix. For a normalized Fourier transform, \( F_{2D} \) is unitary, and hence the Hermitian is the inverse transform. Diagonal elements of \( \Lambda \) constitute the discrete NPS of the noise.
The isotropic correlation structure of the noise fields leads to a rotationally symmetric NPS. The magnitude of all correlation structures were scaled so that the root-mean-squared (RMS) contrast relative to the uniform background was 15%. The NPS of white noise is constant,

$$[A_{WN}]_{kl} = N_{WN},$$  \hspace{1cm} (3)

where $k$ indexes the frequency components of the image. The NPS of the low-pass noise has a functional form of

$$[A_{LN}]_{kl} = N_{LN} \left( 1 + \frac{\alpha}{1 + (\rho_k/2\rho_0)^3} \right), \hspace{1cm} (4)$$

where $\rho_0$ was set to the frequency of the first harmonic ($\rho_0=1/64$ for $64 \times 64$ pixel images), and $\alpha$, the parameter controlling the relative strength of the correlated component, was set to 565.5. The second term in Eq. (4) falls off approximately as $\rho_k^{-3}$, as $\rho_k$ gets large relative to $\rho_0$. This gives the NPS a power-law form at low- to mid-range frequencies where $\rho_k$ is larger than $\rho_0$, but not so large that the NPS is dominated by the first term. At higher frequencies, the NPS asymptotes to a constant value. The NPS of the high-pass texture is essentially the inverse of the low-pass NPS scaled to equate RMS contrast,

$$[A_{HN}]_{kl} = \frac{N_{HN}}{1 + \alpha/(1 + (\rho_k/2\rho_0)^3)}, \hspace{1cm} (5)$$

Here the low- to mid-range frequencies are approximately a power law (with an exponent of 3) that also asymptotes to a constant value at higher frequencies. The low-pass noise texture is essentially a white-noise field with a small band of frequencies that have been amplified up to a factor of 20,

$$[A_{BN}]_{kl} = N_{BN} \left( 1 + 20 \exp \left( -\frac{1}{2} \frac{(\rho_k - \rho_0)}{\sigma_{BN}} \right)^2 \right). \hspace{1cm} (5')$$

The center frequency of the band is $\rho_0=0.0627$, and the bandwidth is fairly narrow with $\sigma_{BN}=0.0078$. These parameters were chosen to concentrate this band of high noise power precisely in the frequencies with the largest difference signal component for the identification task. Plots of the NPS as a function of spatial frequency are given in Fig. 1D.

B. Observer Models

We use two models of visual task performance as a reference for comparing human observers who differ in how they handle noise correlations. The models assume that in each trial a decision variable, $\lambda_i$, is formed for both the target image, $\lambda^+$, and the alternative image, $\lambda^-$. A correct decision is made when $\lambda^+ - \lambda^- > 0$.

We assume a linear observer, which means that the decision variable is formed by a weighted sum of the stimulus with additive internal noise. For a generic stimulus (either + or −), the functional form of a linear observer is given by

$$\lambda = w^T g + \varepsilon, \hspace{1cm} (6)$$

where $w$ is a column vector representing the spatial weights used by the observer and $\varepsilon$ is a scalar internal noise component. In this model, classification images can be interpreted directly as an estimate of $w$ [26]. The internal noise component, $\varepsilon$, is assumed to be a zero-mean Gaussian random variable with variance $\sigma^2$. It is also assumed to be independent between the two classes (+ and −) and independent across trials. The linearity assumption applies rigorously for the two specific models we use. For humans, we will assess the linearity assumption through a number of different methods described in the Results and Discussion section below.

Since the goal of this work is to investigate the effect of correlation structure on visual processing, we consider a model that fully adapts to the correlation structure of the noise and one that does not adapt. Adaptation to noise structure is often cast as the ability to "prewhiten" correlations in the noise before applying a matched filter, and hence we will follow convention and refer to these as prewhitening or nonprewhitening observer models [15,17,25]. The nonprewhitening matched filter uses the difference signal as its spatial weighting,

$$w_{NPW} = s. \hspace{1cm} (7)$$

By contrast, the spatial weighting of the prewhitening matched filter transforms difference signal by the inverse covariance,

$$w_{PW} = \Sigma^{-1}s. \hspace{1cm} (8)$$

The prewhitening and nonprewhitening observer models converge in the case of white noise where the covariance matrix is proportional to the identity matrix.

For stationary Gaussian textures of the sort used in this work, the prewhitened matched filter (absent any internal noise) is equivalent to the ideal observer [16]. Thus the prewhitened matched filter represents the optimal spatial weighting for the task. Figure 2A shows the prewhitened matched filter for all seven tasks. Within both the detection and identification tasks, there is considerable difference in the optimal spatial weighting across correlation structures. These differences are highlighted by the radial frequency profile plots in Figs. 2B and 2C. These show that the optimal weighting in the presence of low-pass noise in these experiments is generally to shift to higher spatial frequencies. In the presence of high-pass noise, the opposite holds, with more weight being shifted to lower spatial frequencies. For band-pass noise in the identification task, the optimal strategy is to form a bimodal profile with peaks on either side of the noise band.

C. Estimation of Classification Images

Our approach to estimating classification images from 2AFC experiments has been described in detail previously [32,34], and hence we will provide only a brief review for clarity here.

At every trial of a 2AFC experiment, two stimuli are presented and an observer response is recorded. We typically convert the response to a trial outcome, or score, by assigning it a value of 1 if the observer responded correctly and 0 otherwise. For each of the stimuli, we will presume that the noise fields used have been stored in some form and are therefore available for analysis. Thus at the $j$th trial ($j=1, \cdots, N_T$), observed random variables are the outcome, $o_j$, the target noise field, $n^+_j$, and the alternative noise field, $n^-_j$. The average of the outcomes is
used as an estimate of proportion correct, $\hat{P}_C$, for the experiment. We form an estimate of the classification image by considering a combination of these terms [32,34],

$$\Delta q_j = \frac{N_T}{N_T - 1}(o_j - \hat{P}_C)\Sigma^{-1}\Delta n_j,$$

where $\Delta n_j = n_j^+ - n_j^-$ is the difference in noise fields. The sign and magnitude of the trial weight term, $o_j - \hat{P}_C$, depends on whether the outcome was correct or not and the overall level of performance in the task. The initial term involving $N_T$ on the right-hand side of Eq.(9) corrects for degrees of freedom in the trial weight and is approximately 1 for the large number of trials used in classification image experiments.

It has been shown [34] that under the linear assumptions in Eq.(6), the expectation of $\Delta q_j$ (over both external and internal noise) is proportional to the spatial weights, $w$, by the relationship

$$\langle \Delta q \rangle_{n,e} = \frac{e^{-d'/2\lambda^2}}{\sqrt{\pi(\lambda^2 w^2 + \sigma_e^2)}} w.$$

(10)

The constant of proportionality incorporates the spatial weights and is also dependent on the performance of the observer through the detectability index [38], $d'$, defined under the linear model [32] in Eq. (6) as

$$d' = \frac{w^s}{\sqrt{\lambda^2 w^2 + \sigma_e^2}}.$$

(11)

Under Gaussian assumptions on $\lambda$, detectability is related to proportion correct through the cumulative normal distribution function,

$$P_C = \Phi(d'/\sqrt{2}).$$

(12)

For human observers, the expectation in Eq. (10) is estimated by a sample average over all the trials,

$$\Delta q = \frac{1}{N_T} \sum_{j=1}^{N_T} \Delta q_j.$$

(13)

Equation (10) is also useful for predicting classification images using linear models [39] and matching them to human-observer data. For a given model $w$, this can be accomplished by first setting the internal noise variance, $\sigma_e^2$ in Eq.(11) to match human-observer performance measured as $d'$, determined by solving Eq. (12), and then computing the right-hand side of Eq.(10).

D. Efficiency with Respect to the Ideal Observer

The previous subsection specified both proportion correct and detectability as measures of performance. Another useful metric of performance is efficiency with respect to the ideal observer [40–42]. Taking a measure of performance that is relative to the ideal observer allows normalization for the amount of relevant information in the stimuli. Thus efficiency reports how well the available information is being used.

Efficiency is generally defined [40] as the ratio of the threshold signal energy needed to achieve a target level of performance by the ideal observer to the same quantity for the observer under study (e.g., a human observer). Under the linear model described in Eq. (6) and the Gaussian assumptions leading to Eq. (11), it is equivalent [41] to compute efficiency, $\eta$, from $d'$ values for the ideal observer, $d'_{IO}$, and the observer of interest, $d'_{Obs}$, as

$$\eta = (d'_{Obs}/d'_{IO})^2.$$

(14)

Under these assumptions, where the ideal observer is linear as in Eq. (8), ideal observer detectability is given by
processing. For these reasons, we will analyze our frequency-selective mechanisms as the basis for visual tion, many important models rely heavily on spatial-characterized in the spatial-frequency domain. In addi-

Fig. 3 shows that the tasks used in this work are well 
made use of empirical properties of the observed classifi-
cation images to reduce the effects of estimation error.

E. Spatial Frequency Analysis of Classification Images

The two main assumptions in this analysis are that the spatial support of the classification images (up to estimation error) is of limited spatial extent and that the classification images are rotationally symmetric. Both of these assumptions are discussed below in the Results and Discussion section.

Figure 3 shows the conceptual steps used to obtain a spatial-frequency plot from a classification image. The classification image has a spatial window applied to it (fourth-order Butterworth window with width of 0.6° of visual angle). This window attenuates noise outside the central area of the classification image. The windowed classification image is then transformed to the frequency domain by a 2D FFT oriented to the center of the image. This produces both real and imaginary Fourier components. However, under the assumption of a rotationally symmetric classification image, the imaginary components simply represent antisymmetric estimation error and can be neglected. Relative to the real component, the imaginary component of the 2D FFT in Fig. 3 does not appear to contain a substantial signal. The final step of the process is a radial averaging of the real component to get a 1D representation of spatial frequency. The entire process can be thought of as a linear transformation of an image (a 4096-dimensional vector) to a radial plot with 33 dimensions.

While Fig. 3 shows the conceptual process for obtaining a spatial-frequency plot from a classification image, it is actually somewhat advantageous to perform the operation on each \( \Delta q \) vector used to estimate the classification image in Eq. (9) as described by Abbey and Eckstein [32]. This allows us to also estimate an error covariance matrix on the radial averages that can be used to generate the error bars seen in the plots and for testing various hypotheses.

3. METHODS

A. Stimulus Generation

Stimuli for the observer studies were created by generat-
ing an initial image of Gaussian white noise and then fil-
tering this to produce one of the power spectra seen in
Fig. 1. The filtering was done by multiplication in the FFT domain by the square root of the NPS and then inverse transforming back to the spatial domain, where the target or alternative profile was added. Because of the cyclical nature of the FFT, spatial correlations in the images “wrap around” the edges of the images. It would be somewhat preferable not to have these cyclical correlations, which are not representative of natural images in general. However, using the FFT convolution allows us to perform the necessary ideal-observer computations with the inverse of the large covariance matrix in Eq. (9) relatively easily using Fourier methods. Since the target is located in the center of the image, we assume that the edge effects do not influence observer strategy or performance significantly.

**B. Psychophysical Studies**

We report the results of three psychophysical studies designed to evaluate human-observer templates for detecting a Gaussian “bump” signal in the high-pass, low-pass, and white-noise images described above. All experiments were conducted on a DOME monitor, photometer calibrated to a linear luminance scale in a darkened room. The monitor luminance ranged from 0.0 to 79.1 cd/m², with a mean luminance of 31.3 cd/m² corresponding to 100 gray levels (gl) in all experiments. Noise contrast, defined as the ratio of the pixel standard deviation to the mean luminance, was fixed in all conditions at 15% (a pixel standard deviation of 15.0 gl or 4.7 cd/m²). Pixel size was 0.3 mm on the monitor screen, and a viewing distance of approximately 1 m was maintained throughout the experiments by placing a table between the observer and the monitor. With this experimental setup, the width (and height) of a pixel occupied 0.017 deg (1.03 arcmin) of visual angle. Spatial frequencies ranged between the horizontal and vertical Nyquist limits of ±29.1 cycles per degree visual angle (cpd).

In each trial, the two alternatives were displayed sequentially for 500 ms each as shown in Fig. 4, with the rest of the monitor set to the mean background level except for the upper left corner where an image of the mean target was displayed for reference. The time interval containing the signal-present image was randomized across trials with both intervals having equal probability. Each trial began with 1000 ms of lead-in time initiated by a mouse click from the observer. During this time, a blank image at average luminance (100 gl) was displayed. After this lead-in, the first stimulus was displayed for 500 ms and then overwritten by an interstimulus white noise image, which was displayed for 1000 ms to disrupt any persistence effects. The white noise image was then overwritten by the second image of the trial. After 500 ms of display time for the second image, the white noise image was used again to overwrite the stimulus image, and the observer was queried for a decision. Decisions were obtained from the observer by a mouse click in the appropriate side of the image (left side for the first displayed image and right side for the second displayed image). While there was no time limit for obtaining an observer response, the observers typically rendered a decision less than 1 s after the final image in the trial sequence was displayed. The forced-choice image display software was written in the IDL development environment (Research Systems Inc., Boulder, Colorado) by the authors.

Three subjects participated in all three experiments. Two of the subjects, DV and CH, were naive to the goals of the research. The other subject (CA) is one of the authors. All observers have had prior experience in visual tasks of the sort reported here, with normal or corrected visual acuity. The observers began each study by completing 4 training sessions of 50 trials, with decreasing signal amplitude in each session. This initial training was followed by the evaluation of a psychometric curve for each observer consisting of 10 sessions of 100 trials at 5 different fixed contrast levels. The results of the psychometric function were used to evaluate the linearity of the observers (see below) and to determine the signal amplitude for the classification image study so that observer performance would be approximately 85% correct (detectability = 1.5). A total of 2,000 trials were used in the white, low-pass, and band-pass noise conditions. In high-pass noise, initial results showed that large errors at lower spatial frequencies would make interpretation difficult, and hence a total of 4000 trials/study were used. In all cases, the trials were divided into sessions of 100 trials. Each study took two to four days of observer time (observers were limited to two hours per day of experimental results with regular breaks to avoid fatigue), and on subsequent days, observers began with a short session of 10 trials to refresh them to the task.

**4. RESULTS AND DISCUSSION**

In this section we present results of the seven detection and discrimination experiments described above. We begin by looking at performance in terms of efficiency with respect to the ideal observer. We then present classification images and an analysis of the assumption of a linear decision variable. These are followed by a spatial-frequency analysis of the classification images.

**A. Observer Performance Results**

Parameters of the various detection and identification tasks are listed in Table 1, along with the average perfor-
Table 1. Parameters of Experiments

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Dt: WN</th>
<th>Dt: LN</th>
<th>Dt: HN</th>
<th>Id: WN</th>
<th>Id: LN</th>
<th>Id: HN</th>
<th>Id: BN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target Width</td>
<td>0.12°</td>
<td>0.12°</td>
<td>0.12°</td>
<td>0.13°</td>
<td>0.13°</td>
<td>0.13°</td>
<td>0.13°</td>
</tr>
<tr>
<td>Alternative Width</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.16°</td>
<td>0.16°</td>
<td>0.16°</td>
</tr>
<tr>
<td>Target Contrast</td>
<td>6.2%</td>
<td>21.3%</td>
<td>2.4%</td>
<td>19.3%</td>
<td>53.6%</td>
<td>7.8%</td>
<td>36.2%</td>
</tr>
<tr>
<td>Alternative Contrast</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>12.3%</td>
<td>34.3%</td>
<td>5.0%</td>
<td>23.1%</td>
</tr>
<tr>
<td>Average PC</td>
<td>84% ±3%</td>
<td>83% ±2%</td>
<td>83% ±2%</td>
<td>84% ±3%</td>
<td>85% ±4%</td>
<td>88% ±3%</td>
<td>85% ±2%</td>
</tr>
</tbody>
</table>

*Listed are the width (full width at half-max. in degrees), contrast (%) of the Gaussian profiles used as target and alternative in the classification image experiments, and performance (percent correct ±1 std. dev.) averaged across subjects.

mance across the three subjects (±1 standard deviation). The contrast values, which vary considerably between different experiments, were determined from pilot psychometric studies. It is also clear the pilot studies were reasonably accurate in determining a target contrast that would result in 85% correct on average. The average percent correct over all observers in all experiments is 84.5% (range: 80.1% to 90.4%), although there appears to be some variation depending on experiment, as shown in the table, and observer (average performance across observers ranges from 82.1% to 87.0%). While these ranges do indicate some residual variability, they are generally consistent with the targeted 85% correct level determined from the pilot experiments.

By contrast, the efficiency data in Fig. 5 have a substantially larger range, predominantly due to the varia-

tion across experiments (average efficiency ranges from 1.9% to 50.4%). The highest efficiency is observed in the white noise and low-pass noise experiments, where average efficiency is over 40%. Observers were substantially less efficient in the high-pass noise experiment where average efficiency is less than 10%. Discrimination in band-pass noise had somewhat reduced efficiency at 28.9%. This large range of efficiency indicates that observers are not able to equally access relevant information with different noise textures, and it replicates the low-efficiency findings of Barrett and co-workers for high-pass noise [25]. The purpose of the classification image analysis is to see to what extent this phenomenon can be explained by the spatial weighting used by the observers.

We have also investigated the effects of interval bias [46]. In our 2AFC performance data, defined as proportion correct for targets presented in the first interval minus proportion correct for targets presented in the second interval. Bias values are generally small with the largest observed values less than 0.11 in percent correct units (data not shown). Bias correction following the approach of Klein [46] results in at most a 1.2 percentage point change in efficiency, and hence we will neglect bias in our analysis.

B. Classification Images

The classification images estimated from our psychophysical studies are displayed in Fig. 6. It is clear from this figure that the estimates themselves are noisy quantities despite the relatively large number of trials that constituted each experiment. As described previously [34], the correlation structure of errors in the estimated classification images is inversely related to the correlation structure of the noise in the stimuli; the low-pass noise
The experiment produces classification-image estimates with high-pass errors, and the high-pass noise produces classification-image estimates with low-pass errors.

All of the classification images appear to have a central area of higher intensity indicating positive weighting of the image in this area. In some cases there is a visibly darker region surrounding the center. This would indicate negative weighting or inhibition from the surrounding area. These issues will be examined in more detail below after first considering the linearity assumption that underlies much of classification image analysis.

C. Linearity of the Decision Variable
A crucial point for interpreting classification images is the linearity of the decision variable. If the decision variable is effectively linear [47], as in Eq. (6), then the classification image is an unbiased estimate of the spatial weighting used to perform the task. If observers employ a highly nonlinear decision variable, it may still be possible to observe a significant classification image. However, it will not be as clear how that classification image relates to the decision variable or even whether it captures all the spatial information influencing task performance (in this case higher-order methods may be appropriate [31,48,49]). For this reason, a goal in the design of stimuli for these experiments was to induce, to the extent possible in the subjects, linear decision variables. In this section we evaluate linearity by three methods shown in Fig. 7.

As described above, we collected psychometric functions for each observer in each experimental condition. One hallmark of a linear decision variable is a linear relationship between detectability ($d'$) and target contrast. Burgess [35] uses the negative offset of a line fitted to the psychometric function as a measure of deviation from linearity. A strict linear relationship requires that such a line pass through the origin (i.e., $d'=0$ at zero contrast).

We have fitted two-parameter lines (slope and offset) to our psychometric data, and we reject linearity if the offset is significantly less than zero. Figure 7A shows an example of psychometric functions for the three subjects in one of the experiments. Here, two of the subjects (DV and CA) are well fitted by a line through the origin, but the third (CH) is not. A summary of the p-values for a one-sided test (offset <0) is given in Table 2.

A second test for nonlinearity uses the classification image data themselves. This test compares the classification images derived from the target and alternative noise fields separately. As described by Abbey and Eckstein [32], under a linear decision variable the two should be identical up to estimation error. They also propose tests for significant differences based on the Hotelling $T^2$ statistic. Figure 7B gives the frequency domain plots of subject CH for the detection task in the white noise. Shown are the classification image derived from all the data along with the classification images derived from the target and alternative data only. Also plotted is the difference between the target and alternative classification images, which is significantly different from zero in this case.

Our final assessment of linearity assesses how well classification images predict the absolute efficiency of the observer, using the formula in Eq. (16). Accurate prediction of efficiency indicates that the linear classification image is effectively describing the relevant components of the observer decision variable, consistent with a linear decision variable. Figure 7C shows a scatter plot of absolute and predicted efficiency for all subjects and experiments. We see generally good agreement between the two measures, indicating consistency between the linear classification images and the measured efficiency of our subjects.

A summary of the statistical inference results for all three tests for a nonlinear decision variable is found in

![Fig. 7. Three tests of a linear decision variable. A, Psychometric functions ($d'$ versus peak signal contrast) from the detection task in high-pass noise. These data are considered evidence against a linear decision variable when the y-axis intercept of a linear fit is significantly different from zero. This experiment was chosen because subject CH has a nonlinear psychometric function by this criterion. B, Example of the classification image test for nonlinear decision variables. The null hypothesis is that the difference between an estimated template derived from target-only and alternative-only noise fields is zero. This example (subject CH for the detection task in white noise) has also been chosen because the null hypothesis is rejected. C, Template efficiency test for a linear decision variable for all experiments and subjects. The null hypothesis is that absolute efficiency (computed from the subject's performance) and efficiency predicted from the template by Eq. (16) are equal.](image-url)
Table 2, which lists the number of tests with nominally significant values ($p < 0.05$). Of the 63 total comparisons, only six achieve statistical significance. However, these values have not been corrected for multiple comparisons and hence must be interpreted with some care. A Bonferroni correction over the 21 comparisons in each test requires a putative $p$-value of 0.0024, leaving only one case where significance is achieved (noted by the asterisk). There is also little correspondence between the tests, with only one case (Detection: HN) where significance is achieved in more than one test. Given the limited number of cases where linearity can be rejected and the low correspondence between the different tests, we consider the linearity assumption to be a reasonable approximation in these tasks. However, we note that better characterization and understanding of relationships among the various tests is needed for a more complete assessment of the linearity assumption.

D. Radial Plots of Observer Templates

Figures 8 and 9 show the spatial frequency content of the classification images after radial averaging in the Fourier domain as described in Fig. 3. We also plot the difference

![Figure 8](image-url)
signal and the template of the ideal-observer for comparison. In the case of white noise, the difference signal and the ideal observer template are identical, and hence we simply plot the difference signal. Human-observer data is plotted with error bars representing one unit of standard error. As in the template estimates in Fig. 6, the standard errors are highly influenced by correlations in the noise, but they tend to fall off as spatial frequency gets higher because the radial average contains more points. Even with averaging, high-pass-noise standard errors are quite large at low frequencies, reflecting a lack of stimulus noise in these frequencies. At zero frequency, template estimates are not plotted because the magnitude of error renders them uninterpretable.

For the white noise experiment (the first row of Figs. 8 and 9), the human-observer template deviates by a small amount from the difference signal at low spatial frequencies. For the detection experiment shown in Fig. 8, there is a drop in the weighting of the lowest spatial frequencies (<2 cpd), and for the identification experiment in Fig. 9 there is a slight boost in the lowest relevant frequencies (1–3 cpd). Both of these characteristics have been observed previously [50,39]. A comparison of spatial-frequency weighting in the low-pass experiments shows that human observers fall somewhere between the difference signal (i.e., the nonprewhitening matched filter) and the optimal prewhitening matched filter. Subjects do not appear to suppress low spatial frequencies (<5 cpd) as...
much as the ideal observer, yet the template values are considerably reduced from the difference signal. By contrast, human observers do not appear to weight the higher spatial frequencies (<5 cpd) as much as the ideal observer, yet they do considerably more than the difference signal would suggest. At spatial frequencies above 10 cpd, the low-pass noise plots have negative or oscillatory regions that likely reflect the effects of a limited spatial extent.

In the high-pass noise experiments, the human observers are closer to the difference signal at somewhat higher spatial frequencies (>4 cpd) and do not exhibit any of the high-frequency suppression found in the ideal observer. The high-pass experiment is somewhat difficult to interpret at low spatial frequencies because of the large error bars, despite doubling of the number of trials in the experiment. Nonetheless, in all cases we observe that spatial-frequency weights drop making a reasonably good match to the difference signal in the identification task. These results suggest that observers are not using the low-frequency data even though they are the most informative part of the stimuli and apparently can be used in other tasks such as detection in white noise. One possible explanation for this inconsistency may be the relatively low contrast of targets in the high-pass noise experiments. We have recently found evidence for low-frequency suppression at low contrast [51].

For the identification task in band-pass noise, none of the observers are able to synthesize the bimodal frequency weighting used by the ideal observer. Two of the three observers (DV and CH) appear to place more weight on the lower spatial frequencies, while the third (CA) appears to more heavily weight the high-frequency peak. These results are roughly consistent with an observer constrained to weighting a band of spatial frequencies, which is then roughly applied to one or the other of the two peaks used by the ideal observer.

Table 3 gives the p-values for hypothesis tests comparing the white-noise classification image to each of the correlated noise textures. There is a significant difference between white-noise and low-pass noise for all observers in both detection and identification. For high-pass noise, all observers but one show no significant difference between the white-noise and high-pass noise classification images. All subjects show a significant difference between the white-noise and band-pass noise classification images.

### E. Implications for Visual Processing

Our results show that human observers can modify the summation of spatial frequencies in response to correlations in the noise. However, this modification is context dependent, depending on target and alternative profiles, on the structure of the noise, and possibly on signal contrast. This seems appropriate from the perspective of natural-scene statistics, where the power-law spectrum makes them predominantly low pass. A visual system that has been adapted to work in such an environment would presumably be able to prewhiten and perform other optimal transformations of the image. Conversely, high-pass noise is not a common part of our visual experience, and hence the visual system may lack effective summation mechanisms for this case.

### F. Implications for Image Processing

Our results also suggest a principled approach to image processing to reduce the effect of noise. In low-pass noise, human observers are using low spatial frequencies relative to the ideal observer and underutilizing higher frequencies. In this case, we would predict that a contrast enhancement filter that suppressed low spatial frequencies of the signal and enhanced high spatial frequencies would tend to improve observer performance. For high-pass noise, observers are not adequately suppressing higher spatial frequencies, and hence we predict that suppressing them by filtering would be beneficial. Note that these predictions assume that the observer would not re-adapt their spatial weighting based on the filtering. We leave the evaluation of these predictions to future works, but note that contrast enhancement of images with significant low-frequency content such as digital mammograms is well recognized [52], as is smoothing in computed tomography to reduce the effect of ramp-spectrum quantum noise [53].

### 5. CONCLUSIONS

In noise-limited detection and discrimination tasks, human observers use different weightings of spatial information depending on the correlation structure of the noise. This idea has been supported previously [15–20] with performance and efficiency data. In this work we confirm and quantify the effect using the classification image methodology adapted to correlated noise to analyze 2AFC experiments. Classification images allow us to observe (linear) spatial weighting directly, as opposed to inferring it indirectly from performance effects.

We report results from a total of seven experiments looking at detection or discrimination of Gaussian targets in four different noise textures representing white, low-pass, high-pass, and band-pass correlation structures. In each task, target contrast was adjusted based on pilot studies to achieve approximately 85% correct in 2AFC tri-

### Table 3. Effect of Noise Correlation

<table>
<thead>
<tr>
<th>Subject</th>
<th>Detection</th>
<th>Identification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WN vs. LN</td>
<td>WN vs. HN</td>
</tr>
<tr>
<td>DV</td>
<td>0.001</td>
<td>0.28</td>
</tr>
<tr>
<td>CA</td>
<td>0.000</td>
<td>0.95</td>
</tr>
<tr>
<td>CH</td>
<td>0.000</td>
<td>0.15</td>
</tr>
</tbody>
</table>

*aHypothesis tests for differences in the frequency classification images plotted in Figs. 7 and 8. P-values less than the nominal value of 0.05 are in bold type.*
als with stimulus durations of 500 ms in each interval. Efficiency with respect to the ideal observer ranges from approximately 2% to over 50%, with the lowest values in the high-pass noise experiments and the highest values in the white- and low-pass noise experiments. Discrimination in the band-pass noise experiment had an intermediate efficiency around 30%.

The classification images show that observers are clearly tuning their sensitivity—albeit suboptimally—based on the signal and the texture of the noise. We find significant differences among classification images for different noise textures in both detection and discrimination tasks. Comparing results from the white-noise and low-pass noise experiments, we find that observers give more weight to higher spatial frequencies in the low-pass noise experiment. This is reasonable from a signal detection perspective since the low-pass texture has much more noise at low spatial frequencies, making them less informative. However, in comparison with the spatial weighting of the ideal observer, human observers are able only to partially tune this process. Nonetheless, it is clear that observers modify their spatial-frequency weighting in the presence of correlated noise.

In the high-pass noise experiments, lower spatial frequencies contain the majority of information needed to perform the task. While there are some difficulties measuring low spatial frequencies in classification images from high-pass noise, human-observer classification images do not appear to be any better tuned to these frequencies than the white-noise case. Therefore these studies replicate previous findings showing that human observers appear able to at least partially decorrelate (or prewhiten) low-pass noise but not high-pass noise. We argue that this makes sense from a natural-scene statistics, which are predominantly low pass and form the majority of our visual experience. We also hypothesize that the limitation in visual processing of high-pass noise may be internal noise at low spatial frequencies.

Optimal spatial weighting in discrimination with band-pass noise required a strongly bimodal weighting of spatial frequencies with peaks at 1.8 and 5.5 cpd. While all the observers’ classification images exhibited significant differences in band-pass noise compared to white noise, none showed evidence of a bimodal template. Instead, observers appear to adopt a unimodal weighting falling to one side or the other of the high-noise band. This also suggests that observers have limitations in their ability to prewhiten the correlations in the image, forcing them to choose a side of the noise band.

APPENDIX A: PREDICTED EFFICIENCY FOR CORRELATED NOISE
In this appendix, we derive the formula for predicted efficiency presented in Eq. (16) of the text. The derivation relies on two previous results [34]. The first is the ensemble mean of the classification image, given in Eq. (10). For now, let us denote this mean by the vector quantity \( \mathbf{\mu}_C \). The second is the covariance matrix of the estimated classification images which is well approximated as

\[
\Sigma_C = \frac{2P_C(1-P_C)}{N_T} \Sigma^{-1},
\]

recalling that \( \Sigma \) is the covariance matrix of the noise.

Now consider a cross correlation of the difference signal \( \mathbf{s} \), with the estimated classification variable \( \Delta \mathbf{q} \). The result is a random variable with mean and variance given by

\[
\mathbf{s} \Delta \mathbf{q} \sim (\mathbf{s} \mathbf{\mu}_C \mathbf{s} \Sigma_C)^{-1}.
\]

Substituting for \( \mathbf{\mu}_C \) and \( \Sigma_C \) using Eqs. (10) and (A1) allows us to rewrite this as

\[
\mathbf{s} \Delta \mathbf{q} \sim \left(\frac{e^{-d_{\mathbf{d}_C}/2^2}}{\sqrt{\pi(s\Sigma w + \sigma^2_{\mathbf{d}_C})^T}} 2P_C(1-P_C) \mathbf{s} \Sigma^{-1} \mathbf{s}\right).
\]

We recognize the definition of observer \( d' \) from Eq. (11) in the mean value of the cross correlation (since \( s'=w'=w'\mathbf{s} \)). We also note the square of the ideal observer detectability, \( d'_{\text{IO}} \), from Eq. (15) in the variance component. As a result, dividing the cross correlation by \( d'_{\text{IO}} \) yields

\[
\frac{\mathbf{s} \Delta \mathbf{q}}{d'_{\text{IO}}} \sim \left(\frac{e^{-d_{\mathbf{d}_C}/2^2}}{\sqrt{\pi}} d'_{\text{Obs}} 2P_C(1-P_C) \frac{\mathbf{s} \Sigma^{-1} \mathbf{s}}{N_T}\right).
\]

We now consider the square of the scaled cross correlation in Eq. (A4). The expected value of the square of a random variable is its mean squared plus its variance. Therefore the expected value of the scaled cross correlation squared is

\[
\left\langle \frac{\mathbf{s} \Delta \mathbf{q}}{d'_{\text{IO}}}^2 \right\rangle = \frac{e^{-d_{\mathbf{d}_C}/2^2}}{\pi} \frac{2P_C(1-P_C)}{N_T} \eta + \sum \eta,
\]

where the efficiency, \( \eta \), is obtained from the ratio of detectability values as in Eq. (14). Solving Eq. (A5) for efficiency yields

\[
\eta = \frac{e^{-d_{\mathbf{d}_C}/2^2}}{d'_{\text{Obs}}} \left( \left\langle \frac{s' \Delta q}{d'_{\text{IO}}}^2 \right\rangle - \frac{2P_C(1-P_C)}{N_T} \right),
\]

This formula is turned into an estimate (or prediction) of efficiency by replacing the expected cross correlation squared, \( P_C \), and \( d'_{\text{Obs}} \) by observed values. The resulting estimate is

\[
\hat{\eta} = \frac{e^{-d_{\mathbf{d}_C}/2^2}}{d'_{\text{Obs}}} \left( \left\langle \frac{s' \Delta q}{d'_{\text{IO}}}^2 \right\rangle - \frac{2\hat{P}_C(1-\hat{P}_C)}{N_T} \right).
\]

ACKNOWLEDGMENTS
The authors thank Darko Vodopich and Cedric Heath for participating in the psychophysical experiments. This work was supported by the National Institutes of Health under grants CA118294 and HL53455.
REFERENCES


54. Note that Murray et al. [37] describe cross correlation with the template of the ideal observer. This is correct only for the white-noise case they considered.