Available online at www.sciencedirect.com
SCIENCE DIRECT®
Vision Research 45 (2005) 1835–1846
Vision Research
www.elsevier.com/locate/visres

What is the signal in noise?

Dennis M. Levi *, Stanley A. Klein, Inning Chen

University of California, Berkeley, School of Optometry and The Helen Wills Neuroscience Institute, Berkeley, CA 94720-2020, USA

Received 19 August 2004; received in revised form 12 January 2005

Abstract

Visual perception is limited by both the strength of the neural signals, and by the noise in the visual nervous system; however, little is known about what aspects of the input noise the human visual system is sensitive to, i.e., what is the signal in noise? In order to investigate this question we asked observers to discriminate differences in the strength of one-dimensional white noise. We measured their response consistency and classification images and compared the results with an ideal energy detector. Our results and modelling show that discrimination of noise is limited by the observer's template (i.e., the weighted combination of energy in each stimulus component) plus higher order nonlinearities (systematic noise), and by sources of random internal noise. We found that systematic noise is present only near detection threshold. Surprisingly, we found that the human template is "adaptive"—its shape depends on the spatial frequency band of the noise—suggesting that sensitivity to spatial noise is not simply determined via passive filtering.

Keywords: Psychophysics; Classification images; Response consistency; Visual noise; Discrimination; Modelling; Double-pass

1. Introduction

It has been recognized for well over a century that visual perception is limited by both the strength of the neural signals, and by the noise in the visual nervous system. For example, Gustav Fechner, the father of psychophysics, recognized that the absolute limit for detecting a dim light resided in the "augenschwartz", the dark light of the eye (quoted in Pelli, 1981). The notion that internal noise in the visual system acts like light, even in the absence of a stimulus, i.e., dark light, led Barlow (1957) to formulate a very influential model of visual detection which posits that visual sensitivity is limited "by the difficulty of distinguishing a weak signal from the background of spurious signals, or 'noise', which occurs without any light signal at all". Barlow (1957) quantified the dark noise by determining the amount of actual light that produced the same amount of noise in the eye that is present in the dark, using the now widely used prescription for quantifying the noise, known as the equivalent input noise technique (Pelli, 1990). Engineers often specify amplifiers as a noiseless "black box" with an equivalent noise source added at its input. Analogously, the visual system can be specified by the noise added to the display that acts like the visual system's internal noise, i.e., the equivalent input noise. For detecting faint lights in the dark, the noise may have its origins in the spontaneous reactions of the photopigment rhodopsin, in the probabilistic nature of light, or in neural events in the retina or visual pathways which act like light. However, for everyday pattern vision, contrast is the coin of the realm, i.e., the visual system signals variations in contrast (i.e., changes in luminance over space relative to the local mean luminance) rather than variations in intensity, and Pelli (1990) Pelli (1981) provided a new description of the equivalent input noise as a contrast function. Contrast noise is explicitly or implicitly incorporated

* Corresponding author. Tel.: +1 510 642 3414; fax: +1 510 642 7806.
E-mail address: dlevi@berkeley.edu (D.M. Levi).

0042-6989/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2005.01.020

into all extant models of spatial vision, and the equivalent input noise has been extensively quantified by measuring contrast thresholds on a background of white noise [i.e., random fluctuations in luminance over space, time, or both] (Pelli, 1981) (Dosher & Lu, 1999; Eckstein, Ahumada, & Watson, 1997; Pelli, 1990; Pelli & Farell, 1999). In this paper, we are interested in what information observers use to judge the contrast of noise and what factors limit these judgements.

In trying to understand the factors that limit visual performance, it is common to compare the performance of human observers with that of an ideal observer, a machine that 'knows' the stimulus exactly, and that has no internal noise (Burgess, Wagner, Jennings, & Barlow, 1981; Eckstein et al., 1997; Green, 1960). This type of modeling is informative because it allows one to specify human performance relative to ideal performance, and this measure, known as efficiency, can be compared across tasks and stimuli. It is straightforward to develop an ideal observer for *discriminating* noise contrast, since a simple energy detector is ideal (Green & Swets, 1966—and see Appendix A) and the statistics of that energy detector are well known (Green & Swets, 1966). It is not possible to develop an assumption-free ideal observer model for noise *detection* because it requires knowledge of the number of photons absorbed. For that reason we focus on discrimination. The detection runs with multiple stimulus levels in addition to a blank are included to examine discrimination at very low contrasts.

One way to assess what information human observers use to judge the contrast of noise is the response classification method. Response classification provides an important new tool for learning about what information an observer uses to make perceptual decisions (Eckstein & Ahumada, 2002). By keeping track of both the pattern of noise and the observer's responses on each trial it is possible to compute the correlation between the noise and the observer's response. The result is a classification image that shows which aspects of the image influence the observer's performance. Thus, the classification image may be thought of as a behavioral receptive field (Gold, Murray, Bennett, & Sekuler, 2000). Classification images, first derived in audition (Ahumada & Lovell, 1971), provide an important new tool for measuring the "template" an observer uses to accomplish visual tasks such as detection of patterns (Ahumada & Beard, 1999), Vernier acuity (Beard & Ahumada, 2000; Levi & Klein, 2002), stereopsis (Neri, Parker, & Blakemore, 1999), and illusory contours (Gold et al., 2000).

An important study by Kersten (1987) suggests that humans are quite efficient at detecting noise over a wide range of stimulus spatial frequencies (from 1 to 6 octaves in bandwidth). Kersten's study is important because it raises questions about the now well-accepted multiple-channel model of visual detection. The multiple-channel model asserts that there are a number of narrow (1–2 octaves) bandwidth channels, each sensitive to a different range of spatial frequencies, and there is considerable evidence to support the existence of such channels for detection of simple patterns on a uniform background (see Graham, 1989 for a review). For detection of combinations of a few sinusoids the channels are combined inefficiently (Graham, 1989). However, Kersten's results seem to imply that visual noise is detected by an "adjustable" visual channel (i.e., a channel whose spatial frequency tuning is determined by the noise), just as auditory noise is detected by an adjustable auditory channel (Green, 1960). Here, for the first time, we report the classification image for visual noise. Our classification images bear some similarity to Kersten's adjustable channel.

Green (1964) coined the term "molecular psychophysics" to describe the case in which the *identical* experiment is performed twice, and the consistency of individual trials on the two passes is used to provide an estimate of the observers' internal noise (Ahumada & Lovell, 1971; Burgess & Colborne, 1988; Gold, Bennett, & Sekuler, 1999; Levi & Klein, 2003). The amount of response disagreement between the two tests allows the system's total noise to be parsed into random noise that is independent across multiple presentations of the identical stimulus, and noise that is systematic (100% correlated) across multiple presentations. Here we use a variation of the double-pass method, which we refer to as the "N-pass" method to quantify the ratio of random to systematic noise.

To anticipate, our results and modelling show that discrimination of noise is limited by three factors: a non-optimal template (i.e., the weighted combination of energy in each stimulus component) plus systematic noise in the form of higher order nonlinearities, and by sources of random internal noise.

2. Methods

2.1. Observers

Five normal observers (including two of the authors) participated in this study. Viewing was monocular, with appropriate optical correction. All experiments were performed in compliance with the relevant laws and institutional guidelines.

2.2. Stimuli

Each noise stimulus was presented for 0.75 s, with a mean luminance of 42 cd/m^2 and a dark surround. The noise is a one-dimensional grating consisting of 11 harmonics (either 0.5–5.5 c/deg, 1–11 c/deg or 2–22 c/deg)

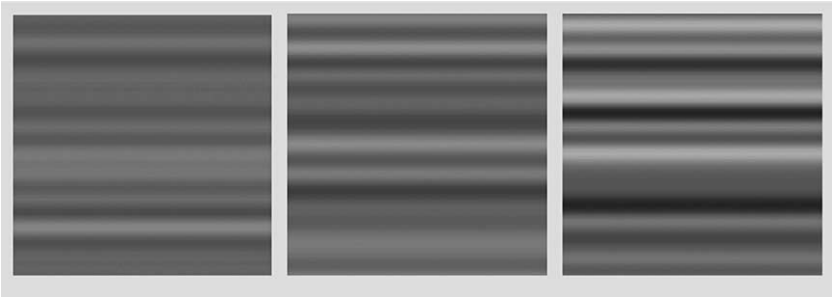

Fig. 1. Noise: one-dimensional white noise shown at low (left) medium (middle) and high (right) contrast.

with phases and amplitudes randomized (Fig. 1). The kth stimulus can be expressed as:

$$\mathrm{stim}_k(y) = N_k \sum_m (a_{k,m}\cos(2\pi m f y) + b_{k,m}\sin(2\pi m f y)) \quad (1)$$

where f is the fundamental spatial frequency and m is an integer ranging from 1 to 11. The Fourier amplitudes, $a_{k,m}$ and $b_{k,m}$ are zero mean, unit variance Gaussian random variables. Therefore, N_k is the expected rms contrast of each of the $n = 11$ components of the external noise for the kth stimulus. The expected rms contrast of the external noise is $N_k\sqrt{n}$. We varied the range of harmonics by varying the viewing distance. For the lowest range (0.5–5.5 c/deg) with $f = 0.5$ c/deg, the noise appeared in a 2.2° square field. At the higher ranges, $f = 1$ or 2 c/deg, the field size was proportionally smaller.

2.3. Psychophysical methods

We used a rating-scale signal detection methodology to measure the observers' performance. For discrimination the expected value of the contrast, N_k on the kth trial was one of 3 suprathreshold stimulus levels, and the observer responded with numbers from 1 (lowest rms contrast) to 5 (highest rms contrast). For detection either a blank ($N_k = 0$) or one of three near-threshold stimuli were shown and the responses were numbers from 1 (blank) to 4 (highest rms contrast). Data were collected in runs of 410 trials, preceded by 20 practice trials. Each threshold and classification image is based on the results averaged over either 3 or 4 separate runs (1230–1630 trials).

2.4. Classification images

We used linear regression to compute the classification coefficients, as described by Levi and Klein (2002, 2003). Details are presented in Eqs. (2)–(7) of Appendix A.

2.5. Response consistency

We used an N-pass method to determine our observers' response consistency (Burgess & Colborne, 1988; Gold et al., 1999; Levi & Klein, 2003). In this method, identical stimuli are used in each pass. Specifically, we saved the random seed from the initial run, and re-used it so that the noise was identical in either two (double-pass), 3 (triple-pass) or 4 runs (quadruple-pass). In the double-pass case, we ran two separate double passes. For triple and quadruple passes we analyzed and averaged all possible pairings. Although the same stimuli were presented in multiple runs the order of presentation changed across runs.

2.6. General response model

An outline of our general model for both detection of noise added to noise and known signals added to noise is sketched in Fig. 2a. The model assumes various sources of noise that degrade human performance. These include an imperfect template (template noise), systematic errors (higher order nonlinearities) and random noise. The computational details of the model are provided in Appendix A.

Fig. 2b (left panel) shows hypothetical predictions of noise threshold, ΔN vs. external noise, N. Both abscissa and ordinate are measure in rms contrast units. So, for example if noise of 4% rms contrast can be discriminated from noise of 2.5% rms contrast at $d' = 1$, then we would say the threshold for seeing a noise increment is 1.5% rms contrast. The gray dotted line is the predicted threshold for an ideal observer, given by Eq. (10) (in Appendix A): Ideal = N/sqrt(44). The red dashed curve is the template prediction. For Fig. 2b we assumed the correlation between the human template and the ideal template to be 0.8, so that an ideal observer using the human template would have a contrast threshold of: $T = I/0.8 = I^*1.25$.

The solid black curve is the hypothetical model prediction of the human threshold. It is simplest to describe the human data in energy units (contrast squared). We assume an energy Weber fraction of $b = 60\%$. That is, external noise of energy $1.6N^2$ can be just discriminated from noise of energy N^2. At low contrast we take the rms noise detection threshold to be th = 1%. When expressed as an equation the human rms threshold H is given by: $\Delta\mathrm{Energy} = (N + H)^2 - N^2 = \mathrm{th}^2 + bN^2$ (Fig. 2b). Or

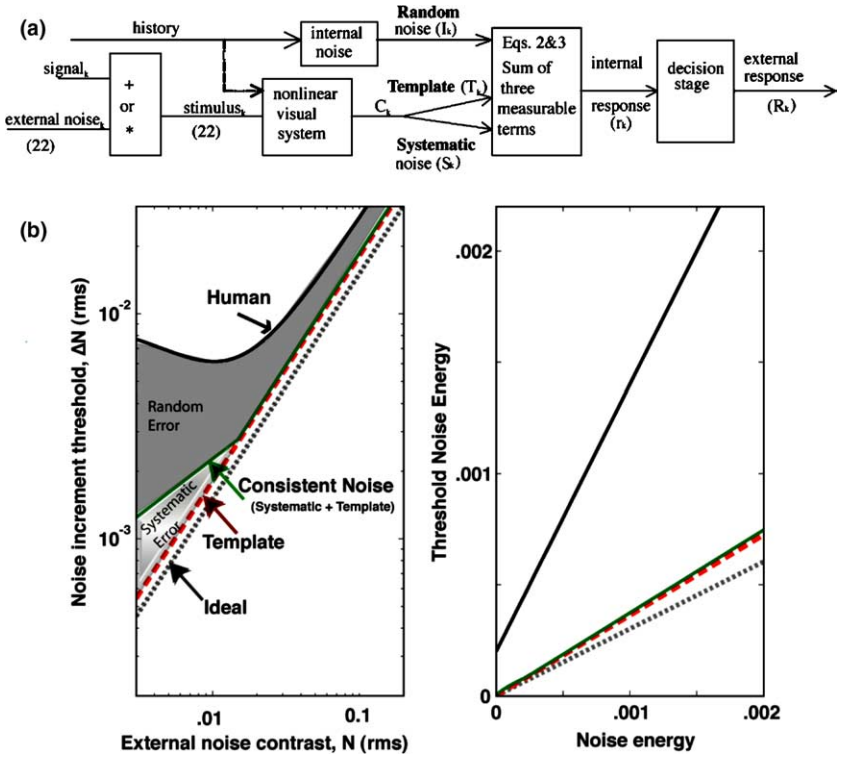

Fig. 2. (a) Noise Model: An outline of our general model for both detection of noise added to noise and known signals added to noise. See Appendix A for details. (b) (left panel) shows hypothetical predictions of noise threshold, ΔN vs. external noise, N. Both abscissa and ordinate are measured in rms contrast units. The gray dotted line is the predicted threshold for an ideal observer. The solid black curve is the model prediction of the human threshold. The dark gray region bounded by the black and green curves reflects the "random error" and the light gray region between the red and green lines shows the "systematic error". The green curve (consistent noise, C) shows the limitations imposed by the sum of the template noise and the systematic noise. The right panel of Fig. 2b simply re-plot the predictions shown in the left panel, but in units of noise energy, in order to show how the energy threshold ΔEnergy can be a linearly increasing function of noise energy, whereas the noise contrast threshold, H, can have a dipper shape.

$$H = \mathrm{sqrt}(\mathrm{th}^2 + (1+b)N^2) - N$$
$$= \mathrm{sqrt}(.01^2 + 1.6N^2) - N$$

The dark gray region bounded by the black and green curves reflects the "random error" and the light gray region between the red and green lines shows the "systematic error". The green curve shows the limitations imposed by the sum of the template noise and the systematic noise, S. The Green line is what is called consistent noise, C, in Appendix A. For external noise contrasts above 1.5% it is taken to be just 3% larger than the Template curve (the 3% is chosen so that the two curves are distinguishable). Below 1.5% noise contrast the systematic noise is taken to fall off like the square root of external noise rather than linearly, in order to match the real data. The ratio of the consistent threshold to the human threshold is given by the 2-pass correlation as specified by Eq. (15) and as shown in Fig. 2b. The specific numbers specifying the curves in Fig. 2b were chosen so that the curves resemble the measured curves in Fig. 3a. The right panel of Fig. 2b simply replots the predictions shown in the left panel, but in units of noise energy, in order to show how the energy threshold ΔEnergy can be a linearly increasing function of noise energy, whereas the noise contrast threshold, H, can have a dipper shape.

3. Results

Fig. 3 shows both human (large black circles) and ideal discrimination thresholds (dotted gray line) as a function of noise contrast. The top panel shows that human noise thresholds (ΔN) follow the well-known "dipper" function (Legge, 1981; Stromeyer & Klein, 1974) of noise contrast, first falling as contrast increases, and then rising more or less in proportion to the noise contrast, indicating that noise discrimination is a more or less constant Weber fraction of the noise contrast once noise contrast reaches about three times the noise detection threshold. Human efficiency (defined as the ratio of ideal to human thresholds squared—Fig. 3 bottom) is lowest at low (near threshold) noise levels (≈3–4%), and increases to about 30% beyond the dipper regime. Eqs. (2) and (3) in Appendix A identified three factors producing efficiency loss: (1) a poorly matched

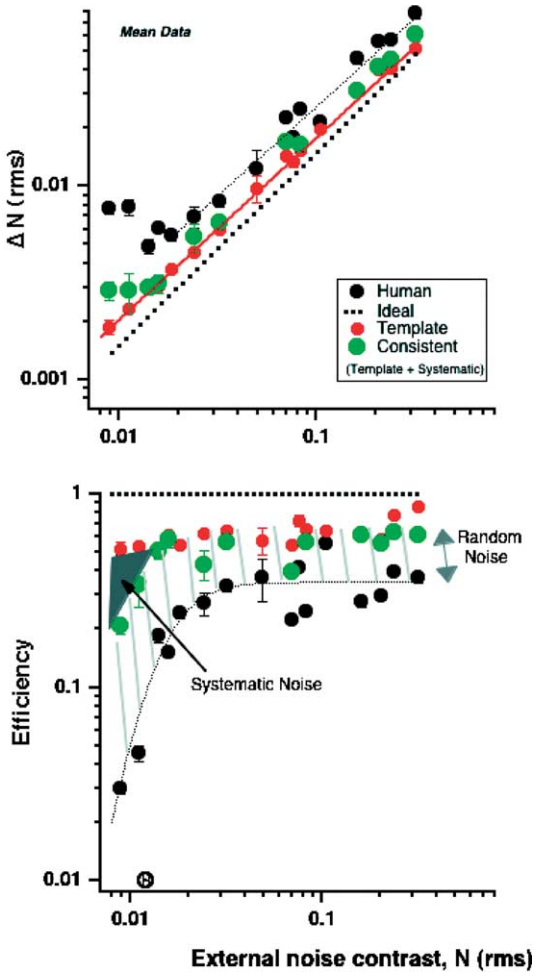

Fig. 3. Noise discrimination thresholds as a function of external noise contrast (noise spatial frequency range of 0.5–5.5 c/deg) for human observers (black circles). Predictions are also shown for the ideal observer (gray dotted line), the "template" observer (red symbols) and for the consistent noise (green circles). The top panel shows noise thresholds (ΔN) and the bottom panel, Efficiency. The open black symbol near the abscissa in the bottom panel shows the observers' threshold for detecting the noise. The stripes and the solid shading represent random noise and systematic noise respectively.

template, (2) high levels of internal noise and (3) higher order nonlinearies (systematic noise) not present in the transducer function. Below we will try to estimate the contribution of each of these factors.

One factor that might limit human performance is the information that the observer uses to solve the task. An ideal observer would use all of the information in the stimulus, equally weighted. Human observers might be expected to be more sensitive to certain spatial frequencies than to others. Perceptual task performance is often modeled as a "template", reflecting the weighted combination of inputs from basic visual mechanisms, with sources of internal noise (Dosher & Lu, 1999).

We assessed the observer's template by measuring their classification images for white noise. The classification coefficients for white noise (Fig. 4), obtained using linear regression, show how each of the spatial frequency components of the external noise affected the observer's responses, over a large range of input noise levels from below threshold, to well above threshold. The classification coefficients reveal that for each of the three frequency ranges (shown by the different colors) the classification image is band-pass—it is more or less proportional to f for low frequencies, and there's a rapid drop in the coefficient amplitude at high frequencies. The classification plots are similar in shape, over a broad range of noise levels, showing the same reduction at both low and high spatial frequencies. The classification template provided a major surprise: the template changes shape by recentering on the range of frequencies in the stimulus. The high frequency falloff shifts to 4.5, 7 and 10 c/deg for stimuli whose cutoff is 5.5, 11 and 22 c/deg. A naive hypothesis would posit that an observer's sensitivity to white noise could be predicted by convolving the stimulus with the observer's detection based contrast sensitivity function (Jamar & Koenderink, 1985; Mostafavi & Sakrison, 1976; Quick, Mullins, & Lucas, 1978). The strong dependence of the band-pass classification images on the stimulus noise range (particularly the shift of the high spatial frequency fall-off shown in Fig. 4) shows that this hypothesis cannot be correct.

We can predict how much the observer's template contributes to the loss of efficiency (square of contrast loss) by computing the performance of the human observer's template (red circles in Fig. 3). The human template is moderately efficient (template efficiency varies from about 50% at low noise levels to close to 80% at high—Fig. 3, bottom panel). A template efficiency of 80% means the human template has an inner product of sqrt(0.8) = 0.9 with the ideal, flat, template. Thus most of the loss is due to other factors, and we investigate those further using multiple runs with the identical stimuli intermixed in each run as discussed in Methods.

By repeating the experiment several times with *identical* noise sequences shown in a randomized order, we measured q, the ratio of consistent to total noise contrasts. At low noise contrasts (near the noise detection threshold), q increases rapidly from zero to about 0.5, and then continues to increase more slowly up to about 0.6 (Fig. 5). Thus, not surprisingly, at low noise contrast random noise dominates, while at high noise contrast, consistent noise dominates.

We calculated q in two different ways: the open symbols in Fig. 5 show q_{Npass} (as described above); the solid symbols show q_{bi} a bivariate ellipse fit to the data (see methods). While the two methods give similar fits, they are not identical (see inset in Fig. 5); our simulations suggest that q_{Npass} overestimates the amount of systematic noise at low noise levels and underestimates it at high levels. We suspect the true value lies between the two estimates.

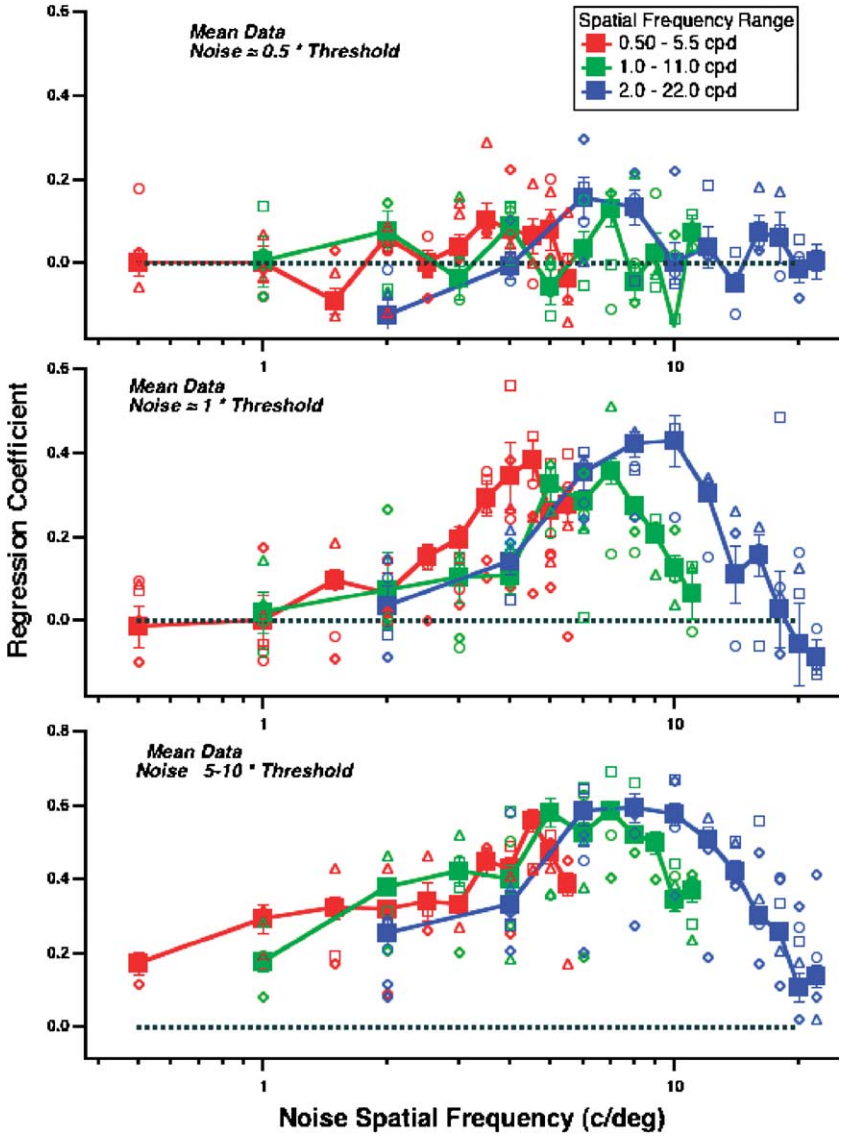

Fig. 4. Noise Templates: Classification Images for noise levels of about 0.5 (top), 1 (middle) and 5–10 times threshold (bottom). The 3 colors represent 3 different noise ranges (red 0.5–5.5, green 1–11 and blue 2–22 c/deg). The filled symbols and lines are averaged across the 4 observers. The small open symbols are for individual observers.

The green symbols in Fig. 3 show the contribution of consistent noise (based on the bivariate fits) to human performance. At low noise contrast, consistent noise reduces efficiency to around 20%, whereas at high noise contrast, consistent noise results in little or no loss of efficiency beyond the mismatched template. The gap between the human performance and the consistent noise prediction, shows the effect of random noise (stripes in the lower panel of Fig. 3). At all noise levels above detection threshold, human discrimination thresholds are about 50% higher than the consistent noise prediction, so random noise reduces human efficiency by about a factor of 2.25 (1.5^2) over the approximately 40-fold range of noise levels tested. We note that this random noise is stimulus dependent or multiplicative, consistent with the Weber's law dependence of noise thresholds on the noise pedestal.

An important feature of Fig. 3 is the gap at the two lowest external noise levels, between the consistent noise and the template noise (solid region). These were the discrimination data collected in the detection runs. The gap represents the presence of a systematic error due to higher order nonlinearities, the C term in Eqs. (2) and (12) in Appendix A. These nonlinearities may be explained by the multichannel model that has been successful in accounting for detection data on a blank field (Graham, 1989). At very low external noise levels the test increments are likely to be detected by medium bandwidth filters whose information is pooled inefficiently, by a pooling exponent greater than 3 as was found by Gra-

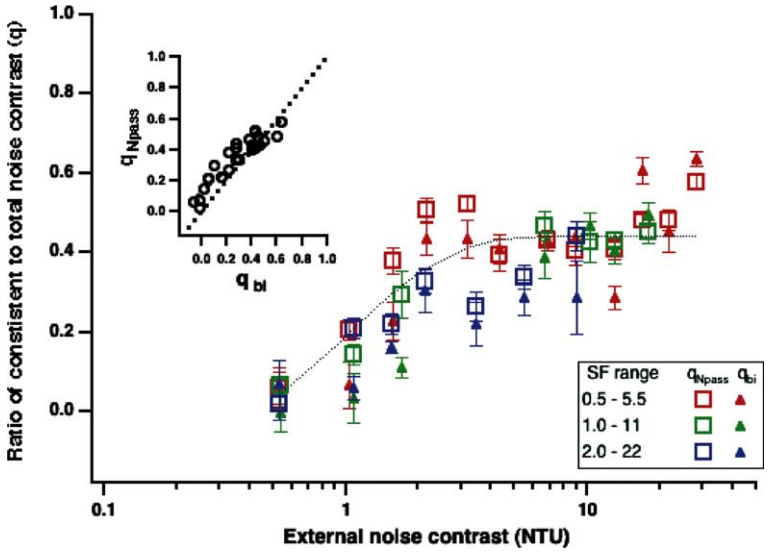

Fig. 5. Ratio of consistent to total noise contrast (q) as a function of rms noise contrast (normalized to the observers' noise detection thresholds, i.e., in Noise Threshold units, NTU). We used two different methods to estimate q (see Methods). Open symbols show q estimated from the correlation coefficient (q_{Npass}). Filled symbols show q estimated from a bivariate fit (q_{bi}). The inset compares the two estimates of q. The 3 colors represent 3 different noise ranges (red 0.5–5.5, green 1–11 and blue 2–22 c/deg). Since each spatial frequency range has a different detection threshold, we plot the data in NTU.

ham and Nachmias (1971) for detection of compound gratings. An example of the output of one of these filters is shown in Eq. (6) of Appendix A.

4. Discussion

4.1. The template for noise

Our classification images reveal the signal in noise, showing which aspects of noise influence an observer's responses. Our results and modeling show that human efficiency is modest, being between 25% and 35%, and help to untangle the various factors that limit human detection and discrimination of noise.

One surprising result of our study is the band-pass shape of the noise classification image *in each noise band* even with noise that is barely detectable ($d' \approx 1$). A number of previous studies examined human sensitivity to noise patterns, and it has sometimes been assumed that detection of the noise occurs after filtering by the eye's contrast sensitivity function (Jamar & Koenderink, 1985; Kersten, 1987). Our results suggest that sensitivity to spatial noise is not simply determined via passive filtering (i.e., it is not simply the input noise convolved with the observer's contrast sensitivity function). Rather, the shape, and particularly the shifting of the high spatial frequency fall-off at each frequency range, suggests that there must be active neural interactions.

One idea is that the shape of the noise template may be the result of high frequencies masking low. Another idea is that the observer is using one octave channels. The higher frequency channels pool over more stimulus components and thus have relatively lower noise. An ideal detector looking at the output of these channels would give the higher frequencies greater weighting. We tested both of these notions by measuring the classification image for noise that was spaced logarithmically in frequency (rather than linearly as before). This has the effect of boosting the low spatial frequency content. Interestingly we still find a strong low-frequency fall-off in the classification image (Fig. 6a) suggesting that the shape of the classification image is not simply a consequence of the high spatial frequency components of the noise inhibiting the low. Neither is it a consequence of greater pooling of information at higher frequencies.

It is not difficult to develop a plausible model that produces classification images similar to those shown in Fig. 4. Since the task is noise contrast discrimination the observer must develop a method for assessing the contrast of the image. The ideal observer would calculate the image mean luminance and then calculate the rms deviation from that mean. It might be difficult for the human to make this calculation because it requires attention to the full image, with a single global mean luminance calculation. Much simpler for the human would be to judge the contrast of nearby stimulus regions. We carried out simulations in which the simulated observer divided the image into 10 regions, blurred those regions so each had a uniform luminance, took the difference in luminance of adjacent blurred regions, and then based its response on the

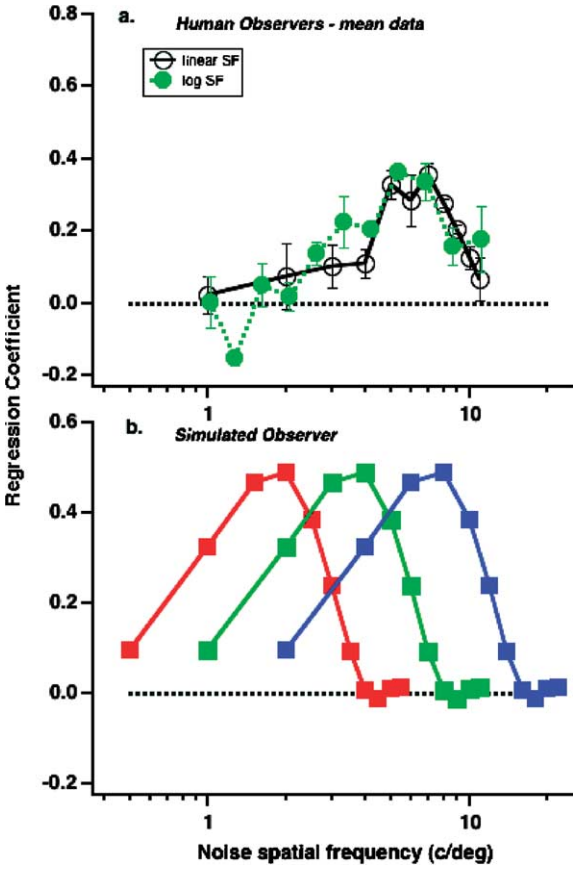

Fig. 6. (a) Classification images for noise (1–11 c/deg) at detection threshold ($d' = 1$). Black symbols –linear spacing; green symbols—logarithmic spacing. (b) Simulated classification images in which the simulated observer divided the image into 10 regions, blurred those regions so each had a uniform luminance, took the difference in luminance of adjacent blurred regions, and then based its response on the rms value of those differences. These classification images are very similar to the Fourier domain classification images of our human observers shown in Fig. 4.

rms value of those differences. The classification images produced by this simulated observer are shown in Fig 6b, and are very similar to the Fourier domain classification images of our human observers shown in Fig. 4. The notion that the classification image is relatively independent of viewing distance is reasonable since a size constancy mechanism whereby the observer does the blurring and differencing based on screen coordinates rather than angular coordinates is not unreasonable.

The notion that the template for visual noise may be an adaptive visual channel is consistent with Green's (1960) inference that the critical bands for auditory noise are adjustable, and would also explain the recent report that human observers can summate both spatial frequency (Taylor, Bennett, & Sekuler, 2003). and orientation information in noise over a very broad range of spatial frequency and orientation bandwidths (Taylor, Bennett, & Sekuler, 2004).

4.2. Internal noise and response consistency

It is often claimed that double-pass methods (Burgess & Colborne, 1988; Gold et al., 1999; Green, 1964) measure the ratio of internal to external noise, whereas we prefer to say they measure the ratio of random to consistent noise. The distinctions among different types of noise can be subtle so we will try to clarify them here.

We first discuss the relationship between internal and random noise, then we discuss the relationship between external and consistent noise.

The relationship between internal noise and random noise depends on how the experiment is done. In our double-pass experiment we were interested in how the subject's response to stimulus k is determined by properties of stimulus k. We were not interested in how response k depended on specific prior stimuli. For that reason we randomized the order of stimuli from run to run, keeping track of the order for the double-pass comparison. Thus any systematic dependence of the kth stimulus on the preceding few stimuli would be erased.

One could have done the experiment without randomizing the order. In fact, in the seminal paper on the double-pass method, Green (1964) measured noise detection on a noise background and did not randomize the order. In that type of experiment the response on a given trial could depend on the stimuli in preceding trials. For example, if a low contrast stimuli were shown on the previous three trials (known to the subject by the feedback) there could be a subjective (incorrect) belief that the next trial would be of higher contrast. This possibility would result in a contribution to the consistency of responses to the kth trial, so Eq. (3) of Appendix A would be rewritten as:

$$rc_k = T_k + S_k + P_k$$

The term P_k would be the subjective factor that depended on the stimuli in the few trials preceding the kth trial. This prior stimulus influence on consistency is indicated by the dashed line in Fig. 2 going from the stimulus history to the consistency section of the model. This dashed line would be present only if the presentation order is the same in the multiple passes. Similarly, a medium contrast stimulus preceded by several high contrast stimuli may be judged to be of lower contrast due to adaptation.

The experimental method determines whether the term P_k is grouped with random noise (our method of scrambling the presentation order in the two passes) or with systematic noise (Green's method of keeping the same order). We prefer our method since we were interested in restricting the systematic aspects of the response to a dependence solely on nonlinear properties of the kth stimulus. This discussion shows that response inconsistency can come from factors other than the internal noise generated by the present stimulus. We prefer using

term random noise rather than internal noise since the former is an operational term defined by the results of a double-pass experiment.

The usual statement that the double-pass method measures the ratio of internal to external noise not only has a problem with the definition of internal noise, it also has a problem of defining external noise. We believe that the name "external noise" should be restricted to properties of the external noise that is independent of the observer. Examples of a single number that characterize the external noise are the trial-by-trial total noise energy or it could be N_k, the average noise spectral density as given in Eq. (1). There are definitions that depends solely on the external noise and do not depend on the properties of the observer. Other authors, however, define external noise to be what we call consistent noise $rc(k_s)$ (Burgess & Colborne, 1988; Gold et al., 1999). We prefer the term consistent noise and reserve the name external noise for a measure of the external noise that is independent of the human observer.

4.3. Efficiency

Our estimate of efficiency is consistent with previous work (Kersten, 1987); however, combining threshold measurements with trial-by-trial analysis (Green, 1964) allows us a unique way of dissecting the sources of noise in the visual nervous system that limit efficiency. Our results and modelling show that the human template for noise is quite efficient, that random noise reduces efficiency by about a factor of 2.25, and that systematic noise only reduces efficiency at low input noise levels.

Acknowledgments

We are grateful to Hope Queener for developing the processing tools and to Peter Neri and Roger Li, and Craig Abbey for helpful comments on the paper. This research was supported by Research grants R01EY01728 and RO1 EY04776 from the National Eye Institute, NIH.

Appendix A. General response model and classification images

Our general model for both detection of noise added to noise and known signals added to noise is sketched in Fig. 2. The first step is to combine the known external noise with the stimulus. In previous classification image experiments the noise and signal are combined additively. In our present experiments, on the other hand, the stimulus level, N_k, multiplies the external noise. The formalism that we use for analyzing these types of experiments works equally well for additive and multiplicative signals.

Our approach for modeling the detection and discrimination of signals in noise is to subdivide the subject's internal response, r_k, into two parts, consistent, rc_k, and random, ri_k.

$$r_k = rc_k + ri_k \tag{2}$$

We refer to the component, rc_k, as the consistent part of the internal response because if the identical stimulus is shown multiple times, the same consistent output would be obtained. The random part, ri_k, (the upper branch in Fig. 2) provides a varying contribution to the internal response, r_k, when the same stimulus is presented multiple times. An N-pass methodology, based on response inconsistency, will be used to identify ri_k. Although random noise is often called internal noise, we prefer to use the operational term random noise to remove methodological ambiguities, as was discussed in Section 4.

In principle, the consistent portion of the response can be a complicated nonlinear function of the stimulus. In practice however, it is found that often (see Section 3) a linear or quadratic approximation to the consistent response provides an excellent fit to the data. For the present contrast discrimination task, the consistent response on the kth trial will be split into a template term representing a weighted contrast energy plus a systematic error containing the higher order nonlinearities, S_k

$$rc_k = T_k + S_k \tag{3}$$

Fig. 2b is a graphical representation of Eqs. (2) and (3). The template term is given by

$$T_k = \sum_m e_{k,m} t_m \tag{4}$$

where the contrast energy at frequency, m, can be obtained from Eq. (1) as

$$e_{k,m} = N_k^2 (a_{k,m}^2 + b_{k,m}^2) \tag{5}$$

The template, t_m, specifies the observer's weighting of the 11 Fourier energy components of the stimulus. An ideal observer would use a constant template, $t_m = t$ (Green & Swets, 1966). For a task requiring detection of a stimulus known exactly, Eq. (5) would contain linear rather than quadratic terms in $a_{k,m}$ and $b_{k,m}$. Linear terms are not expected to contribute to the present noise discrimination task since these terms merely shift the location of the peak activity of the kth energy component without affecting its strength.

The systematic error term S_k, in Fig. 2 and Eq. (3), includes the effects of all additional nonlinear contributions to the response. These higher order terms can be written as polynomials of the stimulus Fourier components, $N_k a_{k,m}$ and N_k bk,m, that would contribute in addition to the quadratic combinations of the Template term in Eq. (5). An example of a contribution to S_k would be

a multichannel model where the energies, $e_{k,m}$ in different frequency bands are summated nonlinearly. For example, a medium bandwidth symmetric mechanism at the center of the image could contribute a term such as:

$$S_k = (N_k(a_{k,5} + 2a_{k,6} + a_{k,7}))^n \quad (6)$$

where n is an exponent characteristic of the mechanism. For quadratic summation, $n = 2$, S_k would include the cross terms $a_{k,5}a_{k,6}$ and $a_{k,6}a_{k,7}$ that are not present in Eqs. (3) and (4). The higher order nonlinearities, S_k in Eq. (3), can in principle, depend not only on the 22 numbers characterizing the kth stimulus, but also on the strength of preceding stimuli, (see Section 4) as for example by an adaptation mechanism had we not randomized the presentation order.

In our experiments, the observer's rating scale response, R_k, is an integer determined by the magnitude of r_k relative to the observer's fixed criteria. In the noise contrast discrimination experiments the observer uses four criteria to define five response categories (by giving integer numbers from 1 to 5). For example, with mean criteria of 1.1, 3.2, 4.8, 6.3 and an internal response of $r_k = 2.8$, the observer's response would be $R_k = 2$ (the second category, between 1.1 and 3.2). The decision stage that converts the internal response, r_k, to the external response R_k is represented by the rightmost box in Fig. 2.

The template classification image, t_m, in Eq. (4) can be obtained by linear regression, using a model specified by Eqs. (2)–(4)

$$R_{k(s)} = \sum_m e_{k(s),m} t_{m,s} + f_s. \quad (7)$$

An index, s, has been added to k to restrict the stimuli to stimulus level s when doing the regression, where s goes from 1 to 3 for both detection and discrimination (the lowest detection level has no noise and thus no template estimate). f_s, an arbitrary additive constant that depends on the stimulus level, s is included as a parameter in the regression. The index, s, is needed to do the linear regression on each stimulus level separately to minimize the distortion produced by unequally spaced criteria. There are occasions to consider the levels separately, as for example when the template specific to a particular level is desired. The templates were found to be relatively independent of stimulus level. Therefore in Section 3, we report the regression coefficients, t_m, averaged over the three stimulus levels, s.

A.1. Signal detection theory (SDT) and thresholds

Consider the task of discriminating stimulus level s from stimulus level s' (where $s < s'$). SDT allows one to specify the discriminability of two noisy signals in terms of a d' measure. This is possible even for the present case of discriminating noise energy where the relevant statistic has a chi-square rather than a Gaussian distribution (Green & Swets, 1966). We can define d' for the task to be

$$d'_{ss'} = (M_{s'} - M_s)/\mathrm{sqrt}(Vt_s) \quad (8)$$

where M_s and Vt_s are the mean and total variance of the internal response, r_k, for stimulus level s. We use the notation Vt for the total variance because later we will need to consider the variance of subcomponents of Eq. (2). By taking the denominator of Eq. (8) to be the variance for reference stimulus level s, we are choosing to define d' to be at the horizontal z-score intercept of the ROC curve (Green & Swets, 1966). This can be done even for non-Gaussian distributions. Green (1964), in what may be the first paper on the topic, developed the formalism for analyzing a task very similar to ours, but using a two-alternative forced choice paradigm, in which the denominator is the square root of the average of the variances of the two alternatives.

For the ideal observer only the first term of Eq. (2) (consistent part) and the first term of Eq. (3) (template part) contribute to the response. The template, $t_{m,s}$, for the ideal observer is a constant, independent of frequency, m, or stimulus level, s, because the noise is white, with each component contributing equally to the discrimination. The numerator and denominator of Eq. (8) are related to the mean and variance respectively of a chi square distribution since for the ideal observer the template response is the sum of squares of 22 random numbers. For the task of discriminating noise specified by contrasts N_s and $N_{s'}$ the ideal observer's d', based on the variance of a chi square distribution is (Green & Swets, 1966, Eq. (6.67))

$$d'_{\mathrm{ideal}} = \mathrm{sqrt}(4n)\Delta N/N, \text{ for small values of } \Delta N/N. \quad (9)$$

where $n = 11$, and $\Delta N = N_{s'} - N_s$, $N = (N_{s'} + N_s)/2$. If we had quantified the noise strength by its energy rather than contrast, Eq. (9) would have been (Green, 1964; Green & Swets, 1966):

$$d_{\mathrm{ideal}} = \mathrm{sqrt}(2n)\Delta E/E$$

Since we define threshold at $d_{\mathrm{ideal}} = 1$, Eq. (9) implies the ideal threshold noise increment is

$$\Delta N_{\mathrm{thresh}} = N/\mathrm{sqrt}(4n) = 0.15N. \quad (10)$$

for $n = 11$. Since N is on both sides of Eq. (10), Eq. (10) holds both for N defined as the component contrast as in Eq. (1), or for it defined as the rms total noise contrast. This ideal observer prediction is shown as the dotted gray line in Fig. 3.

Ideal Observer assumptions can be loosened slightly by going to the "Template Observer" that is identical to the Ideal Observer, except that the template, t_m, does not need to be a constant. The value of t_m can be

obtained from the linear regression based on Eq. (7). Following the logic of Levi and Klein (2002), Eqs. (5)–(7) we are able to derive d_T, the Template Observer's d' in terms of d_{ideal}, the Ideal Observer's d' and the template efficiency, TE

$$d_T^2 = \text{TE} d_{\text{ideal}}^2 \tag{11}$$

where TE is the square of the normalized overlap between the Template and Ideal weightings (Levi & Klein, 2002).

For a general ideal observer template given by w_m, the template efficiency is (Levi & Klein, 2002)

$$\text{TE} = \left(\sum w_m t_m\right)^2 \Big/ \left(\left(\sum w_m^2\right)\left(\sum t_m^2\right)\right) \tag{12}$$

where the summations go over $n = 11$ components for our experiments. This is the square of the cosine of the angle between two vectors. For the present case where the ideal template is a constant, the template efficiency is

$$\text{TE} = \left(\sum t_m\right)^2 \Big/ \left(n \sum t_m^2\right) \tag{13}$$

In Section 3 we show that except for the highest frequencies the human template is roughly proportional to frequency ($t_m = m$). Eq. (13) gives TE = 0.78 for $t_m = m$, which asymptotically becomes TE = 3/4 as the number of components become large. This value is close to what we report in Fig. 3.

A.2. Response consistency

The template (first) term of Eq. (3) was experimentally isolated by the classification image method. We now discuss how we experimentally isolated the full model (first term of Eq. (2)) shown in Fig. 2 using a response consistency N-pass method. In this method, identical stimuli are used in each pass. Specifically, we saved the random seed from the initial run, and re-used it so that the noise was identical in either two (double-pass), 3 (triple pass) or 4 runs (quadruple pass). In the double-pass case, we ran two separate double passes. For triple and quadruple passes we analyzed and averaged all possible pairings.

For the yes–no or 2AFC methodologies a double-pass experiment supplies one item of information not available in a single pass. It supplies the fraction of trials for which the same answer was given. For our methodology using multiple stimulus levels and multiple ratings the double-pass method offers more than one new item of information since one learns the fraction of times the answers differed by a particular integer (e.g. 0, 1 or 2), for each stimulus level.

The first step in analyzing double-pass data is to modify Eq. (2) to include p, an integer that specifies the pass

$$r_k(p) = rc_k + ri_k(p) \tag{14}$$

where rc_k was given by Eqs. (3) and (4) and components of ri_i was considered in Section 4. Eq. (14) shows the random response changes from pass $p = 1$ to $p = 2$, whereas the systematic response is independent of pass. In discussing the various quantities in Eqs. (14) and (15) we will use the names "noise" and "response" interchangeably. Although rc_k, the consistent response, is not true noise since it is knowable, it is commonly called noise since it is a function of the external noise.

The ratio of consistent (Vc) to total (Vt) response variance is specified by the cross-correlation between the total responses of the two passes, Correl($r_k(1)$, $r_k(2)$)

$$q^2 = \text{Correl}(r_k(1), r_k(2)) = Vc/Vt \tag{15}$$

The derivation of Eq. (15) is based on the assumptions that the consistent and random responses are uncorrelated and also that the random responses of the two passes are uncorrelated. The variance of the random response noise, Vi, is given by the variance of Eq. (14):

$$Vi = Vt - Vc \tag{16}$$

Just because ri is uncorrelated with rc does not mean that the two are independent. In fact they are highly dependent. The double-pass experiments show that an increase in the external noise results in an increase in the random noise. Previous studies have shown that an increase in external noise produces a disproportionate increase in both random and consistent noise (Burgess & Colborne, 1988), but those studies that lacked the classification template were unable to distinguish the components of consistent noise (systematic noise vs. a mismatched template).

By the logic discussed following Eq. (8) we are finally able to relate d_c, the d' of the systematic observer with no random noise, to d_h, the d' of the human observer:

$$(d_c/d_h) = q \tag{17}$$

Thus the correlation, q^2, represents the loss of efficiency in going from a consistent observer to the human observer.

The correlation, q^2, in Eq. (15) cannot be directly calculated because the internal response, r, is not available. It must be inferred from the rating responses R. The correlation was calculated at each noise level by two independent methods: (1) Finding the best fit to the double-pass rating data using a bivariate Gaussian constrained by d' values and ROC criteria. The two dimensions of the bivariate Gaussian were the paired responses in the two matched runs. (2) Directly calculating the standard correlation coefficient based on the paired human responses, R. Simulations show that the error in using R rather than r is usually small. This is partly because by restricting the analysis to a single stimulus level the range of ratings is limited. The two methods gave similar but not identical results (see Fig. 4).

References

Ahumada, A. J., & Beard, B. L. (1999). Classification images for detection. *Investigative Ophthalmology & Visual Science, 40*(Suppl), s3015.

Ahumada, A. J., & Lovell, J. (1971). Stimulus features in signal detection. *Journal of the Acoustical Society of America, 49*, 1751–1756.

Barlow, H. B. (1957). Increment thresholds at low intensities considered as signal/noise discriminations. *Journal of Physiology, 136*, 469–488.

Beard, B. L., & Ahumada, A. J. (2000). Response classification images for parafoveal Vernier acuity. *Investigative Ophthalmology & Visual Science, 41*(Suppl), s804.

Burgess, A. E., & Colborne, B. (1988). Visual signal detection. IV. Observer inconsistency. *Journal of the Optical Society of America A, 5*, 617–627.

Burgess, A. E., Wagner, R. F., Jennings, R. J., & Barlow, H. B. (1981). Efficiency of human visual signal discrimination. *Science, 214*, 93–94.

Dosher, B. A., & Lu, Z. L. (1999). Mechanisms of perceptual learning. *Vision Research, 39*, 3197–3221.

Eckstein, M. P., & Ahumada, A. J. (2002). Classification images: A tool to analyze visual strategies. *Journal of Vision, 2*, 1–2.

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*, 2406–2419.

Gold, J. M., Bennett, P. J., & Sekuler, A. B. (1999). Signal but not noise changes with perceptual learning. *Nature, 402*, 176–178.

Gold, J. M., Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2000). Deriving behavioural receptive fields for visually completed contours. *Current Biology, 10*, 663–666.

Graham, N. (1989). *Visual pattern analyzers*. New York: Oxford University 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.

Green, D. M. (1960). Auditory detection of a signal in noise. *Journal of the Acoustical Society of America, 32*, 121–131.

Green, D. M. (1964). Consistency of auditory detection judgements. *Psychological Review, 71*, 392–407.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Jamar, J. H., & Koenderink, J. J. (1985). Contrast detection and detection of contrast modulation for noise gratings. *Vision Research, 25*, 511–521.

Kersten, D. (1987). Statistical efficiency for the detection of noise. *Vision Research, 27*, 1029–1040.

Legge, G. E. (1981). A power law for contrast discrimination. *Vision Research, 21*, 457–467.

Levi, D. M., & Klein, S. A. (2002). Classification images for detection and position discrimination in the fovea and parafovea. *Journal of Vision, 2*, 46–65.

Levi, D. M., & Klein, S. A. (2003). Noise provides some new signals about the spatial vision of amblyopes. *Journal of Neuroscience, 23*, 2522–2526.

Mostafavi, H., & Sakrison, D. J. (1976). Structure and properties of a single channel in the human visual system. *Vision Research, 16*, 957–968.

Neri, P., Parker, A. J., & Blakemore, C. (1999). Probing the human stereoscopic system with reverse correlation. *Nature, 401*, 695–698.

Pelli, D. G. (1981). Effects of visual noise. Doctoral dissertation. Cambridge University, Cambridge, UK.

Pelli, D. G. (1990). In C. Blakemore (Ed.), *Visual coding and efficiency*. Cambridge: Cambridge University Press.

Pelli, D. G., & Farell, B. (1999). Why use noise? *Journal of the Optical Society of America A, 16*, 647–653.

Quick, R. F., Mullins, W. W., & Lucas, R. N. (1978). Contrast thresholds of random patterns. *Photographic Science and Engineering, 22*, 72–75.

Stromeyer, C., & Klein, S. A. (1974). Spatial frequency channels in human vision as asymmetric (edge) mechanisms. *Vision Research, 14*, 1409–1420.

Taylor, C. P., Bennett, P. J., & Sekuler, A. B. (2003). Noise detection: Bandwidth uncertainty and adjustable channels. *Visual Sciences Society abstract*.

Taylor, C. P., Bennett, P. J., & Sekuler, A. B. (2004). Noise detection: Optimal summation of orientation information.*Visual Sciences Society abstract*.