Signal detection in power-law noise: effect of spectrum exponents

Arthur E. Burgess* and Philip F. Judy

Brigham and Women's Hospital, Harvard Medical School, 75 Francis St., Boston, Massachusetts 02115, USA
**Corresponding author: ab2boston@gmail.com*

Received March 5, 2007; revised July 4, 2007; accepted July 10, 2007;
posted August 9, 2007 (Doc. ID 80574); published September 26, 2007

Many natural backgrounds have approximately isotropic power spectra of the power-law form, $P(f)=K/f^{\beta}$, where f is radial frequency. For natural scenes and mammograms, the values of the exponent, β, range from 1.5 to 3.5. The ideal observer model predicts that for signals with certain properties and backgrounds that can be treated as random noise, a plot of log (contrast threshold) versus log (signal size) will be linear with slope, m, given by: $m=(\beta-2)/2$. This plot is referred to as a contrast-detail (CD) diagram. It is interesting that this predicts a detection threshold that is independent of signal size for β equal to 2. We present two-alternative forced-choice (2AFC) detection results for human and channelized model observers of a simple signal in filtered noise with exponents from 1.5 to 3.5. The CD diagram results are in good agreement with the prediction of this equation. © 2007 Optical Society of America

OCIS codes: 330.0330, 330.1880, 330.4060, 330.5510.

1. INTRODUCTION

Experiments have demonstrated that the human performance of a number of noise-limited visual tasks can be successfully evaluated by comparison with the ideal Bayesian observer based on signal detection theory [1,2]. The comparison of human and ideal observer performance is sometimes done using observer efficiency. Tanner and Birdsall [3] defined efficiency using the signal energy, E, needed by each observer to perform a specified task at a specified level of accuracy (for example, 90% correct detection of an exactly known signal in noise added to a known background). Signal energy is proportional to the square of signal amplitude or contrast, so these can be used as an alternative definition. The Tanner and Birdsall absolute efficiency measure at a selected performance level is defined as

$$\text{efficiency} = \frac{\text{Signal energy required by the ideal observer}}{\text{Signal energy required by the human observer}} = \left[\frac{\text{Signal amplitude required by the ideal observer}}{\text{Signal amplitude required by the human observer}}\right]^2. \quad (1)$$

It should be noted that one of the authors used a different equation in a very early paper [4] on human observer efficiency for signal amplitude discrimination. That equation was based on detectability indices, d', and is evaluated for a selected signal energy (or amplitude). Using subscripts H and I for human and ideal observers, respectively, the efficiency equation at a selected signal energy is

$$\text{efficiency} = \left[\frac{d'_H}{d'_I}\right]^2. \quad (2)$$

The two definitions give identical values only if the observer d' values are proportional to signal amplitude. Nachmias and Sansbury [5] showed that this is the case for signal amplitude discrimination but is rarely true otherwise. Burgess *et al.* [4] had found this to also be true for signal discrimination in white noise, so their use of Eq. (2) was justifiable. However, we suggest that Eq. (1) is the preferable definition because it is measured at a selected performance level. In general, the relationship between human d' values and signal amplitude is not proportional because there is a nonlinearity at low amplitude. The relationship does become linear for d' values above approximately 0.5. Burgess and Ghandeharian [6] found that, for aperiodic signals, the linear regression fit to d' values above 0.5 intercepts the d' axis at negative values. The intercept value depends on the signal type and the offset is greater for periodic signals such as sine waves. When the offset is large the use of Eq. (2) can give very low human efficiencies. This arises if the selected signal energy gives a model observer d' of ~2 or 3 and the human d' determined at the same signal energy is in the region where the relationship between human d' values and signal amplitude is nonlinear. We regard the efficiency evaluation at a selected performance level (d' equal to 2, for example) as a more realistic procedure because one is more likely to be interested in human performance at d' values above approximately unity.

It has been found that human efficiencies are usually in the range from 0.25 to 0.5 for aperiodic signals. Barlow [7] first did observer visual signal detection experiments using random dot patterns. This was followed by a number of investigations [4,6,8,9] on detecting and discriminating simple spatial patterns (signals) in grayscale white-noise

 © 2007 Optical Society of America

images. Burgess and Colborne [9] showed that the major source of human inefficiency was due to intrinsic decision variability (internal noise). They found that there were two types of internal noise: one, referred to as static, was independent of image noise level while the other, referred to as induced, was proportional to the external noise level. The results of these investigations using white noise suggested that humans could be modeled as suboptimal Bayesian observers. This list of references is far from complete. Many vision scientists have measured human efficiency for a wide variety of signals in white noise. Geisler [10] showed that the ideal observer analysis approach was useful for tracing the flow of discrimination information for arbitrary spatiochromatic stimuli through the early stages of visual system processing. The models of human performance based on signal detection theory were successful for tasks involving white noise. At the same time, Ahumada and Beard [11] showed that a model based on image discrimination could not predict human performance when the two different noise fields were used in a two-alternative forced choice (2AFC) task.

Once the effects of white noise on performance were understood, the experimental investigation in medical imaging turned to evaluating human performance for the detection of signals in images where the noise was spatially correlated. For these tasks, the ideal observer would perform a "prewhitening" (PW) operation to remove spatial correlations in the noise before doing the cross-correlation detection procedure. The goal of human observer experiments was to determine whether we could compensate for spatial correlations. Myers *et al.* [12] used high-frequency emphasis noise images (such as those found in computed tomography) created using ramp filters. Rolland and Barrett [13] used "lumpy" backgrounds (such as those found in nuclear medicine images) created using randomly distributed Gaussian lumps and Poisson noise (which gave a low-pass power spectrum). Burgess *et al.* [14,15] used two-component noise (white plus low-pass filtered). These experimental results suggested that humans could not do complete PW (decorrelation). Myers and Barrett [16] proposed the most successful model class: a PW linear discriminator with spatial frequency channels. This will be discussed in detail later. Wagner and Weaver [17] had suggested another class of observer models referred to as nonprewhitening (NPW) matched filters. In that class, the observer behaves as if the noise is white and uses the signal as a template—no matter what power spectrum the noise has. It should be noted that many common models used in vision science are implicitly NPW. Burgess *et al.* [14,15] demonstrated that NPW models cannot always predict human performance for signal detection in filtered noise. The important point is that the ideal observer uses the spatial statistics of the backgrounds to select the appropriate filter (or template) to employ during signal detection. Experimental results suggest human observers also do this to some degree. Therefore it is worthwhile to consider human performance for the sorts of spectra encountered in important types of scenes and images.

A number of authors, Field [18], Pentland [19], Ruderman and Bialek [20], and Webster and Miyahara [21] among others, measured the power spectra, $P(f)$, of ensembles of luminance images of natural scenes: where f is the radial frequency and spectra averaged over angle. They found that spectra had the power-law form,

$$P(f) = K/f^{\beta} \tag{3}$$

and the exponent, β, had values in the range from 1.5 to 2.0. This topic is of interest because efficient visual scene processing and representation takes advantage of the knowledge of this input statistical structure. For example, Barlow [22] suggested that one important goal for early sensory coding would be the removal of redundant information. Also, it should be noted that a power-law spectrum indicates that the spatial variations do not have a defined correlation distance—they are scale invariant.

The detection of mass lesions in mammograms is limited by normal patient structure in the images. The structure has been found to be nearly random, with a power spectrum that is approximately isotropic and a radial frequency dependence of the power-law form. A number of investigators [23–26] have found that the average value of the exponent, β, is approximately 3. It has been found that individual mammograms have exponents ranging from 1.5 to 4.0. Bochud *et al.* [24] compared human and model observer 2AFC detection of a simulated mass in digital mammogram backgrounds and filtered noise backgrounds with the same power spectrum. They found that mammographic backgrounds did not degrade 2AFC detection performance as much as would be expected on the assumption that the structure can be treated as a form of pure random noise. Burgess *et al.* [25] compared human 2AFC detection for mammograms and $1/f^3$ noise with matched statistics and also found that performance was significantly better with mammograms. This agreed with the findings of Bochud *et al.* [24] and suggests that mammogram structure cannot be considered to be "pure" noise. Burgess *et al.* [25] also conducted search experiments with the two types of backgrounds. For this case the results were very different: performance was the same for the two types of backgrounds. This suggests that for search, mammogram structure can be considered to be pure noise. The data were presented in the form of a contrast-detail (CD) diagram, which is a plot of the log (signal amplitude) required for a selected detection accuracy as a function of the log (signal diameter). They showed that under aperiodic signals and certain conditions, the ideal observer theory predicts a linear relationship between the spectral exponent, β, and CD diagram slope, m:

$$m = (\beta - 2)/2. \tag{4}$$

For the case of β greater than 2, the CD diagram slope is predicted to be positive. This was a very surprising result, because for all previous visual signal detection investigations, the signal amplitude (or contrast) required for detection decreased as signal size increased. Our human observer experimental results [25] agreed with this prediction. In mammograms, contrast thresholds increased with increasing signal size for signals larger than ~1 mm, with CD slopes of ~0.3 for humans and 0.4 for model observers. Given the average power-law exponent of ~3, one would have expected a CD slope of ~0.5 for mammograms. However, the mammogram situation is

complicated by the fact that the power spectrum varies from image to image (and possibly within an image) over a range of power-law exponents from ~1.5 to 3.5. The CD diagram slope for $1/f^3$ filtered noise backgrounds was 0.44 for human observers, which is in good agreement with the predicted result of in Eq. (4).

As was mentioned earlier, Eq. (4) is valid only under certain conditions. One is that as the size increases, the 2D spatial profile must be scaled or, as an alternative description, the form of its amplitude spectrum remains constant. Another condition is that the integral describing the detectability index for the ideal observer must converge. Several other papers have been published concerning human observer performance with $1/f^3$ noise backgrounds. First we will describe papers concerned with the effects of signal properties. Chakraborty and Kundel [27] determined the CD diagrams of Gaussian and smoothed disk signals in mammographic backgrounds. For the Gaussian signals, with a constant 2D profile, they found a CD slope of 0.2. For the disk signals with blur held proportional to the disk radius and hence a constant 2D profile, the threshold contrast was almost independent of radius. For disk signals with fixed blur edges, for which the 2D profile was not constant, they observed a CD slope of −0.28. Burgess [28] calculated the performance of a variety of observer models using designer signals (described in Section 3) with a number of edge profiles determined by a parameter v. Human observer performance was also evaluated by experiment. The human CD slope for unfiltered disks (a designer signal with $v=0$) was −0.06 and −0.17 for a disk filtered using a Gaussian filter with a standard deviation of 1 pixel. The next group of papers concerns experiments done using both different (unpaired) noise realizations for the two background fields and twin (identical or paired) background fields. Johnson *et al.* [29] presented 2AFC Gaussian signal detection results for uniform (noiseless) backgrounds, and the unpaired and paired $1/f^3$ filtered noise conditions. They obtained a CD diagram slope of −0.32 for the noiseless background case. For the case of the different $1/f^3$ filtered noise realization the slope was +0.22, which is less than our value of +0.45. But it was still a positive slope. There were some differences in experimental methods. Johnson *et al.* [29] used a fixed viewing distance and did not provide signal location cues, while in the other experiments [25,27,28] observers were free to vary viewing distance and location cues were provided. We [30] presented preliminary CD slope results for both different and twin backgrounds with a range of filtered noise exponents (1.5–3.5). For these experiments, viewing distance was not constrained and location cues were provided. The results for $1/f^3$ noise did not agree with those of Johnson *et al.*, for either different or twin backgrounds. Johnson *et al.* published a subsequent paper [31] using revised experimental methods and obtained results that were in qualitative agreement with ours [30] in that they obtained a positive CD slope for their twin $1/f^3$ noise backgrounds. Nafziger *et al.* [32] attributed the previous disagreement mainly to signal location uncertainty since Johnson *et al.* [29] had not provided the observer with any fixation cues. The purpose of this work is to determine whether the CD diagram slope relationship, in Eq. (4), predicts observer results or humans detecting a simple signal in filtered noise images over a wider range of filtered noise exponents (1.5–3.5).

2. THEORY

The theoretical section is divided into two parts. In Subsection 2.A we present an analysis of the variation of signal contrast threshold as a function of signal size. This will be done using the ideal observer model with continuous 2D function descriptions in the Fourier domain. This approach does not apply precisely to our experiments that were done using discrete rather than continuous images. In Subsection 2.B we present a description of the detectability index calculation for another model observer. In principle, we could have done an analysis of the ideal observer (PW matched filter) model by numerical integration in the Fourier domain. We chose not to do this because we wanted to determine model observer performance based on the actual images used for our observer experiments. To do this for the PW matched filter (ideal observer) model, we would have needed to do a Monte Carlo simulation using spatial domain templates. This would have required us to first do inverse filtering in the frequency domain to obtain the isotropic 2D filter with radial dependence, $T(f)=S(f)/P(f)$, where $S(f)$ is the radial amplitude spectrum of the isotropic signal. The template, $t(x,y)$, would then be the inverse 2D Fourier transform of the isotropic filter. When we tried this, we were unable to obtain satisfactory templates. They appeared to have serious artifacts. So we chose to use the channelized Fisher–Hotelling (FHC) model since the channelized signal vector and the channels covariance matrix were easily determined. In Subsection 2.B we will describe the FHC observer model using vector-matrix notation. This approach is valid for discrete images.

A. Contrast Threshold versus Signal Size

In this subsection, images will be described as 2D functions, $g(x,y)$, with additive signals, $s(x,y)$, imaging system noise, $n(x,y)$, and statistically defined backgrounds, $b(x,y)$, where (x,y) are spatial coordinates. The total image with a signal added at position (x_0,y_0) is described by

$$g(x,y) = s(x-x_0,y-y_0) + n(x,y) + b(x,y). \tag{5}$$

For mathematical convenience, the noise and backgrounds will be assumed to be wide-sense stationary random processes. Burgess *et al.* [25] showed that for mammogram and power-law noise backgrounds there is good agreement between predictions based on this assumption and predictions of models that do not assume stationarity. For spatially correlated noise or backgrounds, detection model description is most convenient in the spatial frequency domain. Since the signal to be used and background spectra for our experiments were isotropic (no angular dependence) all functions will be presented in polar form. The signal amplitude spectrum will be described by its radial frequency dependence, $S(f)$, and the stochastic background component will be described by its power spectra, $P_b(f)$. Imaging system noise will be assumed to be small compared with background fluctuations and will be neglected. The ideal PW observer signal-to-noise ratio

(SNR), or equivalently the detectability index (d'), equation for the signal-known-exactly (SKE) detection task is

$$\mathrm{SNR}^2 = (d')^2 = 2\pi \int_0^\infty \frac{|S(f)|^2 f df}{P_b(f)}. \quad (6)$$

The equation describing the dependence of the SNR on signal size (for a selected functional form) for the PW observer will be calculated for a special case to illustrate the general form of the expected result. The signal will be assumed to be isotropic with spatial domain polar equation, $s_R(\rho) = \alpha s(r/R)$, where ρ equals r/R, r is the radial distance from the center, and R is a positive real-valued size scaling factor. The profile, $s(r/R)$, has unit amplitude, and α is an amplitude-scaling factor. The Hankel transform of this scalable signal, which does not change the 2D profile as size increases, is

$$S_R(f) = \alpha R^2 S(Rf). \quad (7)$$

The stochastic background will be assumed to be a wide-sense stationary random process with a 2D power spectrum slice described by the equation $P_b(f) = K/f^\beta$. Substituting in Eq. (6), the signal detectability index for the PW observer under SKE conditions becomes

$$(d')^2 = 2\pi \int_0^\infty \frac{|S_R(f)|^2 f df}{P_b(f)} = \frac{2\pi\alpha^2 R^4}{K} \int_0^\infty |S(f)|^2 f^{(1+\beta)} df. \quad (8)$$

A detection threshold amplitude, A_t, can be defined using the criterion of d' equal to some arbitrarily selected constant value. Burgess *et al.* [25] showed that the relationship between log (amplitude threshold) and log (signal size scale) then becomes

$$\log(A_t) = C + m \log(R), \quad \text{with } m = (\beta - 2)/2. \quad (9)$$

The plot of $\log(A_t)$ versus $\log(R)$ is referred to as the contrast-detail (CD) diagram. The parameter C is a constant for a given scalable signal profile and fixed power spectrum parameters. The parameter m is the slope of the CD diagram. As an example, for a power-law exponent of 3, the predicted CD diagram slope is 0.5. There are limitations to the application of this heuristic argument about CD diagram slopes. First, it is necessary to observe the assumption of profile scaling. If the (radius normalized) signal profile changes as the radius changes, then the analysis is no longer valid. An alternative way of stating the requirement is that the frequency-normalized form of the signal spectrum must not change. For example, the scaled spectrum form would change when a filter of fixed bandwidth smooths a scalable signal. A more serious limitation is that the energy spectra of some signals do not decay fast enough with frequency to allow the convergence of the integral in Eq. (8). This is particularly important for the analysis of digital signals, where the integral should approach its limiting value by the Nyquist frequency (0.5 cycles/pixel). The signal used in our experiments will be described later.

B. Channelized Fisher–Hotelling Observer Model

We will use a spatial domain method to assess the performance of a model observer for the signal and filtered noise backgrounds used in our human observer experiments. Fiete *et al.* [33] suggested this approach, which they referred to as the Hotelling observer. We refer to it as the Fisher–Hotelling (FH) observer model because it is, it fact, a Fisher linear discriminant function analysis. The only difference between the Fisher and Hotelling analysis is the manner in which the covariance matrix is obtained. Hotelling assumed that the covariance matrix is known *a priori*, while Fisher estimated the covariance matrix from ensemble data. For the complete FH model the image and its components are described by vectors $\mathbf{g}$, $\mathbf{s}$, $\mathbf{n}$, and $\mathbf{b}$ in lexicographic order with the dimension equal to the number of image pixels, N_p. Noise and background fluctuations are described by their covariance matrices, $\mathbf{K}_n$ and $\mathbf{K}_b$. The complete FH model is difficult to use because of the large size of the covariance matrices (N_p by N_p). Barrett *et al.* [34] suggested a method of overcoming this dimensionality problem: by representing the image data in the vicinity of the signal using smooth functions of the pixel values rather than the pixel values themselves. This is similar to the receptive fields (or equivalently spatial frequency channels) assumed to be used by the human visual system. Barrett *et al.* [34] chose to use Laguerre–Gauss (LG) functions for the radial component, the product of Laguerre polynomials and Gaussians. The details of the LG functions will be described later in Section 3. Typically a total number, N_c, of five to eight radial basis functions (channels) are adequate for the simple isotropic signals and the power spectra found in medical imaging signal detection investigations. Each channel is described in the spatial domain by a basis function $t_k(x,y)$ or alternatively by a vector, $\mathbf{t}_k$. The channels set $\{\mathbf{t}_k, k = 1,2,\ldots,N_c\}$ is combined to form an $N_p \times N_c$-dimensional matrix, $\mathbf{T}$, whose columns are the individual channel vectors. The signal response vector after the channels is

$$\mathbf{s}_C = \mathbf{T}^t \mathbf{s}, \quad (9')$$

with a dimension N_c. Similarly the response vectors, $\mathbf{g}_C$, $\mathbf{n}_C$, and $\mathbf{b}_C$ describe the image, noise, and background as seen through the channels. The covariance matrix of the image data after the channels becomes $\mathbf{K}_C = \mathbf{T}^t \mathbf{K}_g \mathbf{T}$, with a dimension $N_c \times N_c$ and $\mathbf{K}_g = \mathbf{K}_n + \mathbf{K}_b$. However, one can determine the elements of $\mathbf{K}_C$ directly from the image collection. We used this approach and will describe it in Section 3. The detectability index (or alternatively the SNR) for this FHC observer model is given by

$$(d')^2 = \mathrm{SNR}^2 = \mathbf{s}_C^t \mathbf{K}_C^{-1} \mathbf{s}_C. \quad (10)$$

3. METHODS

A. Human Observer Experiments

Many of the human observer experimental method details have been described in detail elsewhere [25]. The experiments were done using a MacIntosh computer with a DOME model MD2 display board (Planar Systems, Inc., Beaverton, Oreg.) and a Clinton model DS2000HB grayscale monitor [1600(H) × 1200(V)] (Clinton Electronics, Loves Park, Ill.) with a maximum luminance of 210 cd/m^2. The task was the 2AFC detection of a designer signal (described later) in $1/f^\beta$ power-law noise. The two

noise fields were 512×512 pixels with a pixel size of 0.32 mm on the monitor. An example image display is shown in Fig. 1.

There were 30 experimental conditions: five noise power-law exponents and six signal diameters. Two very experienced observers (the authors) took part in the experiments. Each observer first did one block of 128 practice trials and then two blocks of 128 data collection trials per experimental condition. We compared the thresholds for pairs of data collection blocks to check for learning effects and to estimate the standard deviations of thresholds for the observers. The results will be given later.

A fixed experimental condition was used for each block of 128 decision trials and the signal amplitude was varied using a staircase procedure. Our approach is a simplified version of the QUEST Bayesian adaptive psychometric method developed by Watson and Pelli [35]. We used a fixed sequence of signal amplitude steps spaced at ~10% intervals, so spacing would be constant on a log scale. We used a constrained "1-up/6-down" procedure. The goal was to keep the observer near the amplitude corresponding to ~90% correct. A value in this vicinity gives the lowest coefficient of variation of the estimate of d' for a given number of decision trials [36]. By a constrained procedure, we mean that the threshold amplitude was estimated periodically during trials and used to select a "best" step number. The 1-up, 6-down procedure is then forced to stay in a step range of $\pm N$ around the most recent best step number. The initial choice for the best step was based on practice trial results. The initial sequence of data collection was done using a step range of ±5, and a new value for the best step was estimated after each 16 trials using the regression analysis method described later. After 64 trials the step range was reduced to ±3. The step adjustments were made separately for each observer. A high contrast reference copy of the signal was shown above the two backgrounds, and the two alternative signal locations were identified using unobtrusive circle cues [37] centered on the signal to assist the observer. There was no time limit for a decision trial, and feedback as to the correctness of response was provided.

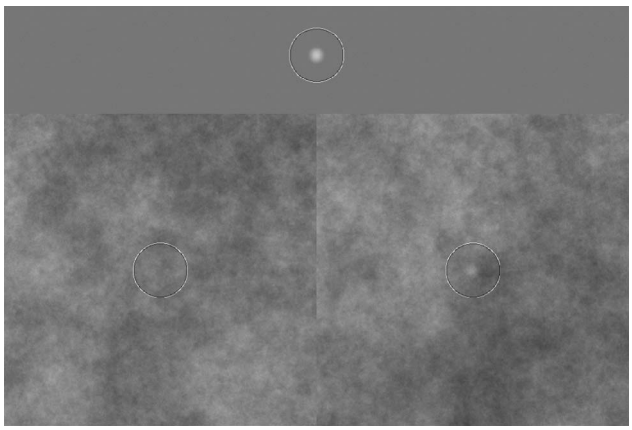

Fig. 1. Example image for the 2AFC detection task with a nodule signal diameter of 32 pixels and a power-law noise exponent of 2.0. A high amplitude reference version of the signal is shown above the noise fields. The location reference circles are exaggerated in the manuscript image but are unobtrusive low contrast, concentric bipolar circles on the experimental display monitor.

The observer was free to select any desired viewing distance. We found that the selected viewing distance varied with signal size and ranged from ~30 cm for the smallest signal diameter (0.26 cm=8 pixels) to greater than 100 cm for the largest signal (4.1 cm=128 pixels). We defined the threshold as the signal amplitude (contrast) required for 92% correct performance ($d'=2$) to the 2AFC detection task. There were two reasons for selecting this value. One is that Judy and Swennson [38] found that for the detection of disc signals in computed tomography noise, the observer's subjective estimates for just detectable contrast corresponded to the value that gave a detectability index of 2 as measured objectively using the receiver operating characteristic (ROC) method [2]. The second reason is that d' equal to 2 is near the value of d' used as a goal in the staircase method used. As was mentioned earlier, collecting data near d' equal to 2 minimizes the coefficient of variation of the estimate of d'.

The data analysis was done as follows: For each amplitude step in each experimental condition, the percentage of correct responses was calculated and converted to a detectability index, d'. If the percentage correct for a particular step was zero or 100% for a particular step, this step data was not used. This situation occurred for steps that included ~12% of decision trials (based on 256 trials per condition). A weighted regression fit of d' versus signal amplitude was used to determine the threshold amplitude. The weighting was done using a binomial statistics estimate of the variance of d' for that particular value of d' and the number of associated trials. Given the relatively small number of trials for each step (an average of 42), the slope of the regression line could be quite variable. To stabilize the regression line, it was forced to pass through d' equal to −0.3 at an amplitude of zero. This intercept had been determined from previous experiments [6] using the method of constant stimulus with a number of signal amplitudes and 1024 trials per datum. We had previously determined that threshold estimation was not sensitive to the particular intercept used. There was no significant difference between an intercept of −0.3 and an intercept of zero. Thresholds were determined for each of the two observers, and a two-observer average was calculated. The average result was obtained by assuming that there was no systematic difference in the results for the two observers. As will be seen in Section 4, this was a reasonable assumption. So for each experimental condition, response results for the two observers were combined to obtain an overall percentage of correct responses. This was then converted to a d' score, and a regression fit was used to obtain the two-observer average threshold amplitude.

B. Signal and Noise Properties

We used an isotropic designer signal with the spatial domain equation

$$s(r) = A\ \mathrm{Rect}(2\rho)[1-\rho^2]^v, \tag{11}$$

where ρ is a normalized distance (r/R), R is the nodule radius, A is the signal amplitude, and the value of v was chosen to be 2. Five exponents were used for the power-law noise, with values of β from 1.5 to 3.5 in steps of 0.5. This covered the range of exponents found for mammo-

grams. The noise backgrounds were produced by frequency domain filtering of zero-mean, Gaussian white noise images. Care is needed in production of power-law filtered noise because of the problem of "data wrap-around" due to the implicitly periodic nature of the discrete Fourier transform when doing the inverse transform from the frequency domain. We started with 2048 $\times$2048 pixel white-noise images, transformed to the frequency domain and filtered using the isotropic filter, $H(f)=C/f^{\alpha}$ where f is radial frequency, and with α equal to $(\beta/2)$. We then did the inverse DFT and selected the central 512$\times$512 region for use in observer experiments. Note that the filter function would become infinite at zero frequency. To prevent this, we set that filter value, $H(0)$, to be twice that of the first nonzero radial frequency. For our initial work with $1/f^3$ filtered noise [25], we tried a number of ratios for $H(0)$ relative to $H(f)$ at the first nonzero radial frequency, from 1 to 8. The detectability index for the ideal observer was evaluated by numerical integration using these values of $H(0)$, Eq. (8), and the designer signal described earlier. We found no significant difference in results. The $1/f^3$ noise was linearly transformed to have an ensemble mean of zero and a pixel standard deviation of 790 when measured within a square 192 pixels on a side. It should be noted that the pixel standard deviation for power-law noise is a function of the measurement area [25]. The values of the spectral density scaling factor, K, for other power-law exponents were selected so that the CD diagrams for the FHC observer would all cross at a signal diameter of 32 pixels. A total of 1024 images were produced for each exponent. Local mean values of power-law noise are highly variable, and it was desirable to keep potential signal locations away from extreme luminance values in the displayed image. So the data values of each image were shifted to give a mean of zero within the central 192$\times$192 square. This size was 50% larger than the diameter of the largest signal, and the mean value shift would ensure that the potential signal regions would fall within the center of the displayed image gray scale. For display, two randomly selected noise fields were loaded into the display image array, and the signal was added to the appropriate field. The display image data were then linearly transformed to fit within an 8-bit range with a mean of 128. The same linear transformation was used for all display images under all experimental conditions. Therefore the same (contrast) amplitude scale applied to all signals and noise fields. Example noise fields for the five power-law exponents are shown in Fig. 2.

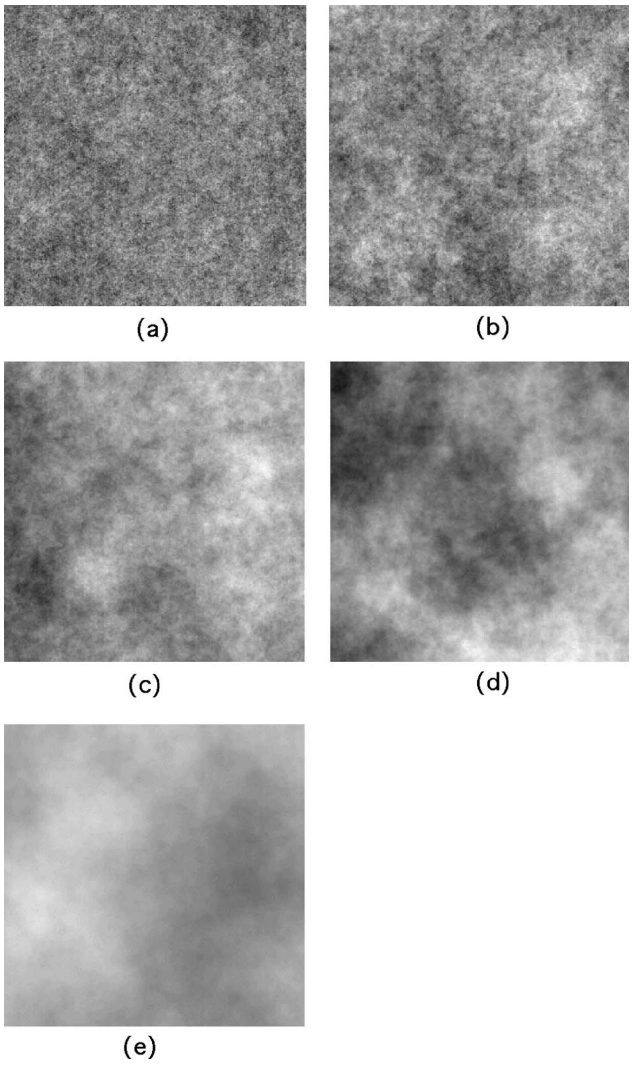

Fig. 2. (a)–(e) Examples of power-law noise background images with isotropic power spectrum $P(f)=1/f^{\beta}$. Exponents are (a) 1.5, (b) 2.0, (c) 2.5, (d) 3.0, (e) 3.5.

C. Channelized Fisher–Hotelling Model Performance Evaluation

As an independent check on the ideal observer model prediction, the performance for the channelized FHC model was determined using the same filtered noise images that were used for the human observer experiments. The signal was isotropic, so isotropic basis functions could be used. The equation for the radial dependence of the nth-order basis function (in orthonormal form) is

$$B_n(r)=\frac{2\sqrt{\pi}}{a}\exp\left(-\frac{\pi r^2}{a^2}\right)L_n\left(\frac{2\pi r^2}{a^2}\right), \tag{12}$$

where $L_n(x)$ is the nth-order Laguerre polynomial. We had previously found [25] that for nodule detection in $1/f^3$ noise, the SNR of the FHC observer model depended on the value of the parameter a and that the optimum was ~0.75 times the nodule support radius, R. Let $\alpha=a/R$. Calculations of SNR were done using all signal diameters and filtered noise exponents. This investigation showed a broad peak in the SNR versus α relationship. The SNR values had maxima near $\alpha=0.8(\pm0.1)$ with a weak dependence on filtered noise exponents and nodule size. Therefore a value of α equal to 0.8 was used for all investigations. The SNR of the FHC observer model also depends on the number, N_c, of basis functions used. We had previously found [25], for nodule signal detection in $1/f^3$ noise, that no increase in SNR occurred on going from five to six LG basis functions. So the spatial domain calculations were done using Eq. (10) with six basis functions. The vectors and the covariance matrices were determined using the following procedures: The signal response vector element values were obtained by cross correlating the signal vector with each LG basis function.

The covariance matrix, $\mathbf{K}_C$, for the collection of background images (with no signal present) was estimated by cross correlating each basis vector with each region of interest (ROI) to obtain the result $\mathbf{g}_C = \mathbf{T}^t\mathbf{g}$, and then using the following equation (with $E[\ldots]$ indicating ensemble averaging):

$$\mathbf{K}_C = E[(\mathbf{g}_C - E[\mathbf{g}_C])(\mathbf{g}_C - E[\mathbf{g}_C])^t]. \quad (13)$$

4. EXPERIMENTAL RESULTS

We checked for learning effects and determined threshold statistics in the following manner: First, the ratio of threshold values for the two blocks of trials for each experimental condition and observer were calculated. Then the mean and standard deviation of the set of 30 ratios for each observer were determined. The means and population standard deviations of the ratios for observers 1 and 2 were 0.98 (±0.21) and 1.00 (±0.19), respectively. This indicates that there were no significant learning effects. Not all trials were used in the analysis because some staircase steps had either zero or 100% correct responses. On average 13% of trials in each block of 128 were not used. The population standard deviations of the ratios were 15% higher than would be expected based on binomial statistics for 111 trials per observer per condition per block. When the two blocks per condition were combined, the average fraction of trials used was slightly higher (88%) because of better sampling statistics. Based on the results from the population standard deviations of the 128 block ratios, we estimate that the coefficients of the variation of the human observer threshold estimates for 256 trials (225 valid trials) are 10% for one observer and 7.1% for the average data for two observers.

The results are presented in Figs. 3(a)–3(e) with power-law spectrum exponents from 1.5 to 3.5. Each figure shows the logarithm of the threshold amplitude (defined using d' equal to 2) as a function of nodule signal diameter. In each figure there are three threshold data sets: data for each of the two observers shown by separate symbols, a dashed line showing a regression fit to the average results for the two human observers, and a solid line showing a regression fit to the results for the FHC model observer. The error bars for the data of one observer indicate 95% confidence intervals using a coefficient of variation of 10%. Consider the case of threshold data as a function of signal size for $1/f^{1.5}$ noise shown in Fig. 3(a). The expected CD slope for this exponent is −0.25 based on the ideal observer model. The slope for the FHC model is −0.25 while the slope for the average performance for two human observers is −0.29. The CD results for the other power-law exponents are shown in Figs. 3(b)–3(e). The average ratio of human to FHC model thresholds is 1.23 for β equal to 1.5. This result can be used to estimate a human efficiency of 0.66 relative to the FHC observer. We had previously found [25] that the absolute efficiency of the FHC model was 90% for filtered noise with an exponent of 3. For analysis in this paper, we will assume that this FHC model efficiency holds for all power-law exponents and use it to estimate human absolute efficiency (relative to ideal observer performance). This gives an estimate of 0.60 for the absolute efficiency with an exponent

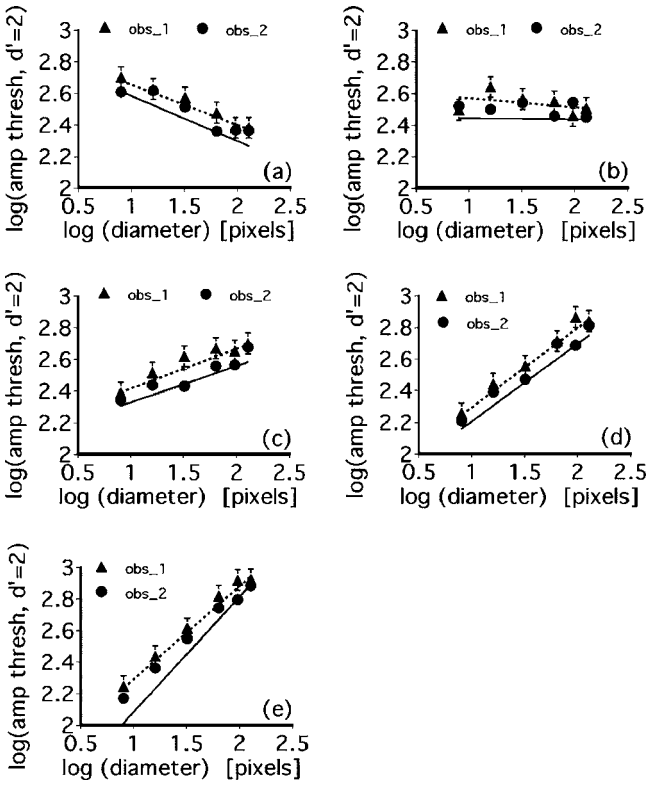

Fig. 3. (a)–(e) The CD diagram (contrast threshold for d' equals 2 versus signal diameter) results for two human observers (symbols), regression fits (dashed lines) to the averages for the two observers and regression fits the FHC model results (solid lines) for five power-law exponents: (a) 1.5, (b) 2.0, (c) 2.5, (d) 3.0, (e) 3.5. The error bars for observer 1 are 95% confidence limits.

of 1.5. The estimated absolute efficiencies for other exponents were in the range 0.4 and 0.6. The detailed slope and efficiency results are summarized in Table 1.

5. CONCLUSION

We have presented results for the 2AFC detection of an aperiodic signal in power-law noise with five different exponents that cover most of the range found in mammograms. The signal was a designer signal with an exponent of 2. The filtered noise was obtained by frequency domain filtering of white noise images. The goal of the work was to determine whether the predicted relationship, $m = (\beta - 2)/2$, between CD diagram slope and power-law exponent held for humans and a channelized FHC model observer. A summary of the human CD slopes (two observer averages) as a function of power-law exponent is shown in Fig. 4. Six signal diameters (8, 16, 32, 64, 96, and 128 pixels; 0.26–4.1 cm on the monitor) were used to estimate the CD diagrams. The coefficients of variation of the average threshold amplitudes (for two observers) were 7.1%. We found excellent agreement between the slopes predicted for the ideal observer and the measured slopes for the FHC observer for all noise exponents. There was also good agreement between the predicted slopes and human results for all exponents except 3.5. For that exponent, there was a significant difference between the predicted value of 0.75 and the human observer values of ~0.59. We have no explanation for this discrepancy. It is

Table 1. Summary of the CD Diagram Slopes for the Five Power-Law Noise Exponents Used in the Experiments[a]

Noise Exponents	Ideal Slope	FHC Slope	Obs_1 Slope	Obs_2 Slope	Average Slope	Human Efficiency (%)
1.5	0.25	−0.25	−0.28	−0.26	−0.29	60
2.0	0.00	0.00	−0.05	−0.03	−0.06	57
2.5	0.25	0.23	0.24	0.25	0.25	56
3.0	0.50	0.49	0.50	0.48	0.50	59
3.5	0.75	0.73	0.60	0.59	0.59	41

[a]The first two columns give model CD diagram slopes. The next three columns give the slopes for the individual human observers and the two observer average. The last column gives the estimated human absolute efficiencies, assuming that the FHC model observer efficiency is 90%.

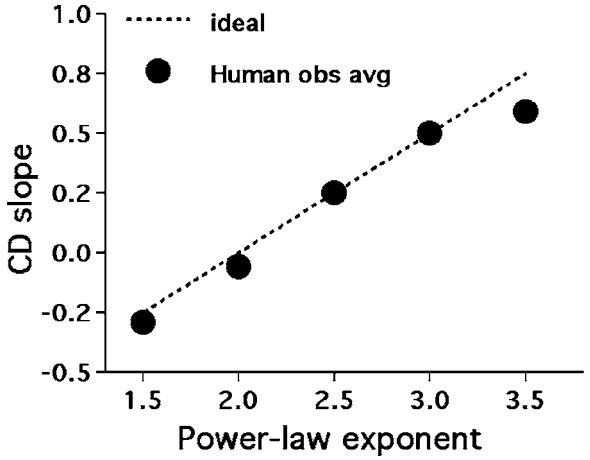

Fig. 4. Summary of estimated average human observer CD slopes as a function of the noise power-law exponent.

interesting to note that the predicted CD slope is zero for a power-law exponent of 2. That exponent is commonly found for natural scenes. A CD slope of zero means that the amplitude (contrast) threshold in such scenes were independent of signal size if they could be treated as random noise.

In summary, we conclude that the ideal observer model does a very good job of predicting the CD diagram slope for human observers for power-law noise with exponents over the range of 1.5 to 3.5. This is the range of interest in mammography. We also find that human observer efficiency is between 40% and 60% for detection of the aperiodic signal that was used.

ACKNOWLEDGMENTS

This research was supported by National Institutes of Health grant CA87738.

REFERENCES

1. J. A. Swets, *Signal Detection and Recognition by Human Observers* (Wiley, 1964).
2. D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics* (Wiley, 1966). (Reprinted by Peninsula, 1988.)
3. W. P. Tanner and T. G. Birdsall, "Definitions of d′ and η as psychophysical measures," J. Acoust. Soc. Am. **30**, 922–928 (1958).
4. A. E. Burgess, R. F. Wagner, R. J. Jennings, and H. B. Barlow, "Efficiency of human visual discrimination," Science **214**, 93–94 (1981).
5. J. Nachmias and R. V. Sansbury, "Grating contrast: discrimination may be better than detection," Vision Res. **14**, 1039–1042 (1973).
6. A. E. Burgess and H. Ghandeharian, "Visual signal detection. II. Signal-location identification," J. Opt. Soc. Am. A **1**, 906–910 (1984).
7. H. B. Barlow, "The efficiency of detecting changes in density in random dot patterns," Vision Res. **18**, 637–650 (1977).
8. A. E. Burgess, "Visual signal detection. III. On Bayesian use of prior knowledge and cross-correlation," J. Opt. Soc. Am. A **2**, 1498–1507 (1985).
9. A. E. Burgess and B. Colborne, "Visual signal detection. IV. Observer inconsistency," J. Opt. Soc. Am. A **5**, 617–627 (1988).
10. W. S. Geisler, "Sequential ideal-observer analysis of visual discriminations," Psychol. Rev. **96**, 267–314 (1989).
11. A. J. Ahumada and B. L. Beard, "Image discrimination models predict detection in fixed but not random noise," J. Opt. Soc. Am. A **14**, 2471–2478 (1997).
12. K. J. Myers, H. H. Barrett, M. C. Borgstrom, D. D. Patton, and G. W. Seeley, "Effect of noise correlation on detectability of disk signals in medical imaging," J. Opt. Soc. Am. A **2**, 1752–1759 (1985).
13. J. P. Rolland and H. H. Barrett, "Effect of random background inhomogeneity on observer detection performance," J. Opt. Soc. Am. A **9**, 649–658 (1992).
14. A. E. Burgess, X. Li, and C. K. Abbey, "Visual signal detectability with two noise components: anomalous masking effects," J. Opt. Soc. Am. A **14**, 2420–2442 (1997).
15. A. E. Burgess, "Visual signal detectability with two-component noise: low-pass filter effects," J. Opt. Soc. Am. A **16**, 694–704 (1999).
16. K. J. Myers and H. H. Barrett, "Addition of a channel mechanism to the ideal-observer model," J. Opt. Soc. Am. A **4**, 2447–2457 (1987).
17. R. F. Wagner and K. E. Weaver, "An assortment of image quality indices for radiographic film-screen combinations—can they be resolved?" Proc. SPIE **35**, 83–94 (1972).
18. D. Field, "Relations between the statistics of natural scenes and the response of cortical cells," J. Opt. Soc. Am. A **4**, 2379–2394 (1987).
19. A. Pentland, "Fractal-based descriptions of surfaces," in *Natural Computation*, W. Richards, ed. (MIT Press, 1988), pp. 279–299.
20. D. L. Ruderman and W. Bialek, "Statistics of natural images: scaling in the woods," Phys. Rev. Lett. **73**, 814–817 (1994).
21. M. A. Webster and E. Miyahara, "Contrast adaptation and the spatial structure of natural images," J. Opt. Soc. Am. A **14**, 2355–2366 (1997).
22. H. B. Barlow, "Three points about lateral inhibition," in *Sensory Communications*, W. A. Rosenblith, ed. (MIT Press, 1961) pp. 782–786.
23. B. Zheng, Y.-H. Chang, and D. Gur, "Adaptive computer-

aided diagnosis scheme of digitized mammograms," Acad. Radiol. **3**, 806–814 (1996).
24. F. O. Bochud, J. F. Valley, F. R. Verdun, C. Hessler, and P. Schnyder, "Estimate of the noisy component of anatomical backgrounds," Med. Phys. **26**, 1365–1370 (1999).
25. A. E. Burgess, F. L. Jacobson, and P. F. Judy, "Human observer detection experiments with mammograms and power-law noise," Med. Phys. **28**, 419–437 (2001).
26. J. J. Heine and R. P. Velthuizen, "Spectral analysis of full field digital mammography data," Med. Phys. **29**, 647–661 (2002).
27. D. Chakraborty and H. L. Kundel, "Anomalous nodule visibility effects in mammographic images," Proc. SPIE **4324**, 68–76 (2001).
28. A. E. Burgess, "Evaluation of detection model performance in power-law noise," Proc. SPIE **4324**, 123–132 (2001).
29. J. P. Johnson, J. Lubin, J. S. Nafziger, and D. P. Chakraborty, "Visual discrimination modeling of lesion detectability," Proc. SPIE **4686**, 248–255 (2002).
30. A. E. Burgess and P. F. Judy, "Detection in power-law noise: spectrum exponents and CD diagram slopes," Proc. SPIE **5034**, 57–62 (2003).
31. J. P. Johnson, J. Lubin, J. S. Nafziger, E. A. Krupinski, and H. Roehrig, "Channelized model observer using a visual discrimination model," Proc. SPIE **5749**, 199–210 (2005).
32. J. S. Nafziger, J. P. Johnson, and J. Lubin, "Effects of visual fixation cues on the detectability of simulated breast lesions," Proc. SPIE **5749**, 566–571 (2005).
33. R. D. Fiete, H. H. Barrett, W. E. Smith, and K. J. Myers, "Hotelling trace criterion and its correlation with human-observer performance," J. Opt. Soc. Am. **4**, 945–953 (1987).
34. H. H. Barrett, C. K. Abbey, B. Gallas, and M. P. Eckstein, "Stabilized estimates of Hotelling observer detection performance in patient structured noise," Proc. SPIE **3340**, 27–43 (1998).
35. A. B. Watson and D. Pelli, "QUEST: a Bayesian adaptive psychometric method," Percept. Psychophys. **33**, 113–20 (1983).
36. A. E. Burgess, "Comparison of ROC and forced-choice observer performance measurement methods," Med. Phys. **22**, 643–55 (1995).
37. H. L. Kundel, C. F. Nodine, L. Toto, and S. Lauver, "A circle cue enhances detection of simulated masses on mammographic backgrounds," Proc. SPIE **3032**, 81–84 (1997).
38. P. F. Judy and R. G. Swennson, "Detectability of lesions of various sizes on CT images," Proc. SPIE **535**, 38–42 (1985).