

# Human discrimination of fractal images

David C. Knill

Department of Psychology, Brown University, Providence, Rhode Island 02912

David Field\*

Physiological Laboratory, Cambridge University, Cambridge CB2 3E6, UK

Daniel Kersten†

Department of Psychology, Brown University, Providence, Rhode Island 02912

Received November 7, 1988; accepted January 20, 1990

In order to transmit information in images efficiently, the visual system should be tuned to the statistical structure of the ensemble of images that it sees. Several authors have suggested that the ensemble of natural images exhibits fractal behavior and, therefore, has a power spectrum that drops off proportionally to  $1/f^\beta$  ( $2 < \beta < 4$ ). In this paper we investigate the question of which value of the exponent  $\beta$  describes the power spectrum of the ensemble of images to which the visual system is optimally tuned. An experiment in which subjects were asked to discriminate randomly generated noise textures based on their spectral drop-off was used. Whereas the discrimination-threshold function of an ideal observer was flat for different spectral drop-offs, human observers showed a broad peak in sensitivity for  $2.8 < \beta < 3.6$ . The results are consistent with, but do not provide direct evidence for, the theory that the visual system is tuned to an ensemble of images with Markov statistics.

## INTRODUCTION

One goal of the human visual system is the efficient transmission of information in images of our environment. From this point of view, the visual system should be studied as a communication device.<sup>1-4</sup> It is a combination of a physical channel and a coding device. The physical channel consists of the eye and the neural components of the system. The coder is the organization of these components (e.g., the receptive fields and pattern of interconnections of neurons). The information capacity of the visual system is limited by physical constraints on its components, including both optical constraints, such as the diffraction limit on resolution and photon noise, and constraints on the nervous system, such as the number of available neurons, their dynamic range, intrinsic noise, and a limit on the available metabolic resources. Because of this capacity limitation, the coding scheme should be matched to the statistical structure of the ensemble of images that it will see. For humans, this is the ensemble of images of both the natural and synthetic environments. We will refer to these images generically as natural images. As a trivial example of the relationship between the visual system and the statistics of the images that it sees, consider a world in which all surfaces have the same spectral reflectance properties. Our color-vision mechanisms would merely provide a redundant coding of the information in images of this environment. They would reflect an inefficient allocation of the capacity of the communication channel.

These ideas have motivated a number of recent computational models of neural function in the visual system. Srinivasan *et al.*<sup>5</sup> show that a lateral inhibitory type of receptive field can provide an optimal code for ensembles of images

that have an exponentially decaying spatial autocorrelation function. The spatial-frequency tuning and selectivity of neurons in primary cortex may also be related to efficient coding.<sup>6,7</sup> More-recent research<sup>8-10</sup> has concentrated on the modeling of developmental processes that can lead to the evolution of codes in neural networks with repeated exposure to example images. One can relate the receptive fields in any of these models to the problem of efficient coding of images from a defined ensemble.

A full understanding of the performance of the visual system in its role as a communication device requires that one study it in the context of the environment in which it operates. A logical question to ask is: How effective is the visual system in transmitting information about natural images? Any given transmission system (coder and channel) will more efficiently transmit the information in images from some ensembles than from others; thus we can say that it is tuned to a particular ensemble.<sup>11</sup> This observation leads to a natural generalization of the question: For what ensemble(s) of images is the visual system an efficient transmission device? As a stochastic ensemble is defined wholly by its statistical structure, the question could be rephrased to ask what the statistical structure of this ensemble is. One can then ask whether this structure matches that of natural images and, if not, why.

In this paper we describe a psychophysical experiment designed to explore the statistical structure of the ensemble of images to which the human visual system is tuned. As an experimental measure, we use subjects' performance on a discrimination task for images drawn from ensembles with different correlational structures. Here we have simplified the question: What is the correlational structure of the ensemble of images for which the visual system is an efficient

transmission device? Clearly a full answer to the more general question would require looking at differences in other statistics as well.

The first step in our study requires a specification of which ensembles to use in the experiment, that is, which ensembles will form the basis of our search. A model for the statistics of natural images will motivate our selection.

### Natural Image Statistics: The Fractal Model

The ensemble of images that we consider contains those generated by the perspective projection of three-dimensional scenes onto an imaginary, finite-sized planar surface. We imagine the projection to be through an ideal camera in which the effects of diffraction, aberrations, and discrete sampling are absent. Let  $I(x, y)$  be a two-dimensional random field specifying the light flux incident upon the image plane. We will assume that  $I$  is stationary and isotropic.<sup>12</sup> We are interested in characterizing the correlational structure of  $I$ , given by the autocorrelation function  $R_I(\Delta r)$  or, equivalently, by its power spectrum<sup>14</sup>  $P_I(f_r)$ .  $\Delta r$  is the Euclidean distance between points in an image, and  $f_r$  is radial spatial frequency.

Little empirical research has been done on estimating the correlational structure of natural images, though several models have recently been proposed. Julesz<sup>15</sup> refers to his doctoral dissertation, in which he estimated the autocorrelation function for television images and found it to drop off exponentially with the distance between points. Srinivasan *et al.*<sup>5</sup> also suggest an exponential model for natural images. In their paper, however, they present data from only one image, and their proposal is based on a qualitative analysis of the empirically determined autocorrelation function.

We would like to consider a broader class of ensembles than those with exponential autocorrelation functions. The autocorrelation function of such an ensemble is given by

$$R_I(\Delta r) \propto e^{-k\Delta r}, \quad (1)$$

where the constant  $k$  specifies a spatial scale for luminance changes in images from the ensemble. A search through this set (i.e., a search through the set of values of  $k$ ) amounts to a search for the natural scale of images. Objects in our environment have a large range of sizes. Moreover we view them at a range of distances, so that the projection of any one object will vary in size between images. One might expect, therefore, that the statistics of natural images would be scale invariant, or at least vary smoothly across scales. This is the defining characteristic of a class of functions known as random fractals<sup>16,17</sup> (see Appendix A for a more comprehensive discussion of fractals). Differences in the one parameter characterizing fractal ensembles, their fractal dimension, reflect differences in their statistical structure at all scales. The fractal model thus provides a more appropriate set of ensembles for our investigation.

The particular model that we will consider is fractional Brownian motion, as originally defined by Mandelbrot.<sup>16</sup> This model is a generalization of the well-known Brownian motion, or Wiener process. Fractional Brownian motion is a Gaussian process; therefore it can be fully specified by its power spectrum. The power spectrum of two-dimensional fractional Brownian motion is given by

$$P_I(f_r) \propto \frac{1}{f_r^\beta} \quad (2)$$

for an isotropic ensemble. The exponent  $\beta$  is related to the fractal dimension  $D$  by

$$\beta = 8 - 2D. \quad (3)$$

Figure 1 shows the power spectra of two fractal ensembles, one with a fractal dimension of 2.5 and one with a dimension of 3.0. Most of the information in these ensembles is in the lower frequencies, since the variance at those frequencies is highest. As the fractal dimension is increased, the variance, and thus the information content, becomes spread more evenly across the spectrum.

The fractal model, to some degree, subsumes Gaussian ensembles with exponential autocorrelation functions. Such an ensemble behaves like a fractal ensemble with a fractal dimension of 2.5 at scales below that defined by its space constant  $k$ ; that is, its power spectrum exhibits power-law behavior at frequencies greater than  $k^3$ .

Pentland<sup>18</sup> proposed that fractional Brownian motion might be used to model textures in images. More recently, Field<sup>7</sup> has suggested that the ensemble of natural images has a fractal dimension of 3. He bases this on the sample power spectra of several images of natural scenes (e.g., mountains, water, rocks) that seemed to drop off as  $1/f_r^2$  and on the argument that the average rms contrast in images should not vary with viewing distance (a constraint uniquely met by a power spectrum of  $1/f_r^2$ ).

### Human Discrimination of Fractal Dimension

In the Introduction, we posed the question, What is the correlational structure of the ensemble of images to which the visual system is optimally tuned? In the experiment presented here, we consider the class of fractal ensembles, in particular, ensembles that approximate fractional Brownian motion. The correlational structure of these ensembles, as specified by their power spectra, is completely characterized by one parameter, the fractal dimension  $D$ . Because the fractal dimension is linearly related to the exponent of spectral drop-off,<sup>19</sup> we can replace the fractal dimension with the exponent  $\beta$  in our discussion without loss of generality. For convenience we will refer to ensembles as being specified by this exponent, which we will call the spectral drop-off. We can now rephrase the question: What is the spectral drop-off of the ensemble to which the visual system is optimally tuned? The use of fractional Brownian motion as the specific fractal model for our study has the advantage that the statistical structure of these ensembles, being Gaussian, is completely specified by their power spectra. We therefore do not have to consider the effects of changes in higher-order statistics between ensembles, making the model ideal for asking questions focused on the correlational structure of images.

The central idea behind the experiment is that human subjects' performance on psychophysical tasks using images from different ensembles is dependent on the ability of the visual system to transmit information about the images in those ensembles. If their performance is averaged over a large number of representative samples from each ensemble, it provides a measure of this ability for the different ensem-

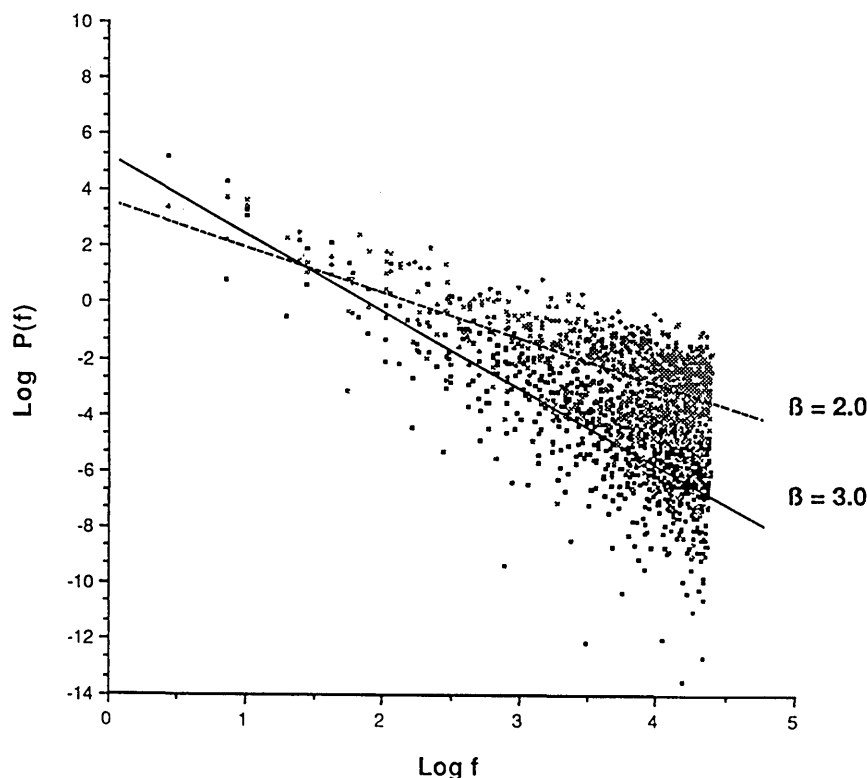


Fig. 1. One-dimensional slices through the power spectra of two fractional Brownian motion ensembles, one with fractal dimension  $D = 2.5$  ( $\beta = 3.0$ ) and the other with fractal dimension  $D = 3.0$  ( $\beta = 2.0$ ). On a log-log plot, the spectra fall off linearly, with a slope given by exponent  $\beta$ . The scatter plots show the actual spectra of sample images drawn from these two ensembles, with the power along different orientations collapsed onto one dimension.

bles. We use an experimental task in which subjects are asked to discriminate between images drawn from ensembles with different spectral drop-offs. Specifically they are asked to determine which of two randomly generated images is drawn from an ensemble with lower spectral drop-off (higher fractal dimension). Figure 2 shows examples of three such images. In this example the difference in spectral drop-off from one to another of the three images is rather large ( $\Delta\beta = 1.0$ ), and the image in Fig. 2(c) is easily seen to come from an ensemble with a lower drop-off. As fractal dimension, and thus spectral drop-off, is correlated with roughness, the task is essentially one of determining which of two images has a rougher texture.

Though the experiment presented here is designed to investigate a question about the information coding and transmission characteristics of the visual system, it bears an obvious relation to research in texture discrimination. Early research on texture discrimination focused on phenomenological tests of Julesz's<sup>20,21</sup> conjecture that first- and second-order image statistics could be used for texture discrimination, but third- and higher-order statistics could not. Counterexamples to this conjecture caused a switch in emphasis toward investigation of textures composed of featural elements.<sup>22</sup> Though it has been clearly demonstrated that humans could discriminate patterns based on their second-order image statistics, little research has been done to provide a detailed characterization of this ability. Pratt *et al.*<sup>23</sup> showed that humans can discriminate Gaussian textures with exponential autocorrelation functions based on the

space constant  $k$  of the function; however they did not obtain any quantitative measure of performance. More recently, Rensink<sup>24</sup> measured the discriminability of one-dimensional fractal textures based on their spectral drop-off and found that performance was best for the largest drop-off used,<sup>25</sup>  $\beta = 4$ . In the current experiment, we extend the investigation to two-dimensional random textures and collect data for a larger number of spectral drop-offs. (Rensink used four spectral drop-offs, between 1 and 4.) Furthermore we show, by using simulations of an ideal observer, that the discrimination thresholds collected are absolute measures of performance.

#### Ideal Observer for Texture Discrimination

Performance on a signal discrimination or detection task is a function both of factors in the observer and of the information content of the stimuli. The performance of a statistically optimal, or ideal, observer depends only on the information content of the stimuli. The ideal observer provides a benchmark with which to compare human performance, in order to draw inferences about what aspects of performance are related to factors within the observers. Typically, efficiency measures are derived to reflect the relative sample sizes needed by ideal and human observers to achieve equivalent levels of performance on specific tasks.<sup>27</sup> Efficiencies defined in this way provide a dimensionless measure that may be used to compare human performance not only in a task at different levels of some independent variable but also between tasks.

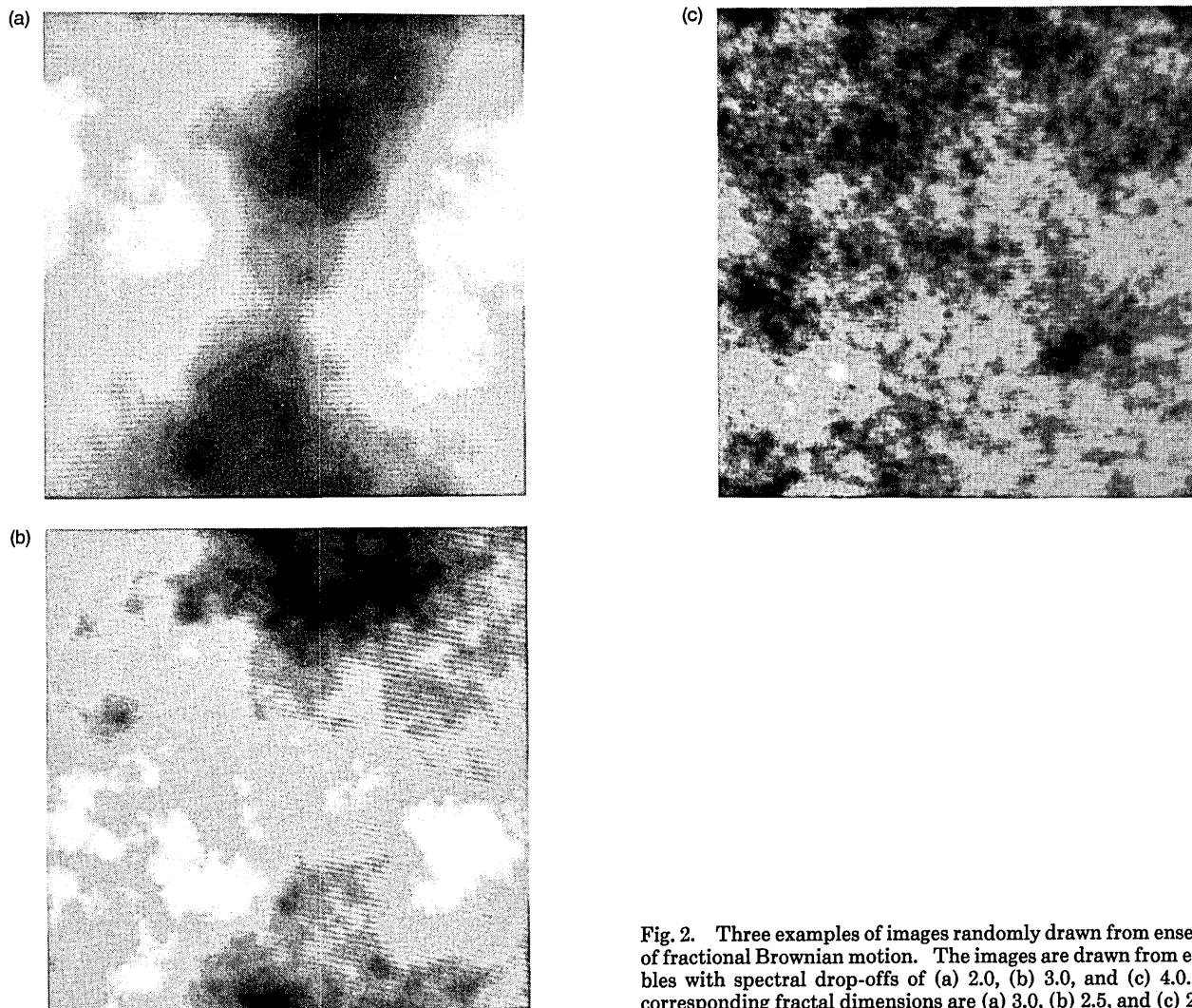


Fig. 2. Three examples of images randomly drawn from ensembles of fractional Brownian motion. The images are drawn from ensembles with spectral drop-offs of (a) 2.0, (b) 3.0, and (c) 4.0. The corresponding fractal dimensions are (a) 3.0, (b) 2.5, and (c) 2.0.

Ideal observers have been used to estimate human efficiency in a number of visual tasks, including the detection of light in scotopic<sup>28</sup> and photopic conditions,<sup>29</sup> the detection of changes in dot density in a display of random dots,<sup>27</sup> and the detection and discrimination of simple luminance patterns in noise.<sup>30,31</sup> In this study, we show that the discrimination threshold of the ideal observer is constant for different spectral drop-offs. We can therefore use changes in human subjects' thresholds for different drop-offs to draw conclusions about what they are tuned to in an absolute sense.

## METHODS

### Apparatus

Stimuli were generated and displayed on a Sun 3/160 workstation with an 8-bit gray-level display. The screen had a power-law nonlinearity, with a gamma exponent of 2.67. We modified the lookup table to linearize the display. The gamma correction<sup>32</sup> lowered the effective resolution of the screen to 7.27 bits. After gamma correction, midgray for the screen (gray level 127) had a luminance of 37 cd/m<sup>2</sup>. The spatial resolution of the screen was 1152 pixels wide by 900 pixels high. The size of the screen was 38.6 cm wide by 30

cm high. For the viewing distances and stimulus size used, this screen resolution provided available spatial-frequency bandwidths of 0.87–28 cycles per degree (c/deg), for a viewing distance of 1 m, and 1.3–42 c/deg, for a viewing distance of 1.5 m. Subjects used the workstation's mouse to give responses.

### Definitions

The rms contrast of an ensemble is given by

$$C = \left[ \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(f_x, f_y) df_x df_y - \mu^2}{\mu^2} \right]^{1/2}, \quad (4)$$

where  $P(f_x, f_y)$  is the power spectrum and  $\mu$  is the mean for the ensemble. If the ensemble is spatially isotropic, then the power spectrum can be represented as a function of radial spatial frequency, and the rms contrast is given by

$$C = \left\{ \frac{2\pi \left[ \int_0^{\infty} f_r P(f_r) df_r - \mu^2 \right]}{\mu^2} \right\}^{1/2}, \quad (5)$$

where  $f_r$  is the radial spatial frequency. The sample rms contrast of an image in the ensemble will vary somewhat

around  $C$ . It is given by the same equation, after replacing  $P(f_r)$  with the sample power spectrum  $\hat{P}_i(f_r)$ . The use of the term rms contrast to refer to an ensemble average differs from its use in most psychophysics literature, in which it is generally used to refer to the contrast of a particular image. The definition used here is necessary, however, to maintain consistency with other statistical terminology and to distinguish between what is in one case an ensemble average and in another a sample statistic.

### Stimuli

Ideally we would have used images of large spatial extent and broad spatial-frequency bandwidth as stimuli; however, computational limits on the generation and storage of large images constrained us to use images that were  $64 \times 64$  pixels. The spatial-frequency bandwidth of these images was 5 octaves, and the need to use a band within the visible limits of human vision limited the possible size of the images in degrees of visual angle. As an example,  $64 \times 64$  pixel images, in order to span a frequency band from 1 to 32 c/deg, would have to subtend a visual angle of 1 deg. The limit on stimulus size raises the possibility that our results would depend primarily on the specific stimulus conditions used (size and spatial-frequency band of images). To control for this problem, we collected discrimination thresholds for two different viewing distances. We also collected thresholds for spectral drop-offs outside the range of those defining fractal ensembles ( $2.0 < \beta < 4.0$ ) in the hope that the extra data would illuminate condition-dependent effects.

The following procedure was used to generate stimulus images. Random  $64 \times 64$  arrays of Fourier coefficients were created by independently sampling from a normal distribution at each frequency, resulting in samples of white noise with unit variance. The coefficients were then scaled by the appropriate factor at each frequency to generate samples from an ensemble with a specified spectral drop-off. The inverse Fourier transform of these arrays was computed by using a fast Fourier-transform routine to generate  $64 \times 64$  pixel sample images<sup>33</sup>. The resulting images were then scaled to give them a rms contrast of 35% (recall that this is the average contrast of images in the ensemble). The mean of the images was fixed at 127 (in gray-scale units). A rms contrast of 35% guaranteed that fewer than 0.5% of the pixels would underflow or overflow (have values less than 0 or greater than 255) and need to be thresholded. Images generated in this way can be considered to be samples drawn from a Gaussian ensemble with a particular spectral drop-offs ( $1/f_r^\beta$ ) and a specific rms contrast. The plot of the power spectra of two such images (Fig. 1) illustrates the variability in the spectra of the stimuli.

The stimuli were displayed on the center of the screen with the background luminance set to 127, the mean luminance of the stimuli.

### Procedure

The experiment consisted of a two-alternative forced-choice task in which subjects were asked to tell which of two randomly generated images came from an ensemble with a lower spectral drop-off. We used a modified staircase procedure, QUEST,<sup>34</sup> to estimate the threshold difference in spectral drop-off needed for subjects to discriminate images correctly 75% of the time. Thresholds were determined for 10

different reference drop-offs between 0.8 and 4.4 (0.8, 1.2, 1.6, 2.0, 2.4, 2.8, 3.2, 3.6, 4.0, and 4.4). Spectral drop-offs between 2.0 and 4.0 correspond to ensembles of fractional Brownian noise and had fractal dimensions between 3.0 and 2.0, respectively. We collected discrimination thresholds for three conditions, which we will call (1) high-contrast, near, (2) low-contrast, near, and (3) high-contrast, far. In the high-contrast, near, condition the rms contrast of images was 35%, and the viewing distance was 1 m. At this viewing distance, the stimulus images spanned a 5-octave spatial-frequency band between 0.87 and 28 c/deg. In the low-contrast, near, condition the viewing distance remained 1 m, but the rms contrast of the images was decreased by 1/2, to 17.5%. In the high-contrast, far, condition the rms contrast of the images was 35%, and the viewing distance increased to 1.5 m, at which images spanned a spatial-frequency band between 1.3 and 42 c/deg.

Experimental trials were organized into twelve sets of ten 50-trial blocks, one block for each reference spectral drop-off. Each block consisted of running QUEST for 50 trials to estimate the discrimination threshold for a given reference drop-off. The resulting data consisted of the raw responses and 12 estimates of threshold for each reference drop-off. The increments in spectral drop-off between reference and test images varied in decibel steps from -10 to 10 dB, with a 0-dB increment of 0.2 (1 dB corresponds to a 12% increase). The increments varied from a minimum of 0.0644 (-10 dB) to 0.62 (+10 dB). Note that the increments, not the actual test drop-offs, varied in decibel steps. Each experimental trial consisted of a sequential presentation of an image that was randomly drawn from an ensemble with the reference spectral drop-off and an image that was drawn from an ensemble with the test drop-off selected by the QUEST procedure. The order of presentation was random. The stimuli were presented for 250 ( $\pm 50$ ) msec, with an interstimulus interval of 500 ( $\pm 50$ ) msec. The interstimulus interval was chosen to make trials reasonably fast while avoiding confounding masking effects. Subjects were asked to press one of two buttons on the mouse to indicate whether the first or second stimulus presented was drawn from an ensemble with lower spectral drop-off. Stimulus presentation and interstimulus interval times varied owing to the fact that the Sun 3/160 was operating as a multiuser workstation and was servicing the file system for a local area network. To control for this variation, the program controlling the experiment timed each presentation and discarded trials for which either of the presentation times differed from 250 msec by more than 50 msec.

Subjects were given visual feedback in the form of a message indicating the correctness of their response. Though sometimes more were necessary, a minimum of four sets of trials were run to train the subjects on the task. The experiment proper was begun when the within-set average threshold for a subject appeared to level off between sets of trials.

The images used in the experiments were generated before running a set of trials, as the time that it took to generate the images made it unfeasible to do so in real time. For each block of trials, 50 random images were generated from an ensemble with the appropriate reference spectral drop-off, and 10 images were generated from each of the ensembles defined by the 21 test drop-offs. Each block consisted of exactly 50 trials, which guaranteed that no reference image

would be presented twice. It was possible, however, that some test images were presented more than once because the QUEST procedure occasionally tested at 1 test drop-off more than 10 times. Such repeats occurred only rarely and were separated by a large enough number of trials, on average, that we expect the effects of repeated presentations to be negligible.

### Ideal Observer

We consider the ideal observer for this task to be the maximum-likelihood estimator of the spectral drop-off of an image. The ideal observer would calculate estimates of the spectral drop-offs of the two images presented in a trial and select the image for which its estimate was the lowest. Appendix B gives details of the derivation of the maximum-likelihood estimator. We estimated the ideal observer's 75% correct thresholds for each of the 10 reference spectral drop-offs by running simulations in which the ideal observer acted as subject for a QUEST procedure that ran 400 trials. Novel random stimuli were generated for each of the trials.

### Data Analysis

We estimated both the ideal observer's and human subjects' thresholds by using the *a posteriori* mean of threshold, conditional on the data. The posterior distribution is given by

$$p(T_i|x_i) = \frac{p(x_i|T_i)p(T_i)}{p(x_i)}, \quad (6)$$

where  $T_i$  represents the threshold for a reference spectral drop-off  $i$  and  $x_i$  represents the raw data collected over all trials for the reference spectral drop-off in one condition. We modeled the forward distribution  $p(x_i|T_i)$  as the Weibull function,<sup>34</sup> the same used in the QUEST procedure for placing experimental trials. We assume no prior bias on the threshold, so the prior distribution of  $T_i$  is uniform, and  $p(T_i)$  is a constant for all  $T_i$ . For a given set of raw data  $x_i$ ,  $p(x_i)$  is a constant, so Eq. (6) simplifies to

$$p(T_i|x_i) = K * p(x_i|T_i), \quad (7)$$

where the constant  $K$  is  $p(T_i)/p(x_i)$ .  $K$  can be calculated by selecting the constant that makes the integral of the posterior probability distribution equal to one:

$$K = \frac{1.0}{\int_0^\infty p(x_i|T_i)dT_i}. \quad (8)$$

The *a posteriori* mean is the mean of the distribution given by Eq. (7), and the standard error of the estimate is the standard deviation of the distribution.

A full characterization of  $p(x_i|T_i)$  requires knowledge of the slope of the Weibull function. We compute maximum-likelihood estimates of the slope  $S_i$  and threshold  $T_i$  from the distribution  $p(x_i|T_i, S_i)$  and use the estimated slope in the computation of  $p(T_i|x_i)$  for the estimation of the *a posteriori* mean threshold.

The *a posteriori* mean differs from the maximum-likelihood estimate, which selects the value  $T_i = \hat{T}_i$  that maximizes  $p(x_i|T_i)$ . In the case presented here, in which no prior constraints are placed on the threshold, maximum-likelihood estimation would correspond to selecting the mode of  $p(T_i|x_i)$ , so-called maximum *a posteriori* estimation. The

maximum-likelihood estimates of thresholds computed from the data in these experiments turn out to be slightly lower than the *a posteriori* means.

### Subjects

There were three observers. DCK and DF are authors, and DR was naïve to the details of the experiment. DCK and DR are emmetropes, and DF is a myope with 20/20 corrected vision. Viewing was binocular with fixation on the center of the screen.

### RESULTS

The results of running simulations with the ideal observer (Table 1) show that its threshold function is statistically flat for the reference spectral drop-offs used in the experiment. The average threshold across the 10 spectral drop-offs for the ideal observer was  $-12.3 \pm 0.49$  dB ( $\Delta\beta = 0.05$ ). Because of the flatness of the ideal-threshold function, the efficiency curves for human subjects have the same shape as the absolute-discrimination threshold curves. Since the shape of the curves is our primary interest, we will show the absolute-threshold curves instead of efficiency curves and comment on calculated efficiency when appropriate.

Figure 3 shows the estimates of discrimination threshold as a function of spectral drop-off for the three subjects in each of the three conditions. The thresholds are given in decibel increments. The axes above each graph delimit the range of drop-offs that define fractal ensembles (2.0–4.0). The threshold curves have two general features: a broad dip in the fractal range and, in some conditions, another dip for the smallest spectral drop-offs. DCK shows the second dip in both of the high-contrast conditions, DF shows it in both the low-contrast, near, and the high-contrast, far, conditions, and DR shows it in all conditions. For all subjects, the dip at low spectral drop-offs was largest in the high-contrast, far, condition.

The discrimination task was designed to require the integration of information at different spatial frequencies; however, a close look at the task shows that it might be performed with a simple strategy that uses the sample rms contrast in only one frequency band. Figure 1 shows that the power spectra of ensembles with different drop-offs, but equal rms contrast, intersect at some frequency  $f_c$ . This fact suggests a strategy for the discrimination task consisting of

**Table 1. Ideal Observer's Thresholds for the Ten Reference Spectral Drop-offs**

Reference Spectral Drop-off	Threshold (dB)
0.8	-12.6
1.2	-11.6
1.6	-11.9
2.0	-11.9
2.4	-12.4
2.8	-12.4
3.2	-12.2
3.6	-12.4
4.0	-13.4
4.4	-12.4

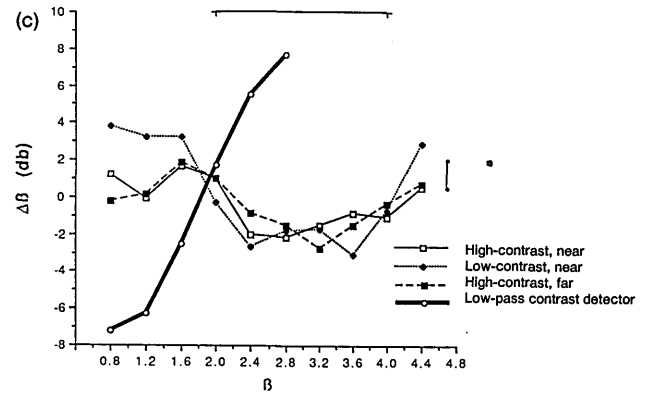
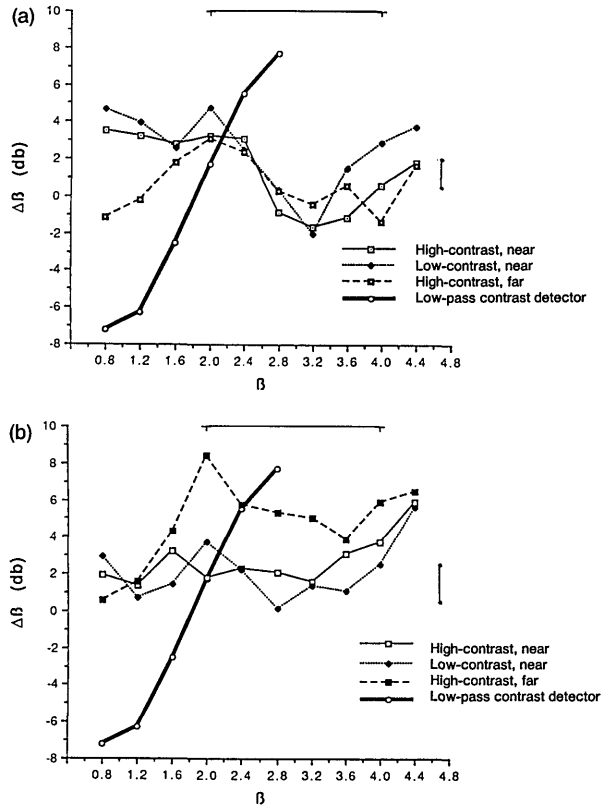


Fig. 3. Discrimination threshold curves for subjects (a) DCK, (b) DR, and (c) DF. The thresholds are plotted as decibel increments versus reference spectral drop-off. A plot of the threshold function for the low-pass rms contrast model is shown as a bold line on each graph. The average standard error is shown on the right-hand side of each graph. An increment of approximately  $-3$  dB is the lowest measured threshold. This value corresponds to a real increment of 0.14.

rms contrast discrimination in either a high-pass (greater than  $f_c$ ) or a low-pass (less than  $f_c$ ) frequency band. A subject looking at the stimuli through a low-pass filter would perform the task by selecting the image with lower sample rms contrast, while one looking through a high-pass filter would select the image with higher sample rms contrast. A comparison of the sample spectra of the images in Fig. 1 indicates that this strategy could succeed for the discrimination task, at least for images drawn from ensembles differing in spectral drop-off by a large amount.

We ran simulations for models that used the contrast-discrimination strategy in both the low-pass and high-pass bands. The models were assumed to have knowledge of the intersection frequency  $f_c$  for each spectral drop-off. The low-pass model used a decision variable given by the ratio of sample rms contrasts in the low-pass bands of the two images:

$$D = \frac{\hat{C}_1}{\hat{C}_2} = \left\{ \frac{\mu_2^2 \left[ \int_0^{f_c} f_r \hat{P}_1(f_r) df_r - \mu_1^2 \right]}{\mu_1^2 \left[ \int_0^{f_c} f_r \hat{P}_2(f_r) df_r - \mu_2^2 \right]} \right\}^{1/2}. \quad (9)$$

The model selects Image 1 if  $D < 1$  and Image 2 otherwise. The high-pass model uses a similar decision variable based on the relative contrasts in the high-pass bands:

$$D = \frac{\hat{C}_1}{\hat{C}_2} = \left\{ \frac{\mu_2^2 \left[ \int_{f_c}^{\infty} f_r \hat{P}_1(f_r) df_r - \mu_1^2 \right]}{\mu_1^2 \left[ \int_{f_c}^{\infty} f_r \hat{P}_2(f_r) df_r - \mu_2^2 \right]} \right\}^{1/2}. \quad (10)$$

The model selects Image 1 if  $D > 1$  and Image 2 otherwise.

The discrimination thresholds of both models drop sharply as the spectral drop-off of the images decreases. The threshold curve for the low-pass model is included in the graphs in Fig. 3 for comparison with subjects' curves. The improvement in subject performance at low drop-offs mirrors that of the low-pass model. Informal oral reports of the subjects suggest the use of a low-pass contrast-discrimination strategy for images with low spectral drop-offs. All three subjects reported selecting images with low perceived contrast when the images had a low drop-off.

The dip in performance for ensembles with high spectral drop-offs, including those specifying fractal ensembles, retains approximately the same shape between conditions. The primary difference is that the dip is more pronounced in the low-contrast and far-viewing conditions. In the far-viewing condition, DCK does show an improvement in performance at a drop-off of 4.0, and DR shows a general decrease in performance for all spectral drop-offs.

The size of the standard errors for the threshold estimates makes the differences in the middle region of the dip insignificant; however, because the shape of the curves is so similar for high spectral drop-offs, we average these threshold estimates across conditions to get a final performance curve. The averaged thresholds are shown with standard-error bars in Fig. 4. Performance is clearly best for drop-offs between 2.8 and 3.6.

Simulations indicate that the ideal observer's absolute threshold for this task is inversely proportional to the square root of the sample size (Table 2), so we define the appropriate measure of efficiency as the squared ratio of the ideal observer's threshold ( $\Delta\beta = 0.05$ ) and the threshold of a subject as



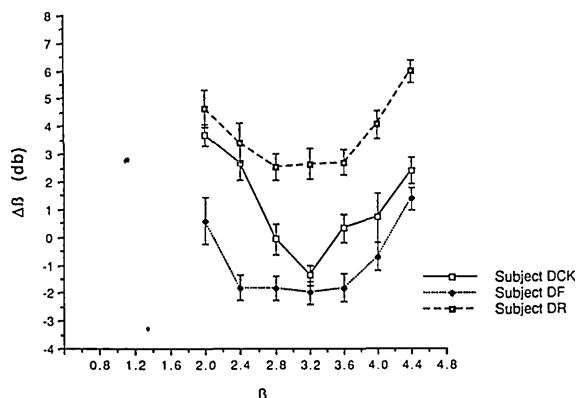


Fig. 4. Discrimination threshold curves with standard-error bars for each subject averaged across the three conditions. Only thresholds for spectral drop-offs that are greater than or equal to 2.0 are shown.

**Table 2. Ideal Observer's Thresholds for Four Different Image Sizes Using  $\beta = 3.2$  as the Reference Spectral Drop-off**

Image Size ( $N_{\text{pixels}}$ )	Threshold (dB)
$16 \times 16$ ( $N = 256$ )	0.17 ( $\Delta\beta = 0.204$ )
$32 \times 32$ ( $N = 1024$ )	-5.6 ( $\Delta\beta = 0.106$ )
$64 \times 64$ ( $N = 4096$ )	-12.2 ( $\Delta\beta = 0.50$ )
$128 \times 128$ ( $N = 16384$ )	-18.7 ( $\Delta\beta = 0.024$ )

$$E = \frac{\Delta\beta_i^2}{\Delta\beta_s^2} \quad (11)$$

DF had a peak efficiency of 13%, DCK had a peak efficiency of 10%, and DR had a peak efficiency of 7%.

## DISCUSSION

The results presented here differ from those of Rensink,<sup>24</sup> which showed peak performance for a spectral drop-off of 4.0. The discrepancy may simply result from differences in discriminability of line textures and two-dimensional textures. Another possible cause for the difference, however, is the difference in the sizes of the stimuli used. Rensink used images that subtended 2 deg by 4 deg of visual angle, whereas the stimuli used here subtended, in the near conditions, 1.15 deg by 1.15 deg of visual angle; therefore most of the rms contrast in Rensink's stimuli was concentrated in a frequency band that was below that tested in this experiment.

The discrimination experiment was designed to investigate the question posed in the introduction: What is the correlational structure of the ensemble of images to which the visual system is optimally tuned? In order to permit any conclusions about the answer to this question to be drawn from subjects' threshold curves, the task must meet several criteria. Subjects' performance on the task should reflect constraints on performance owing to the amount of information in the stimuli reliably transmitted through the system and not constraints on the subjects' ability to use the information in performing the discrimination task. In light of this problem, we should note that the task did seem unnatural to subjects. Humans, on the other hand, are

remarkably adept at learning a wide range of visual tasks, and we do not believe that the particular requirements of the task embody the significant factor in the results.

The second criterion is that subjects' performance should reflect a strategy that uses a broad band of information available in the stimuli.<sup>36</sup> We showed that subjects' performance for ensembles with spectral drop-offs less than 2.0 could be well predicted by a model that used the information in either only a low-pass or only a high-pass frequency band of the stimuli. Specifically, it performed the task based on rms contrast in one of the two bands. For ensembles with spectral drop-offs greater than 2.0, however, the performance of subjects was better than that of this simple model. To have achieved this level of performance, subjects had to have based their decisions on the relative power in different frequency bands of the two images presented in a trial. Results from these conditions may be a more accurate reflection of information transmission limits in the visual system. This is fortunate, as we are particularly interested in fractal ensembles, which have spectral drop-offs between 2.0 and 4.0. The remainder of the discussion will deal with the results for these ensembles.

In the Introduction, we presented two hypotheses about the statistical structure of natural images. Let us assume for a moment that the visual system is optimized to transmit the information in natural images. According to Field's hypothesis, we would expect performance on the discrimination task to be best for images from ensembles with a spectral drop-off of 2.0. A modified version of the hypothesis of Srinivasan (in which the space constant in the exponential is permitted to go to zero) would suggest that the best performance would be for images from ensembles with a spectral drop-off of 3.0. The data from this experiment are most consistent with the modified hypothesis of Srinivasan *et al.*, since discrimination performance was best in a range between 2.8 and 3.6.

A spectral drop-off of 3.0 has a special theoretical significance for Gaussian ensembles, namely, that they are Markov. A Markov ensemble of images is one in which the statistical dependencies between the light intensity at one point in an image and the intensities at all other points is fully characterized by dependencies with the point's nearest neighbors. In fact, fractional Brownian motion with a drop-off of 3.0 is a generalization of a random walk to two dimensions.<sup>16</sup> A Gauss-Markov ensemble that is stationary has an exponentially decaying autocorrelation function,<sup>37</sup> as in the model of Srinivasan *et al.* As noted in the Introduction, the power spectrum of such a process has a spectral drop-off of 3.0 for frequencies higher than that defined by the autocorrelation function's space constant.<sup>19</sup> The results may, therefore, reflect the fact that we are tuned to ensembles of images that are Markov. In a study related to this one, Kersten<sup>38</sup> had subjects predict the gray level of pixels deleted from natural images. He found that performance was well modeled by a number of different nearest-neighbor predictors. More significantly, subjects' performance increased little when information beyond the eight nearest neighbors was provided.

We used fractional Brownian motion for the generation of stimuli in the experiment in part because it is Gaussian and thus wholly defined by either its autocorrelation function or power spectrum. A look at the images in Fig. 2 shows that



fractional Brownian motion is clearly too unstructured to be a complete model of natural images. The images appear as noise and are easily distinguishable from real examples of natural images. This result is due to the existence of higher-order statistical structure in natural images. The prevalence of coherent edges in natural images is one example of this structure. An ensemble that shows power-law behavior in its power spectrum need not even be fractal, if it has higher-order structure. An ensemble of images of randomly positioned circular spots would have a power spectrum with a  $1/f_r^2$  envelope but would certainly not be considered fractal, as it would not exhibit self-similarity across scales. The argument that the human visual system is tuned to ensembles whose spectra exhibit power-law behavior does not, therefore, imply that it is tuned to fractal ensembles.

The discussion above leads to a cautionary note on drawing conclusions that are too general from the results of this study. Though the visual system may be better tuned to ensembles of fractional Brownian noise with a spectral drop-off between 2.8 and 3.6 relative to other ensembles of fractional Brownian noise, one may find that the tuning shifts with the addition of higher-order structure. One way to test this possibility would be to do the same experiments reported here with images that have the phase spectra of some real natural images. Of course, a complete characterization of the statistical structure to which the visual system is tuned would include higher-order structure.

## APPENDIX A

Fractal functions are characterized by a self-similarity over different scales. For deterministic functions, this definition means that a piece of the function  $1/n$ th the size of the original, when blown up by a factor  $n$ , is congruent to the original function (similar up to a rotation or translation of the function). Such a function may be characterized by its similarity dimension,<sup>17</sup> which specifies how the number of self-similar pieces in the function increases as the scale is decreased. Mandelbrot defines the fractal dimension of a function to be equivalent to its similarity dimension.<sup>16</sup>

The similarity dimension of a self-similar function is given by the relation

$$N = n^D, \quad (12)$$

where  $N$  is the total number of self-similar pieces  $1/n$ th the size of the original. From Eq. (12), we see that  $D$  can be computed as

$$D = \frac{\log N}{\log n}. \quad (13)$$

For fractal functions,  $D$  is bounded below by the topological dimension of the function (a curve has a topological dimension of 1; a surface has a dimension of 2, etc.). The fractal dimension, in some sense, corresponds to the roughness of the function. Graphs of functions characterized by high fractal dimensions appear rougher than those with low fractal dimensions.

For self-similar functions, the similarity dimension corresponds to other definitions of fractal dimension (e.g., the Hausdorff-Besicovitch or Minkowski-Bouligand dimensions). These definitions characterize a function's behavior not at all scales but in the limit at infinitely small scales and

are more generally applicable to functions that are not self-similar.

In order to apply these ideas to modeling natural phenomena, Mandelbrot extended them to stochastic ensembles of functions.<sup>16</sup> In this case, self-similarity across scales within an ensemble refers to similarity of the statistics of the functions at different scales; that is, the statistics of a piece  $1/n$ th the size of the original function, after scaling by a factor of  $n^H$  ( $0 < H < 1$ ), are in all respects the same as those of the whole function (technically, these functions are termed self-affine, as the scaling is nonuniform along different dimensions). Fractal models have been successfully applied to a wide range of natural phenomena, including the shape of coastlines, rivers, and surface landscapes,<sup>16</sup> the structure of music,<sup>26</sup> and dynamical systems.<sup>38</sup>

The particular fractal model used in this paper is fractional Brownian motion. It is defined by the following relation:

$$p\left[\frac{I(x + \Delta x) - I(x)}{|\Delta x|^H} < y\right] = f(y), \quad (14)$$

where  $p(\cdot)$  is a cumulative probability distribution function. The random fractal function  $I$  and the index of spatial location  $x$  may be vector valued. The exponent  $H$  is constrained to have a value between 0 and 1. The fractal dimension of the process  $I$  is

$$D = E + 1 - H, \quad (15)$$

where  $E$  is the Euclidean dimension of the function. Relation (14) implies that the increments of  $I$  over different distances  $\Delta x$ , when scaled by a factor  $|\Delta x|^H$ , follow the same probability law. The statistics of  $I$  are therefore said to be invariant over changes in scale.

The spectral drop-off describing the power spectrum of a one-dimensional slice of a fractional Brownian motion ensemble is related to  $E$  and  $H$  by

$$\beta = 1 + 2H. \quad (16)$$

For a radially symmetric ensemble, the power spectrum is fully characterized as a function of radial spatial frequency, in terms of which the spectral drop-off is given by<sup>26</sup>

$$\beta = 1 + 2H + (E - 1) = E + 2H. \quad (17)$$

With appropriate substitutions, we obtain the equation presented in the text for the spectral drop-off of a radially symmetric ensemble with  $E = 2$ :

$$\beta = 8 - 2D. \quad (18)$$

## APPENDIX B

The ideal observer for the discrimination task must compute the maximum-likelihood estimate of spectral drop-off for the two images being compared. In order to derive this estimator, we need to know the distribution of the sample power spectra of images. The discrete Fourier transform of images from a stochastic ensemble may be represented as an array of random complex-valued coefficients, indexed by frequency

$$\mathcal{F}(f_x, f_y) = X_r(f_x, f_y) + iX_i(f_x, f_y). \quad (19)$$

For ensembles that are Gaussian in the space domain,  $X_r(f_x,$

$f_y$ ) and  $X_i(f_x, f_y)$  are independent Gaussian random variables whose variance is given by  $1/2$  the power at the frequency  $(f_x, f_y)$ :

$$\sigma(f_x, f_y)^2 = E[X_r(f_x, f_y)^2] = E[X_i(f_x, f_y)^2] = \frac{1}{2}P(f_x, f_y). \quad (20)$$

The sample power spectrum of an image is given by

$$P_s(f_x, f_y) = X_r(f_x, f_y)^2 + X_i(f_x, f_y)^2 \quad (21)$$

and is itself a random process. When  $X_r(f_x, f_y)$  and  $X_i(f_x, f_y)$  are Gaussian,  $P_s(f_x, f_y)$  has an exponential distribution given by

$$f[P_s(f_x, f_y)] = \frac{1}{2\sigma(f_x, f_y)^2} \exp\left[-\frac{P_s(f_x, f_y)}{2\sigma(f_x, f_y)^2}\right]. \quad (22)$$

The power spectrum of the ensemble is the mean of the sample power spectra of images in the ensemble and is given by

$$P(f_x, f_y) = E[P_s(f_x, f_y)] = 2\sigma(f_x, f_y)^2, \quad (23)$$

which confirms the relation given in Eq. (20). Because the ensembles that we consider are isotropic, we will replace the two-dimensional frequency pair  $(f_x, f_y)$  with radial frequency  $f_r = (f_x^2 + f_y^2)^{1/2}$ . Sums and products, when given, are taken over all  $(f_x, f_y)$  pairs. This will simplify notation somewhat.

For ensembles of fractional Brownian motion, the power spectrum is given by the inverse power law

$$P(f_r) = \frac{k}{f_r^\beta}. \quad (24)$$

Using Eq. (23) and substituting into Eq. (22), we obtain

$$p[P_s(f_r)|\beta, k] = k^{-1}f_r^\beta \exp[-P_s(f_r)k^{-1}f_r^\beta]. \quad (25)$$

The likelihood function for a given sample power spectrum is given by

$$p(P_s|\beta, k) = \prod k^{-1}f_r^\beta \exp[-P_s(f_r)k^{-1}f_r^\beta], \quad (26)$$

since the  $P_s(f_r)$  are independent random variables. The log likelihood function is given by

$$L(P_s|\beta, k) = -\sum \log(k) + \beta \sum \log(f_r) - k^{-1} \sum f_r^\beta P_s(f_r). \quad (27)$$

If the number of Fourier coefficients in the sample is  $N$ , Eq. (27) simplifies to

$$L(P_s|\beta, k) = -N \log(k) + \beta \sum \log(f_r) - k^{-1} \sum f_r^\beta P_s(f_r). \quad (28)$$

We want to find maximum-likelihood estimates of  $k$  and  $\beta$  given by the maximum of Eq. (28). Differentiating Eq. (28) with respect to  $k$  and setting  $\partial L/\partial k$  equal to zero, we obtain for  $k$

$$k = \frac{1}{N} \sum f_r^\beta P_s(f_r). \quad (29)$$

Differentiating Eq. (28) with respect to  $\beta$ , setting  $\partial L/\partial \beta$  equal to zero, and substituting Eq. (29) for  $k$ , we obtain

$$\sum \log(f_r) = \frac{N \sum f_r^\beta \log(f_r) P_s(f_r)}{\sum f_r^\beta P_s(f_r)}. \quad (30)$$

A numerical solution of Eq. (30) gives the maximum-likelihood estimate of  $\beta$ .

An alternative approach to estimating the spectral drop-off is to use a linear-regression method proposed by Pentland.<sup>18</sup> The log power of fractional Brownian noise is linearly related to the log radial frequency with the slope given by  $\beta$ :

$$\log[P(f_r)] = -\log(k) - \beta \log(f_r). \quad (31)$$

He suggests estimating  $\beta$  by doing a linear regression on the log power versus log frequency. This estimate is not equivalent to the maximum-likelihood estimate; however, it is much faster and provides only a slightly worse estimate. (The variance of the estimate is slightly greater than that of the maximum-likelihood estimate.) In the simulations of the ideal observer presented here, we used the maximum-likelihood estimator given by Eq. (30).

## ACKNOWLEDGMENTS

This research was supported by National Science Foundation (NSF) grant BNS-8708532 to Daniel Kersten, NSF grant BNS-85-18675 to James Anderson, and Science and Engineering Research Council grant GR/E 43003 to Horace Barlow. A part of the research was done while David Knill and David Field were at the Physiological Laboratory, Cambridge University, and Daniel Kersten was at the Department of Psychology, Brown University. The authors thank Horace Barlow and Jim Anderson for their support, and they thank David Tolhurst, Yoav Tadmor, and Peter Foldiak for the insights that they shared in many discussions on theoretical and methodological issues related to this research. The research of D. Field was supported by a grant from the Wellcome Trust.

\* Permanent address, Department of Psychology, Uris Hall, Cornell University, Ithaca, New York 14853.

† Permanent address, Department of Psychology, University of Minnesota, Minneapolis, Minnesota 55455.

## REFERENCES AND NOTES

1. F. Attneave, "Informational aspects of visual perception," *Psychol. Rev.* **61**, 183-193 (1954).
2. H. B. Barlow, "Sensory mechanisms, the reduction of redundancy and intelligence," *NPL Symposium on the Mechanization of Thought Processes*, No. 10 (H. M. Stationary Office, London, 1959), pp. 535-539.
3. H. B. Barlow, "The coding of sensory messages," in *Current Problems in Animal Behavior*, W. H. Thorpe and O. L. Zangwill, eds. (Cambridge U. Press, Cambridge, 1961), pp. 331-360.
4. H. B. Barlow, "The Ferrier lecture: critical limiting factors in the design of the eye and visual cortex," *Proc. R. Soc. London Ser. B* **212**, 1-34 (1981).
5. M. V. Srinivasan, S. B. Laughlin, and A. Dubs, "Predictive

- coding: a fresh view of inhibition in the retina," *Proc. R. Soc. London Ser. B* **216**, 427–459 (1982).
6. T. Bossomaier and A. W. Synder, "Why spatial frequency processing in the visual cortex?" *Vision Res.* **26**, 1307–1309 (1986).
  7. D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Am. A* **4**, 2379–2394 (1987).
  8. R. Linsker, "Self-organization in a perceptual network," *IEEE Trans. Comput.* **21**, 105–117 (1988).
  9. D. M. Kammen and A. L. Yuille, "Spontaneous symmetry-breaking energy functions and the emergence of orientation selective cortical cells," *Biol. Cybern.* **59**, 23–31 (1988).
  10. H. B. Barlow and P. Z. Foldiak, "Adaptation and decorrelation in the cortex," in *The Computing Neuron*, R. C. Miall, R. M. Durbin, G. J. Mitchison, eds. (Addison-Wesley, Reading, Mass., 1989).
  11. C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (U. Illinois Press, Champaign, Ill., 1949).
  12. Stationarity implies that the statistics of a random field are invariant over translations of the coordinate system on which it is defined. Isotropy implies that they are also invariant over rotations of the coordinate system. One result of these two assumptions is that the autocorrelation function can be expressed as a function of Euclidean distance. The assumption of stationarity is intuitively attractive, as it can result from viewing scenes from a range of positions, so that shifted versions of any given image are equally likely. A similar argument, however, cannot be made for the isotropy assumption, as we generally view scenes with our heads perpendicular to the ground. In a study related to the question of isotropy, Switkes *et al.*<sup>13</sup> found more power at horizontal and vertical orientations in images of both natural and synthetic scenes. The assumption does, however, simplify our investigation by allowing us to look at the correlational structure of images as a function only of distance between points.
  13. E. Switkes, M. J. Mayer, and J. A. Sloan, "Spatial frequency analysis of the visual environment: anisotropy and the carpentered environment hypothesis," *Vision Res.* **18**, 1393–1399 (1978).
  14. The autocorrelation function of a random field **I** is given by
 
$$R_I((x_i, y_i), (x_j, y_j)) = E[I(x_i, y_i)I(x_j, y_j)],$$
 where  $E[\cdot]$  is the expectation operator. Since we are assuming that **I** is stationary and isotropic, we can write this as a function of the Euclidean distance between points:
 
$$R_I(\Delta r) = E[I(x, y)I(x + \Delta r \cos \theta, y + \Delta r \sin \theta)],$$
 where  $r = (\Delta x^2 + \Delta y^2)^{1/2}$  and  $\theta$  is the angle between the points. The second-order statistical structure of a stationary ensemble of images, given by the autocorrelation function in space, is given by the power spectrum in the frequency domain. The power spectrum is the Fourier transform of the autocorrelation function
 
$$P_I(f_x, f_y) = \int_0^\infty \int_0^\infty R_I(x, y) \exp[i2\pi(fx + fy)] dx dy.$$
 The power at a given frequency is twice the variance of the corresponding Fourier coefficients (real and imaginary) of images in an ensemble. The real and imaginary parts of the Fourier coefficient at a given frequency are uncorrelated and have equal variance. The power at  $f_x = f_y = 0$  is the squared mean of the ensemble. For an isotropic ensemble, the power spectrum may be written as a function of radial spatial frequency,  $P_I(f_r)$ , where  $f_r = (f_x^2 + f_y^2)^{1/2}$ .
  15. B. Julesz, "Spatial frequency channels in one-, two- and three-dimensional vision: variations on an auditory theme by Beke-  
sy," in *Vision Coding and Adaptability*, C. S. Harris, ed. (Erlbaum, Hillsdale, N.J., 1980).
  16. B. B. Mandelbrot, *Fractals: Form, Chance, and Dimension* (Freeman, San Francisco, Calif., 1977).
  17. B. B. Mandelbrot, *The Fractal Geometry of Nature* (Freeman, San Francisco, Calif., 1982).
  18. A. P. Pentland, "Fractal-based description of natural scenes," *IEEE Trans. Pattern Anal. Mech. Intell. PAMI-6*, 661–673 (1984).
  19. The power spectrum of an ensemble with an exponential autocorrelation function clearly shows the effect of the scale constant  $k$ . For an isotropic ensemble, the spectrum is given by
 
$$P_I(f_r) \propto \frac{1}{(k + f_r^{3/2})^2}.$$
 At frequencies much lower than  $k$ , the spectrum approximates white noise; that is, points in the image separated by a distance much greater than  $1/k$  are effectively uncorrelated. For frequencies much greater than  $k$ , the spectrum falls off according to the power law  $1/f_r^3$ . The images of this ensemble exhibit qualitatively different statistical behavior at different scales.
  20. B. Julesz, "Visual pattern discrimination," *IRE Trans. Inf. Theory IT-8*, 84–92 (1962).
  21. B. Julesz, E. Gilbert, L. Shepp, and H. Frisch, "Inability of humans to discriminate between visual textures that agree in second-order statistics revisited," *Perception* **2**, 391–405 (1973).
  22. B. Julesz and J. Bergen, "Textons, the fundamental elements in preattentive vision and perception of texture," *Bell Syst. Tech. J.* **62**, 619–1645 (1983).
  23. W. K. Pratt, O. D. Faugeras, and A. Gagalowicz, "Visual discrimination of stochastic texture fields," *IEEE Trans. Syst. Man Cybern. SMC-8*, 796–804 (1978).
  24. R. A. Rensink, *On the Visual Discrimination of Self-Similar Random Textures*, Department of Computer Science Tech. Rep. 86-16 (University of British Columbia, Vancouver, B.C., Canada, 1986).
  25. Rensink generated line textures by using one-dimensional power spectra. Peak performance was found to be at a spectral drop-off of  $\beta_{1D} = 3$  for these textures. The equivalent two-dimensional spectral drop-off is given by  $\beta_{2D} = \beta_{1D} + 1 = 4$ .<sup>26</sup>
  26. R. F. Voss, "Random fractal forgeries," in *Fundamental Algorithms for Computer Science*, R. A. Earnshaw, ed. (Springer-Verlag, Berlin, 1985), pp. 805–829.
  27. H. B. Barlow, "The efficiency of detecting changes of density in random dot patterns," *Vision Res.* **18**, 637–650 (1978).
  28. H. B. Barlow, "Measurements of the quantum efficiency of discrimination in human scotopic vision," *J. Physiol.* **160**, 169–188 (1962).
  29. A. B. Watson, H. B. Barlow, and J. G. Robson, "What does the eye see best?" *Nature (London)* **302**, 419–422 (1983).
  30. A. E. Burgess, R. F. Wagner, R. J. Jennings, and H. B. Barlow, "Efficiency of human visual signal discrimination," *Science* **214**, 93–94 (1981).
  31. D. Kersten, "Spatial summation in visual noise," *Vision Res.* **24**, 1977–1990 (1984).
  32. We corrected for the nonlinearity by raising the entries of the lookup table to an exponent of 0.375 ( $1/2.67$ ) and rescaling them to give a range of 256 gray levels. The equation used for correcting the lookup-table entries was
 
$$l[i] = i^{0.375} * 250.0^{0.625}, \quad 0 \leq i < 256,$$
 where  $l[i]$  is the  $i$ th entry in the lookup table.
  33. The resulting image statistics had toroidal symmetry, reflecting the symmetry of the fast Fourier-transform algorithm.
  34. A. B. Watson and D. G. Pelli, "QUEST: a Bayesian adaptive psychometric method," *Percept. Psychophys.* **33**, 113–120 (1983).
  35. W. A. Weibull, "A statistical distribution function of wide applicability," *J. Appl. Mech.* **18**, 292–297 (1951).
  36. D. Kersten, "Statistical efficiency for the detection of visual noise," *Vision Res.* **24**, 1977–1990 (1984); *Vision Res.* **27**, 1029–1040 (1987).
  37. H. J. Larson and B. O. Shubert, *Probabilistic Models in Engineering Sciences, Vol. 1: Random Variables and Stochastic Processes* (Wiley, New York, 1979).
  38. D. Kersten, "Predictability and redundancy of natural images," *J. Opt. Soc. Am. A* **4**, 2395–2400 (1987).
  39. H. O. Peitgen and P. H. Richter, *The Beauty of Fractals* (Springer-Verlag, Berlin, 1986).