Role of second- and third-order statistics in the discriminability of natural images

Mitchell G. A. Thomson and David H. Foster

Department of Vision Sciences, Aston University, Birmingham B4 7ET, UK

Several studies of natural-image statistics have concluded that the majority of structural image information is represented in the Fourier domain not by the power spectrum, which can be modeled with a power function of spatial frequency, but by the phase spectrum. Psychophysical experiments are described in which human observers were required to discriminate between visual stimuli differing only in their global second-order statistics. The patterns of sensitivity recorded were found to depend on whether these stimuli were natural images or images with random-phase spectra. These results suggest that discrimination performance is related to changes in higher-order image structure; a bandpass model observer based on third-moment statistics was applied to all stimulus images and was found to provide a good fit to the psychophysical data. © 1997 Optical Society of America [S0740-3232(97)01309-4]

1. INTRODUCTION

The discovery of widespread structural redundancy in natural images has motivated the incorporation of statistical image models into computational theories of visual processing. This approach has produced a variety of candidate encoding strategies,[1–3] many of which effectively filter out predictable correlational structure so as to reduce the dimensionality of the visual input. If measures on the second-order structure of natural-image ensembles can be used to determine the basis vectors of a linear transform, image data can be encoded into a new coordinate system whose axes lie closer to the natural axes of the visual input. Ideally, the encoded representation will show only sparse neural activity in response to an expected stimulus; a reduction in image dimensionality can then be achieved at the next processing stage by thresholding. Field[1] assessed the feasibility of such an approach by investigating the two-dimensional (2-D) amplitude spectra (averaged across all orientations) of an ensemble of natural images; he found that the amplitude falls off typically as the inverse of radial spatial frequency ω, that is, the corresponding power spectra $\hat{H}(\omega)$ fall off as ω^{-2} (see Fig. 1). The contrast energy of visual signals with these statistics is independent of the scale of viewing, and if such signals are encoded by a bank of spatial-frequency-selective mechanisms, or channels, whose spatial-frequency bandwidths are constant in octaves, equal energy will be passed through every channel.[4] Some neurophysiological studies have suggested that the spatial-frequency bandwidths of cortical simple cells are constant when measured in octaves[5]; when a typical natural scene is processed in this way, the outputs of the different channels should have similar variances, irrespective of their preferred spatial frequencies. The sparseness of active cells within this neural representation could make further processing very efficient.

Although such a low-dimensional power-spectral model can greatly reduce the redundancy associated with the power spectra of natural images, it may be too simple a description of natural-image structure: Several studies[6–8] have shown that the power-law model for natural-image power spectra is robust only if the exponent is allowed to vary from image to image, that is, if the spectra are fitted by the more general equation $\hat{H}(\omega) \propto \omega^{-\beta}$. In addition, the model has met with limited success when used to predict human performance in image-discrimination tasks; although second-order measures can explain observers' ability to distinguish natural images from a field of random dots,[1] several authors[6,9,10] have reported that the organization of image phase information appears far more critical to visual perception than those image properties measured by the power spectrum. Wavelet analyses have been used to show that natural images contain structure that is aligned locally in phase space,[11] and there may exist phase-selective mechanisms sensitive to certain phase relationships between harmonically related frequencies.[10,12] It is difficult to envisage a simple model for the encoding of global image phase information, but it has been argued that image discriminability should depend quantitatively on higher-order image correlations as well as those represented by the Fourier power spectrum. These observations, together with the success of stochastic models of global second-order image structure,[1,13] motivate an extension of this work to higher-order statistical domains.[14,15]

The extension considered in the present study is restricted to simple third-order statistical measures. Yellott[16] has demonstrated that the third-order statistical histogram of a band-limited image of restricted support determines the image uniquely and completely up to a translation term. Thus, where the term statistics refers to the stochastic properties of an individual image, third-
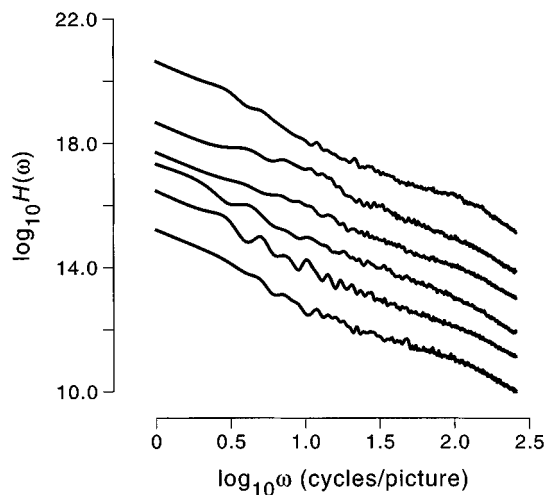
Fig. 1. Power spectra for six digitized images, averaged across all orientations. The spectra have been progressively shifted upward for clarity. The spectra fall off by a factor of roughly $1/\omega^2$; on these double-logarithmic coordinates, this corresponds to a slope of -2.

order statistics must bound (but do not yet solve) the problem of finding a statistical correlate for psychophysical performance in image-discrimination tasks, even though these measures might occupy a lower-dimensional space in the fourth- and higher-order statistical domains. (Notice, however, that where the term statistics is used to describe the generator statistics of an image ensemble, correlations of order 4 and above may also be important; Julesz *et al.*[17] have demonstrated texture discriminability by human observers at the level of fourth-order statistics.)

It is important to emphasize that the global statistical measures used in this study are quite distinct from those used by Julesz and co-workers (e.g., Julesz *et al.*[17]). Klein and Tyler[18] have pointed out that the nth-order statistics definitions of Julesz *et al.* differ from the definitions usually given in textbooks of linear and nonlinear systems theory.[19] Conventionally, the second-order correlation function (SCF) of a zero-mean one-dimensional (1-D) signal is computed by calculating, as a function of a shift τ, the sum of the luminance products of all possible pairs of data points separated by the distance τ; the third-order correlation function (TCF) of a 1-D process is computed similarly but as a function of pairs of shifts (τ_1, τ_2). Accordingly, global third-order image statistics are defined here by the 2-D TCF, global second-order image statistics by the 2-D SCF, and global first-order image statistics by the image mean.

This paper reports a psychophysical experiment designed to reveal the effect of natural phase structure on human observers' ability to detect differences in the second-order statistical structure of natural images. Second-order image statistics were parameterized in terms of the power-law model $\hat{H}(\omega) \propto \omega^{-\beta}$ of image power spectra; the paradigm required observers to discriminate between natural images differing only in their power-spectral falloffs β. The effect of phase structure on performance was determined by conducting the experiment twice, once for images with natural-phase spectra and

once for images whose phase spectra had been fully randomized. Four different model observers were then constructed and compared with the psychophysical data in an attempt to determine the statistical parameterizations of the stimulus images that adequately described human performance.

2. IMAGE STATISTICS

Three libraries of images were acquired for use in this study: a midrange library of 27 images photographed with a SLR camera having a 50-mm lens; a close-up library of 25 images photographed with a SLR camera having a 135-mm lens; and 30 images scanned from the Brodatz collection[20] of natural-image textures. The midrange, close-up, and Brodatz libraries were calibrated for the modulation-transfer function of the image-acquisition system, and the midrange and close-up images were also calibrated for the gamma function of the image-acquisition system. The midrange and close-up libraries covered a wide range of subjects and viewpoints, and image content was not restricted to pastoral scenes.

The power-spectral properties of each image were analyzed by the following method. Each image was subjected to a 2-D fast Fourier transform that used a minimum-bias spectrum-estimation window.[21] The power spectrum was computed, interpolated into polar Fourier coordinates (ω, θ), and averaged across orientation to yield a 1-D function of radial spatial frequency. The power-law model $\hat{H}(\omega) \propto \omega^{-\beta}$ was applied to these data by transforming both power and spatial-frequency axes to logarithmic coordinates and performing a linear regression (by the method of least squares). Two alternative models were also investigated in the log–log power-spectral domain: a second-order polynomial model and a piecewise-linear model. F tests of the residual sum of squares showed that significant ($P < 0.01$) curvature (also reported elsewhere[22]) and two-stage piecewise linearity were common across this image ensemble, but the correlation coefficient associated with the simple linear regressions never fell below 0.95: The power-law model appeared robust in that it accounted for the vast majority of the variance in the power-spectral data. The range of β values spanned by these three image libraries was 1.57–3.10 with a mean of 2.38.

3. METHODS

A. Selection of Original Images
The wide range of natural spectral falloffs reported both here and elsewhere might be considered[1,8] at odds with the single-exponent power-law model that underlies the sparse-distributed encoding strategy discussed in Section 1. This problem was addressed within the framework of a psychophysical paradigm in the following manner. The images were drawn from all image libraries so that the experimental image ensemble could be divided into three classes on the basis of their natural spectral falloffs β; that is, all three images within a given class had identical values of β (see Fig. 2). By recording discrimination thresholds not as a function of natural spectral falloff but rather as changes (denoted as $\Delta\beta$) in β relative to the natural spectral falloff of each class, it was possible to de-
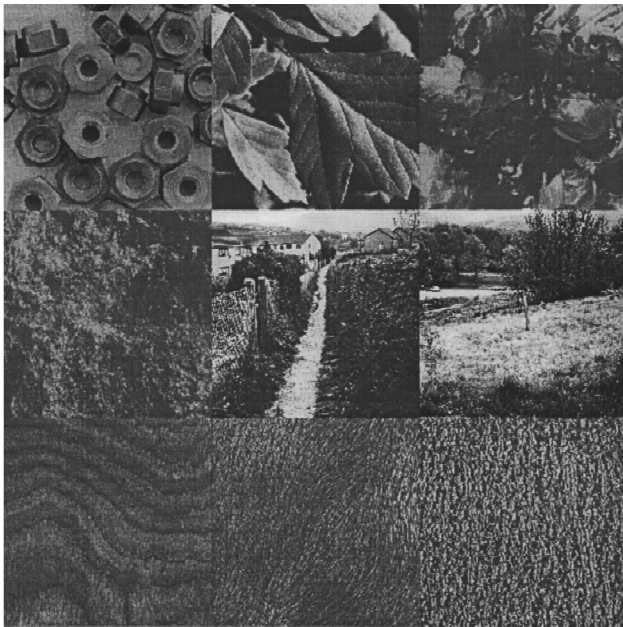
Fig. 2. Experimental image set, divided into three classes on the basis of natural power-spectral falloff β. Top row: $\beta = 2.6$; middle row: $\beta = 2.0$; bottom row: $\beta = 1.4$.

termine whether observers' ability to perform the discrimination task depended on the natural correlational structure (as quantified by the natural spectral falloff β) of each individual scene[7] or whether discriminations were performed relative to some fixed value[8] of β.

To investigate the effects of the removal of natural-phase structure on observer performance, a random-phase ensemble was generated by subjecting each of the nine images shown in Fig. 2 to a 2-D fast Fourier transform, replacing the natural-phase spectrum with a random-phase spectrum (in which the phase coefficients were drawn from a uniform distribution over the interval $[-\pi, \pi]$), and performing the inverse Fourier transform. The resulting images are shown, in their corresponding positions, in Fig. 3.

B. Image-Generation Protocol

A set of filtered images was generated by computer (Sun Microsystems, Santa Clara, Calif., IPX SPARCstation) from each of the nine root images shown in either Fig. 2 (experiment 1) or Fig. 3 (experiment 2). This was done by subjecting each root image to a process of incremental filtering to generate a set of reference images whose values of β were offset by a factor $\Delta\beta$ from that measured for the root image. The values[23] of $\Delta\beta$ used were $-0.9, -0.6, \ldots, 0.9$, and details of this incremental filtering procedure are given in Appendix A.

Each of the seven reference images was then further filtered by a series of increments in β to create a set of eight test images suitable for determining a psychometric function at each value of $\Delta\beta$; a total of 56 images were thus derived from each root image. These images were normalized against that root image for the distribution of power across orientation and for total variance, scaled to a 9-bit luminance range, and then spatially vignetted;

further details of these procedures are given in Appendix A.

C. Apparatus

Stimuli were presented on a high-resolution color monitor (Sony Japan Ltd., Multisync Trinitron model) driven by a color raster system (Ramtek UK Ltd., modified 4660 series); local timing was controlled by a dedicated computer (Sun Microsystems, Santa Clara, Calif., model 3/160). The images were 512×512 pixels in size; with the viewing distance set at 1.3 m, the available spatial-frequency bandwidth was 0.2–38 cycles deg^{-1}. The midgray of the monitor (gray level 256) had a luminance of 30 cd m^{-2}.

D. Procedure

A temporal-sequential image-presentation paradigm was used, since in spatial-sequential paradigms (such as a spatial two-alternative forced-choice paradigm) observers frequently report difficulties in maintaining fixation, tending instead to scan from one image to the other in search of discrimination cues. On each trial the observer was presented, in sequence, with an image, an inter-stimulus field, a second image, and a poststimulus field; he or she then had to decide whether the two images were identical. Each pair of images was presented in the central visual field. The interstimulus and poststimulus fields consisted of uniform gray backgrounds. The presentation time for all images and masks was set at 800 (± 50) ms, an acceptable compromise between the need to maintain consistent levels of observer performance and the need to make the trials reasonably fast. Observers used a two-pushbutton switchbox to signal their responses to the stimuli.

Trials were performed in runs: in each single experimental run, the method of constant stimuli was used to determine a psychometric function for each of the nine
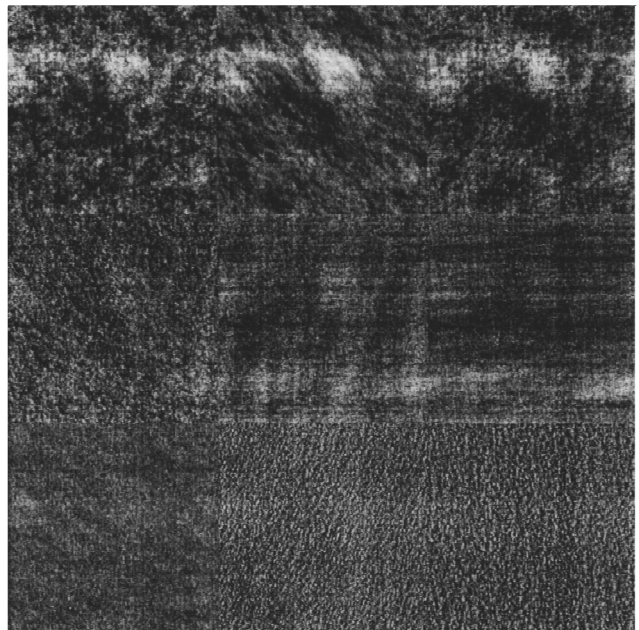

Fig. 3. Experimental image set after full phase randomization; the relative positions of the images are the same as those in Fig. 2.

images around just one value of the reference falloff. An entire experiment thus required seven runs; the reference falloff ($\Delta\beta$) was constant within a single run, but the order of the seven possible values was randomized across runs. On 25% of trials, the first image was the reference image, and the second image was a test image that differed from it by one of the eight increments ($\delta\beta$) in β; on a further 25% of trials, the order of reference and test images was reversed; and in the remaining 50% of trials, the two images were identical reference images. These three conditions, the nine images, and the eight increments were all randomized in each experimental run; each test image was seen a total of 16 times by each observer. A single run consisted of a block of 1280 trials and lasted approximately 2 h; each run was divided into two separate sessions of 1 h each.

E. Subjects

There were two observers, JCH and MGAT (a coauthor), both of whom had normal or corrected-to-normal visual acuity (6/6 Snellen acuity) and less than 0.25 diopters of astigmatism. Observer JCH was unaware of the purpose of the experiment.

4. DATA ANALYSIS

Observers' ability to discriminate between the members of each image pair was measured by an application of the theory of signal detection[24] to the special case of a Yes–No matching paradigm. The theory asserts[25] that the observer decides between two alternative hypotheses (that the two signals presented were the same or that they were different) by comparing the input from each interval and forming different likelihood estimates under the "same" and "different" hypotheses. Under the assumption that these two hypotheses are represented in the relevant decision space by multivariate Gaussian probability density functions of equal variance, the raw hit rates and the false-alarm rates can be linearized by application of the inverse cumulative Gaussian transform; their difference after linearization is the discrimination index d' of signal-detection theory.

If the variances of the d' at each value of $\delta\beta$ are small, then, provided that the hit rate and the false-alarm rate are binomially distributed across the trials, it is possible to estimate the variances of the linearized scores and hence of d' itself.[26] In this way a plot of discrimination index d' against $\delta\beta$ was obtained at each reference falloff $\Delta\beta$. These discrimination plots were fitted (by the method of least squares) with quadratic functions, since some showed significant curvature ($P < 0.01$). Thresholds were computed for various criterion levels of d', and a bootstrap procedure[26] was used to estimate the variances of those thresholds. The final choice of criterion level was judged *a posteriori* on the basis of the variance of the threshold estimate and the stability of the threshold distribution around that criterion value; the optimum level was found to be $d' = 1.5$, which, by virtue of the relationship between Yes–No and two-alternative forced-choice paradigms in psychophysical matching tasks,[25] corresponds to 77% correct in a Yes–No task performed by an unbiased observer.

5. RESULTS

A. Natural-Phase Spectra

Figure 4 shows, for both observers, threshold $\delta\beta$ plotted against filtering step $\Delta\beta$, together with the standard deviations, for the three image classes shown in Fig. 2. The discrimination task seemed most difficult (thresholds were highest) when the correlational structure of each image had been subjected to little or no change, that is, when $\Delta\beta$ was close to zero. Note that the *absolute* spectral falloffs ($\beta + \Delta\beta$) were different for each of the three image classes at the $\Delta\beta = 0$ point, but this does not seem to have influenced the trends shown in the figure, in that the same pattern of sensitivity holds for all three image classes.

B. Random-Phase Spectra

Figure 5 shows a corresponding plot of threshold $\delta\beta$ against $\Delta\beta$ for the random-phase images shown in Fig. 3. Although there are some differences between the two observers in respect of overall performance levels, the trend in each case is the same: the plots show a steady rise in thresholds as the power-spectral falloff becomes increasingly negative. Notice that the ordinate scale of this figure differs from that in Fig. 4; this reflects the fact that thresholds are generally higher for stimuli with random-phase spectra.
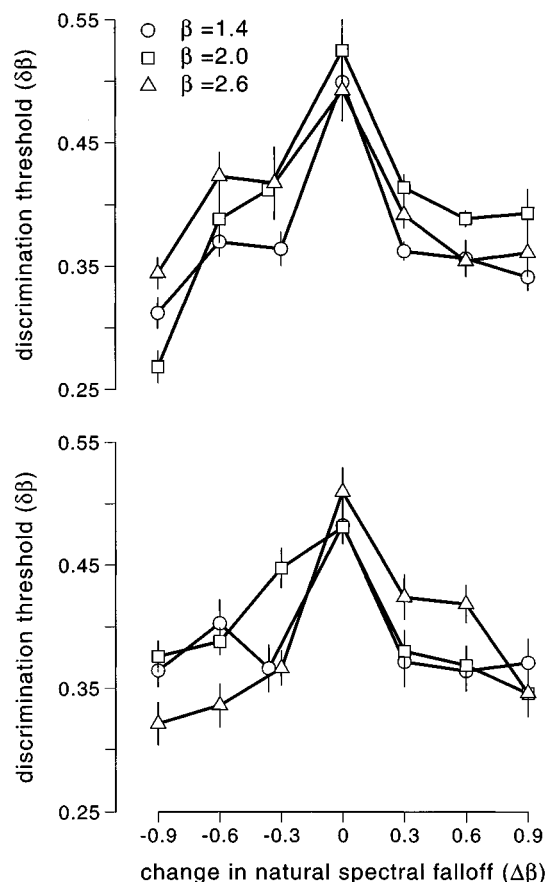

Fig. 4. Discrimination functions for natural-phase stimuli: observers MGAT (top) and JCH (bottom). These functions show observer image-discrimination performance in terms of the discrimination threshold value of $\delta\beta$ (ordinate) as a function of the relative change $\Delta\beta$ in image spectral falloff (abscissa).
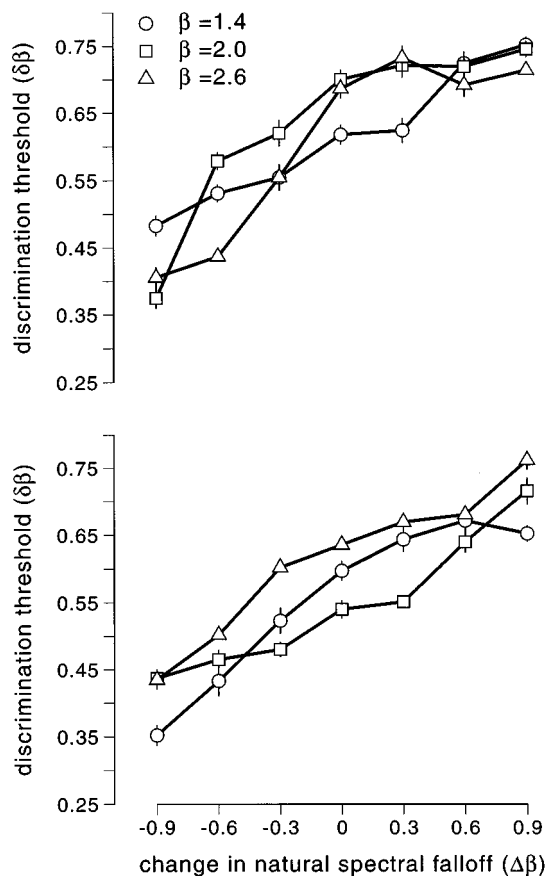
Fig. 5. Discrimination functions for random-phase stimuli: observers MGAT (top) and JCH (bottom). The axes are the same as those in Fig. 4.

6. MODEL OBSERVERS

A. Theory

If observer performance is to be modeled in terms of signal-detection theory, a model observer is required whose specification is complete enough[27] to enable d' values to be computed directly from the stimuli presented in the experiment. The type of model developed here assumed that observers had access only to certain statistical parameterizations of the stimulus data. Similar model observers have been elaborated elsewhere and have been used to fit experimental data gathered from psychophysical procedures such as image detection or discrimination; many of the underlying models are based on some measure of second-order statistical information. For example, one model observer[28] assumes a knowledge of the variance of the stimulus images and uses this information to detect a signal in noise; another[14] assumes a power-law distribution for the power spectra of two stimulus images and performs image discrimination by comparing their spectral exponents.

Two considerations in particular motivate the extension of the model-observer concept to the third-order domain. First, third-order statistics have already been used successfully to model psychophysical data. Klein and Tyler[18] recast the Julesz conjecture in terms of the conventional nth-order correlation functions and chose to measure only zero-shift correlations of order n (i.e., corre-

lations at $\tau_1, \tau_2, ..., \tau_n = 0$); they found that the third moment ($\tau_1 = \tau_2 = 0$) could account satisfactorily for the results of a variety of psychophysical phase-discrimination experiments in many situations in which the second moment failed to do so. The third moment constitutes a highly reduced approach to third-order statistics—it is the first member of a 2-D array of third-order correlations measured at different shifts (τ_1, τ_2) and that together constitute the TCF—but its parsimony is attractive, since it is the simplest global stochastic measure (according to the definitions given in Section 1) that could account for the effects of phase randomization on psychophysical performance. Second, any second-order model observer based on the global SCF must be insensitive to image phase structure and therefore could not account for the effect of phase randomization on the psychophysical results presented here. The effects of phase randomization on image structure are, however, reflected in the differences between the TCF's of a natural image and its phase-randomized counterpart: the phase-randomized images can be considered as finite 2-D samples of a colored Gaussian 2-D random process,[29] whereas stochastic models of natural-phase images can be markedly non-Gaussian in nature.[30] One of the main applications of third-order statistics is in detecting departures from the Gaussian case,[31] and the third moment is often used to compute measures of asymmetry.[32]

Indeed, these arguments lead one toward a comparison of the performance of two model observers: a second-moment model and a third-moment model. It is important to emphasize, however, that local operations on first- or second-order image measures could also account for the psychophysical results: phase randomization and power-spectral filtering can affect local first-order statistics and local second-order statistics to which the SCF is blind. Two additional model observers based on local image measures are therefore also considered. All four models are elaborated in Appendix B, but the principles of each are summarized here:

Second-moment model. This model observer is assumed to process the input data through independent spatial-frequency-selective channels[33]; such a strategy accords with accepted theories of human visual perception. Notice that the output of a bandpass filter generally includes second-order terms on the input; this means that the filter outputs will be sensitive to second-order cross terms in the image and so can be used to quantify relationships between neighboring pixels. The second moments of these filter outputs are then combined as a vector sum.

Third-moment model. Again, the input data are processed through independent channels, but this time it is the third moments of the outputs of the channels that are combined. Thus the model computes a third-order statistic on filter outputs that are themselves sensitive to second-order cross terms in the image; the model is, however, insensitive to third-order cross terms in the image.

First-order model. In this case the pixel values of the two images to be discriminated are compared in the manner of a bit metric.[34] This model can therefore distinguish between two images with different first-order pixel histograms.

Second-order model. As in the first-order model, the pixel values of the two images are first compared, but this operation produces a ratio image, which is then filtered through the same bandpass channels used in the second- and third-moment models. It is therefore best described as a second-order model, since, like the second- and third-moment model observers, it is sensitive to second-order cross terms in the image.

B. Applying the Model Observer to the Image Data

Simulations of the psychophysical experiments were conducted in which each image pair presented to the human observers was supplied as input to the model observers. For the bandpass model observers, signal-to-noise ratios (SNR's) were computed in eight octave-band channels (the images were 512×512 pixels in size) as described by Eqs. (B3), (B6), and (B7) (Appendix B) and were then combined by using Eq. (B5) (Appendix B) to yield a single value of the discrimination index d' for each (unique) stimulus pair. In this way it was possible to construct a model-observer d' plot (analogous to the human observers' linearized psychometric functions) as a function of the stimulus parameter $\delta\beta$. Model-observer thresholds were determined at a criterion level of $d' = 1.5$, exactly as in Section 4.

7. MODEL-OBSERVER PREDICTIONS

It is difficult to compare trends across $\Delta\beta$ in model-observer thresholds with the corresponding trends computed from the psychophysical data, since model-observer discrimination was much more efficient than human-observer discrimination and the model-observer thresholds were consequently much lower (no internal noise was incorporated explicitly into the models). This problem was addressed by normalizing the threshold measured at each value of $\Delta\beta$ to the lowest threshold measured for that observer (human or model). This procedure can be considered equivalent to incorporating additive, Gaussian internal noise into the model observers, such that the decision process was assumed to be noisy after combination of the channels' SNR's.

A. Natural-Phase Images

Figure 6 shows plots of the two human observers' discrimination performance, along with similar plots calculated for the four model observers; all six plots have been normalized against the minimum threshold. The first- and second-order model observers both produce very poor predictions of human-observer performance. Notice how the third-moment model outperforms the second-moment model in accounting for the experimental data.

B. Random-Phase Images

Figure 7 shows the corresponding data for images with random-phase spectra. Second- and third-moment models produce similar results, each providing a good fit for the human observers' data, whereas the first- and second-order models fail again to account adequately for the trends in human performance. The results for the second-moment models in Figs. 6 and 7 are, as expected,
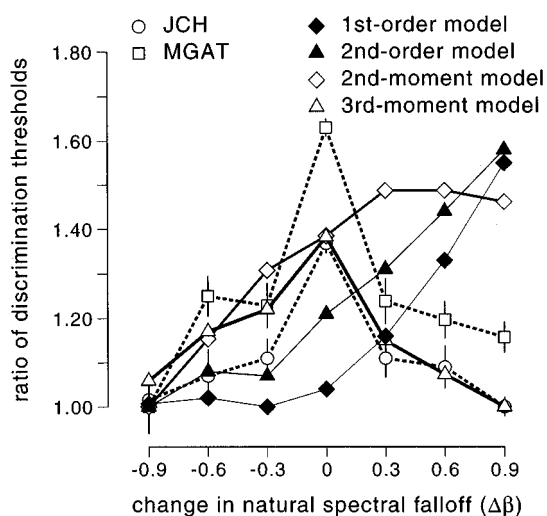

Fig. 6. Discrimination functions for natural-phase stimuli: data obtained from the human observers compared with the results predicted by the four model observers. All discrimination functions have been normalized to minimum threshold; otherwise, the axes are the same as those in Fig. 4.
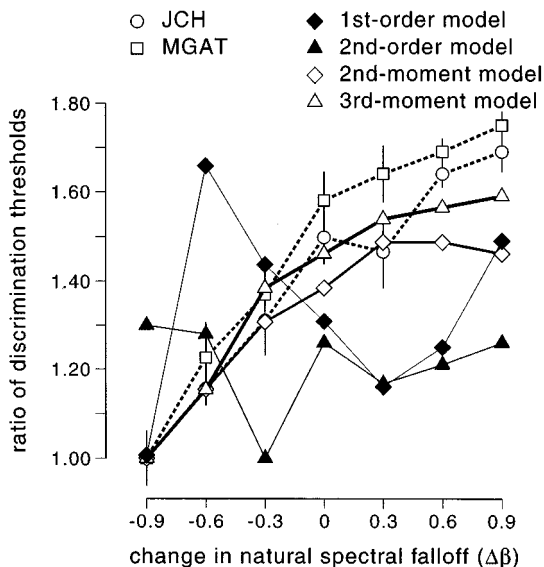

Fig. 7. Same as Fig. 6, but for random-phase stimuli.

identical: phase randomization cannot affect global second-order structure (as defined by the SCF or the power spectrum).

8. DISCUSSION

The results shown in Fig. 4 are in partial agreement with those of Tadmor and Tolhurst;[8] they found evidence for a peak in threshold at some intermediate spectral falloff but reported that the position of this peak depended on absolute spectral falloff rather than on the increment relative to each image's natural spectral falloff. (There were, however, procedural differences between their experimental paradigm and that described here: they used a spatial three-alternative forced-choice paradigm with central fixation, and the images used had their power spectra flattened before spectral filtering.) These

authors also found[8] that this marked trend in performance was largely eliminated by phase-randomizing the stimulus images; Fig. 5 seems to indicate that the trend is not only eliminated but replaced by an upward drift in threshold as the stimulus images' power spectra are made steeper. The results shown in Fig. 5 can also be compared with those reported by Knill *et al.*[14] for a similar psychophysical experiment using Brownian fractal images. (However, their paradigm again differed slightly from the present one: the images that they used were 64×64 pixels in size; subjects were asked to indicate specifically which image had the lower spectral falloff; and discrimination was recorded as a function of absolute spectral falloff.) Knill *et al.* reported that subjects' ability to discriminate between images having two different power-spectral exponents was approximately constant in the region $2.8 < \beta < 3.6$.

Thus, although some controversy must remain over whether the tuning of the human visual system to second-order image properties is relative to an ensemble average statistic or to the natural statistical structure of each image, the destruction of natural-phase structure by means of phase randomization appears to have the effect of either lessening[8] or completely removing (see Fig. 5) this tuning.

Since the filtering applied to the stimulus images produced changes in power-spectral shape only, it is perhaps not surprising that the second-moment model accounts well for human-observer performance with random-phase images, predicting the rise in discrimination thresholds as the spectra steepen; Knill *et al.*[14] described a simpler version of this model (the low-pass contrast detector) that made similar predictions. When natural-phase images were subjected to exactly the same power-spectral changes, however, a third-moment model was necessary to account adequately for observer performance. The success of the third-moment model (and the failure of the second-moment model) in this respect suggests that discrimination may be mediated by perceptual mechanisms capable of characterizing higher-order image structure. This conclusion is supported by the failure of both first- and second-order models to predict the psychophysical data and can be related to the now-established refutation of the Julesz conjecture—that only first- and second-order statistics are used in human texture-discrimination tasks.[17]

Discrimination performance thus appears particularly poor when each image is in either its natural correlational state (results shown here) or the state that is representative of an ensemble average.[8] What advantages accrue from a visual coding system that behaves in this manner? The image-discrimination tasks reported here were designed to measure a type of increment-threshold function; that is, the test and reference parameters were varied along the same dimension (that of image spectral falloff β). If the analyzing mechanisms of the human visual system are tuned to some distribution of β, then poor discrimination performance at intermediate β values may simply reflect the robustness of the internal visual representation in the face of changes to a statistical image parameter for which visual processing is optimized. Moreover, if the visual code is robust, discrimination between

two different natural images (i.e., images that differ in some respect other than spectral falloff) can be made very efficient, since this discrimination can be accomplished by the further processing of an already sparse representation. Tadmor and Tolhurst[35] hypothesized that such a strategy would be of significant biological value to the organism, producing maximum tolerance of image blur yet optimum image discriminability when images are in their natural correlational states. The effects of phase randomization observed here suggest that the sparse representation is processed by some second-stage analyzers capable of characterizing higher-order image structure, but confirmation of this hypothesis will require the design of further experiments in which the stimulus parameter is some measure of higher-order image statistics.

APPENDIX A: INCREMENTAL FILTERING PROCEDURE

So that differences in second-order statistical structure would be the only cue available to the observers in the experiments, the incremental filtering procedure was designed to effect a change $\delta\beta$ in the spectral falloff β of a given natural image while ensuring that other aspects of the stimuli were not affected by the spectral-filtering process. Spectral filtering could affect deviations from the power-law model of power-spectral behavior, as well as the distribution of power across orientation, the luminance range, and the spatial properties of the stimulus in the neighborhood of the image borders. The following subsections illustrate how these problems were addressed.

1. Higher-Order Power-Spectral Behavior

As was noted in Section 2, some natural images have power spectra that depart from linearity when plotted on double-logarithmic axes, and these deviations should be preserved by the spectral-filtering procedure. An incremental filtering technique was developed that computed the radial spatial frequency ω at each (Cartesian) locus in the power spectrum $\hat{H}_R(\omega, \theta)$ of the original root image; the following transformation was then applied to generate the filtered spectrum $\hat{H}_F(\omega, \theta)$:

$$\hat{H}_F(\omega, \theta) = \hat{H}_R(\omega, \theta)\omega^{-\delta\beta}. \tag{A1}$$

In double-logarithmic coordinates this transformation is equivalent to adding $-\delta\beta \log \omega$ to each log-transformed power-spectral value, so the filtering process should alter the slope of the double-log-transformed spectrum while leaving any higher-order behavior unaffected. The following procedure was used to verify that given the root spectral exponent β_R and the filtering increment $\delta\beta$, Eq. (A1) could be used to predict the new spectral exponent β_F of each filtered image. The polar-interpolated power spectra of the root and filtered images were parameterized after double-logarithmic transformation by linear regression; the two spectral falloffs β_R and β_F were then recorded along with their associated residual sums of squares. The variance of the β_F value computed from the least-squares regression was estimated and compared with the 95% confidence interval for a regression line of

slope $\beta_R + \delta\beta$. None of the values of β_F computed in this way fell outside the 95% confidence interval of the expected distribution. F tests were also conducted to compare the residual sum of squares following linear regression of each filtered image's spectrum with that of its corresponding root image; none of these tests reached significance ($P > 0.1$).

2. Distribution of Image Power across Orientation

Since incremental filtering by a factor $\delta\beta$ adds $-\delta\beta \log \omega$ to each log-transformed power-spectral value, this procedure alters the overall image power and hence, by Parseval's theorem, the contrast of the scene. It is therefore necessary to normalize the total power of the filtered image against that of the root image, but, given the orientational anisotropy typically found in natural images,[36] this normalization must be applied separately to each orientation, as changes in the distribution of image power with respect to orientation are potential cues in image-discrimination tasks.[37]

For both the root and filtered spectra, the distribution of image power across orientation was computed by interpolating the spectra into polar coordinates and summing each radial vector across radial spatial frequency ω. Power-normalization factors were estimated by computing the ratio of root-image power to filtered-image power at each orientation. The appropriate normalization could then be applied to the filtered-image spectrum by calculating the orientation at each (Cartesian) spectral locus and interpolating onto the 1-D distribution of power-normalization factors.

After inverse Fourier transformation, each filtered image was normalized for total variance against the corresponding root image; all images had the same mean luminance, so all images in a given set had the same root-mean-square contrast. The requirement for variance normalization restricts the range of possible filtering increments $\delta\beta$, since it is impossible to compute the variance of an image with infinite power.

The effectiveness of this power-normalization procedure was assessed as follows. Root and filtered images were Fourier transformed, and their power spectra were interpolated into polar coordinates. For every pair of root and filtered spectra, the corresponding distributions of power across orientation were compared by a Kolmogorov–Smirnov test. None of the filtered images had a distribution of power across orientation that differed significantly from that of the corresponding root image ($P > 0.1$).

3. Distribution of Image Luminance Values

Although the images were acquired at a luminance resolution of 8 bits, the processes of gamma and modulation-transfer correction produced calibrated images whose floating-point luminance values were not restricted to the range 0–255. These luminance values, which were then further transformed by the spectral-filtering procedure, had to be truncated to the bit resolution of the digital-to-analog converter in the graphics engine before the images could be displayed on an analog monitor. The graphics engine could support an effective luminance range of 9 bits after gamma and spectroradiometric calibration.

The image luminance values therefore had to be scaled and truncated into the range 0–511. To preserve the constancy of root-mean-square contrast across each image set, it was necessary to determine the scaling factors from the single image in that set with the largest luminance range and apply the same linear scaling to every other image in the same family; a similar procedure has been followed elsewhere.[8] The image in question was always that with the flattest power spectrum (most negative $\delta\beta$), since flattening the power spectrum of an image stretches its gray-level histogram. Student's t test was then used to compare the means of the root and filtered images, and F tests were used to compare the variances of the root and filtered images; none of these tests reached significance ($P > 0.1$).

4. Stimulus Properties near Image Borders

Some psychophysical models of the visibility of aperiodic image features[38] allow for the possibility that subthreshold changes in the relative spatial-frequency content of an image could be accompanied by suprathreshold changes in stimulus structure in the neighborhood of the image borders. To prevent the creation of such artifactual discrimination cues in these experiments, a circular extension of one of the variable-parameter spectrum-estimation windows[21] was introduced as a vignette; a similar procedure has been described elsewhere.[39]

APPENDIX B: DERIVATION OF MODEL OBSERVERS

The psychophysical paradigm described in Section 3 falls into the category of fixed-base[25] procedures, which can be related to the simpler Yes–No signal-detection paradigm by assuming that the observer is required in each trial to make a binary decision about whether a difference image $s(x, y)$ [the difference in the luminance patterns of members of a stimulus pair as a function of position (x, y), $0 \leq x, y \leq N - 1$.] is present as well as a background image $t(x, y)$ (the root image). When both difference and background images are defined only statistically, the difference image can be identified with the stimulus signal and the background image can be identified with the stimulus noise. It then remains to define an appropriate metric for comparing the difference and background images. A simple way of assessing first-order luminance differences, used here in the first-order model, is to compute SNR's by means of a bit metric similar to that described by Hubner $et\ al.$[34]:

$$\text{SNR} = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \frac{|s(x, y)|}{t(x, y)}. \tag{B1}$$

The model can be channelized[33] to make a bandpass model observer by feeding its output into a number of bandpass filters, computing the associated SNR's on a channel-by-channel basis, and pooling these bandpass SNR's to yield a single estimate of d'. This procedure requires an appropriate metric, a bank of suitable filters, and some pooling rules; these are defined for the second-order model as follows:

Metric. The first-order ratios of the difference and background images were first computed to produce a ratio image $c(x, y)$:

$$c(x, y) = \frac{|s(x, y)|}{t(x, y)}. \tag{B2}$$

It was this 2-D function $c(x, y)$ that was then subjected to bandpass filtering.

Filtering. It is well known that convolution with an ideal bandpass filter introduces aliasing into the output signal, and it is instead necessary to use a filter function whose Fourier-asymptotic behavior produces good localization on both the space and spatial-frequency axes. Mallat[40] has described a family of filters that have these properties and that can be used to decompose a signal into a family of orthogonal octave-wide subbands by means of the discrete dyadic wavelet transform (DDWT). Mallat's algorithmic implementation[40] of the DDWT, which uses cubic-spline wavelets, was followed here with one modification: to avoid introducing orientation dependency into the transform, it was necessary to use a circular 2-D extension of these wavelets rather than take advantage of the row–column separability of the 2-D convolution integral.[41] For the second-order model observer, a series of bandpass SNR's was computed by decomposing the ratio image $c(x, y)$ into a number of independent spatial-frequency bands and summing spatially over the resulting subband images $c_i(x, y)$ as follows:

$$\mathrm{SNR}_i = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} |(c_i(x, y) - \bar{c}_i)|. \tag{B3}$$

Pooling rules. Although several pooling rules have been discussed in the literature,[42] the assumption that observers base their decisions on the computation of likelihood ratios leads one to consider an integration model,[24] which predicts that the discrimination index d' resulting from the combination of n independent observations in channels with individual discrimination indices d'_1, d'_2, ..., d'_n is given by

$$d' = \left[\sum_{i=1}^{n} (d'_i)^2 \right]^{1/2}. \tag{B4}$$

Denote the bandpass SNR corresponding to the ith octave band by SNR_i; then d' can be computed as

$$d' = \left[\sum_{i=1}^{n} (\mathrm{SNR}_i)^2 \right]^{1/2}. \tag{B5}$$

The second-moment observer behaves rather differently, since the comparison between difference and background images is performed within every channel. First, the images s and t are decomposed separately (with the filters described above) to produce families of subband images s_i and t_i; second, the ratio of the second moments of these subband images is computed:

$$(\mathrm{SNR}_i)^2 = \frac{\displaystyle\sum_{x=0}^{N-1} \sum_{y=0}^{N-1} [s_i(x, y) - \bar{s}_i]^2}{\displaystyle\sum_{x=0}^{N-1} \sum_{y=0}^{N-1} [t_i(x, y) - \bar{t}_i]^2}; \tag{B6}$$

third, the SNR's are pooled as described above.

The third-moment observer differs from the second-moment observer only at the second stage, where the following measure is computed in each spatial-frequency band from the subband images s_i and t_i:

$$(\mathrm{SNR}_i)^2 = \frac{\left| \displaystyle\sum_{x=0}^{N-1} \sum_{y=0}^{N-1} [s_i(x, y) - \bar{s}_i]^3 \right|}{\left| \displaystyle\sum_{x=0}^{N-1} \sum_{y=0}^{N-1} [t_i(x, y) - \bar{t}_i]^3 \right|}. \tag{B7}$$

This SNR is therefore computed as a ratio of signal third-order moment to noise third-order moment, where signal and noise are the bandpass-filtered difference and background images, respectively. These third-moment measures are therefore (highly restricted) samples[43] of the TCF's of difference and background images, respectively; notice that they are not simply equivalent to raising the SNR of the second-moment model to the power of 1.5, since the summations operate after the nonlinearities.

ACKNOWLEDGMENTS

REFERENCES AND NOTES

1. 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).
2. L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," J. Opt. Soc. Am. A **4**, 519–524 (1987).
3. R. J. Baddeley and P. J. B. Hancock, "A statistical analysis of natural images matches psychophysically derived orientation tuning curves," Proc. R. Soc. London Ser. B **246**, 219–223 (1991).
4. Consider an array of idealized (rectangular) channels, each with low-frequency cutoff ω_l, high-frequency cutoff ω_h, and center frequency ω_c; if $\omega_h = 2\omega_l$, the area (in polar coordinates of spatial frequency and orientation) covered by an isotropic band of frequencies between ω_l and ω_h is a constant proportion of $\omega_c{}^2$.
5. R. L. DeValois, D. G. Albrecht, and L. G. Thorell, "Spatial frequency selectivity of cells in macaque visual cortex," Vision Res. **22**, 545–559 (1982).
6. Y. Tadmor and D. J. Tolhurst, "Both the phase and amplitude spectrum may determine the appearance of natural images," Vision Res. **33**, 141–145 (1993).
7. M. G. A. Thomson and D. H. Foster, "Effect of phase refiltering on visual tuning to second-order image structure," Perception **22** (Suppl.), 121–122 (1993).
8. Y. Tadmor and D. J. Tolhurst, "Discrimination of changes in the second-order statistics of natural and synthetic images," Vision Res. **34**, 541–554 (1994).

9. A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," Proc. IEEE **69**, 529–541 (1981).
10. L. N. Piotrowski and F. W. Campbell, "A demonstration of the visual importance and flexibility of spatial-frequency amplitude and phase," Perception **11**, 337–346 (1982).
11. D. J. Field, "Scale-invariance and self-similar 'wavelet' transforms: an analysis of natural scenes and mammalian visual systems," in *Wavelets, Fractals and Fourier Transforms*, M. Farge, J. C. R. Hunt, and J. C. Vassilicos, eds. (Clarendon, Oxford, 1993).
12. M. C. Morrone and D. C. Burr, "Feature detection in human vision: a phase-dependent energy model," Proc. R. Soc. London Ser. B **235**, 221–245 (1988).
13. J. H. Van Hateren, "Theoretical predictions of spatiotemporal receptive fields of fly LMCs, and experimental validation," J. Comp. Physiol. A **171**, 157–170 (1992).
14. D. C. Knill, D. Field, and D. Kersten, "Human discrimination of natural images," J. Opt. Soc. Am. A **7**, 1113–1123 (1990).
15. D. J. Field, "What the statistics of natural images tell us about visual coding," in *Human Vision, Visual Processing, and Digital Display*, B. E. Rogowitz, ed., Proc. SPIE **1077**, 269–276 (1989).
16. J. I. Yellott, Jr., "Implications of triple correlation uniqueness for texture statistics and the Julesz conjecture," J. Opt. Soc. Am. A **10**, 777–793 (1993).
17. G. Julesz, E. N. Gilbert, L. A. Shepp, and H. L. Frisch, "Inability of humans to discriminate between visual textures that agree in second-order statistics—revisited," Perception **2**, 391–405 (1973).
18. S. A. Klein and C. W. Tyler, "Phase discrimination of compound gratings: generalized autocorrelation analysis," J. Opt. Soc. Am. A **3**, 868–879 (1986).
19. M. B. Priestley, *Nonlinear and Nonstationary Time-Series Analysis* (Academic, London, 1988).
20. P. Brodatz, *Textures: A Photographic Album for Artists and Designers* (Dover, New York, 1956).
21. K. M. M. Prabhu and K. B. Bagan, "Variable parameter window families for digital spectral analysis," IEEE Trans. Acoust. Speech Signal Process. **37**, 946–949 (1989).
22. D. J. Tolhurst, Y. Tadmor, and Tang Chao, "Amplitude spectra of natural images," Ophthalmic. Physiol. Opt. **12**, 229–232 (1992).
23. Since β is defined as the magnitude of the falloff of the power spectrum, positive increments in β will produce steeper power spectra.
24. D. M. Green and J. A. Swets, *Signal Detection Theory* (Wiley, New York, 1966).
25. R. D. Sorkin, "Matching procedures in psychoacoustics," J. Acoust. Soc. Am. **34**, 108–113 (1962).
26. D. H. Foster and W. F. Bischof, "Thresholds from psychometric functions: superiority of bootstrap to incremental and probit variance estimators," Psychol. Bull. **109**, 152–159 (1991).
27. W. S. Geisler, "Sequential ideal-observer analysis of visual discriminations," Psychol. Rev. **96**, 267–314 (1989).
28. H. H. Barrett, J. Yao, J. P. Rolland, and K. J. Myers, "Model observers for the assessment of image quality," Proc. Natl. Acad. Sci. USA **90**, 9758–9765 (1993).
29. J. G. Proakis, C. M. Rader, F. Ling, and C. C. Nikias, *Advanced Digital Signal Processing* (Macmillan, New York, 1992).
30. M. G. A. Thomson and D. H. Foster, "Phase perception and the bispectra of natural images," Opt. Photon. News **5** (No. 8, Suppl.), 110 (1994).
31. A. K. Nandi, "Robust estimation of third-order cumulants in applications of higher-order statistics," IEE Proc. F **140**, 380–389 (1993).
32. S. Elgar, "Relationships involving third moments and bispectra of a harmonic process," IEEE Trans. Acoust. Speech Signal Process. **ASSP-35**, 1725–1726 (1987).
33. 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).
34. M. Hubner, M. Caelli, and I. Rentschler, "Visual phase resolution for grayscale textures," Percept. Psychophys. **43**, 319–325 (1988).
35. Y. Tadmor and D. J. Tolhurst, "Is the human visual system optimized for processing natural images?" Perception **19** (Suppl.), A76 (1990).
36. E. Switkes, M. J. Mayer, and J. A. Sloan, "Spatial-frequency analysis of the visual environment: anistropy and the carpentered environment hypothesis," Vision Res. **18**, 1393–1399 (1978).
37. I. R. Paterson, "The information available in natural textures for classification tasks: a computation and psychophysical investigation," Perception **22** (Suppl.), A68 (1993).
38. F. W. Campbell, R. H. S. Carpenter, and J. Z. Levinson, "Visibility of aperiodic patterns compared with that of sinusoidal gratings," J. Physiol. (London) **204**, 283–298 (1969).
39. D. J. Field and B. Olshausen, "Insights into visual coding gained from an understanding of the statistics of natural scenes," Invest. Ophthalmol. Visual Sci. **37**, 3088 (1996).
40. S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," IEEE Trans. Pattern. Anal. Mach. Intell. **11**, 674–693 (1989).
41. One disadvantage of any nth-order bandpass observer model computed in this way is that the DDWT samples the nth-order correlation function of a signal in a manner that is dictated by the Fourier-asymptotic behavior of the wavelets; the requirement for smooth behavior in both spatial and spectral domains means that some contribution to the total nth-order moment must be lost between the channels. Thus computing the outputs of the filters defined by the DDWT decomposition of an image will not be exactly equivalent to summing within annular sectors of the 2-D polar power spectra of an image. The significance of this discrepancy was tested by computing the second-moment model observer in both spatial and spectral domains; the two methods produced almost identical results, and only results obtained by using the spatial method are shown here.
42. R. F. Quick, "A vector-magnitude model of contrast detection," Kybernetik **16**, 65–67 (1974).
43. The nth-order correlation function of a 2-D process is a $2(n-1)$-dimensional process, whereas the output signals produced by the wavelet transform of a 2-D process are necessarily two dimensional. Bandpass model observers of order 3 and higher therefore operate on progressively smaller samples of the nth-order correlation functions, the remaining nth-order information appearing as spatial correlations between (rather than within) channels. It is possible to elaborate a complete nth-order bandpass observer model in the spectral domain, but for a 2-D process this is prohibitively computationally intensive.[44]
44. V. Chandran and S. Elgar, "Bispectral analysis of 2-D random processes," IEEE Trans. Acoust. Speech Signal Process. **35**, 2181–2190 (1990).