NATURAL IMAGE STATISTICS AND NEURAL REPRESENTATION

Eero P Simoncelli

Howard Hughes Medical Institute, Center for Neural Science, and Courant Institute of Mathematical Sciences, New York University, New York, NY 10003; e-mail: eero.simoncelli@nyu.edu

Bruno A Olshausen

Center for Neuroscience, and Department of Psychology, University of California, Davis, Davis, California 95616; e-mail: baolshausen@ucdavis.edu

■ **Abstract** It has long been assumed that sensory neurons are adapted, through both evolutionary and developmental processes, to the statistical properties of the signals to which they are exposed. Attneave (1954) and Barlow (1961) proposed that information theory could provide a link between environmental statistics and neural responses through the concept of coding efficiency. Recent developments in statistical modeling, along with powerful computational tools, have enabled researchers to study more sophisticated statistical models for visual images, to validate these models empirically against large sets of data, and to begin experimentally testing the efficient coding hypothesis for both individual neurons and populations of neurons.

INTRODUCTION

Understanding the function of neurons and neural systems is a primary goal of systems neuroscience. The evolution and development of such systems is driven by three fundamental components: (*a*) the tasks that the organism must perform, (*b*) the computational capabilities and limitations of neurons (this would include metabolic and wiring constraints), and (*c*) the environment in which the organism lives. Theoretical studies and models of neural processing have been most heavily influenced by the first two. But the recent development of more powerful models of natural environments has led to increased interest in the role of the environment in determining the structure of neural computations.

The use of such ecological constraints is most clearly evident in sensory systems, where it has long been assumed that neurons are adapted, at evolutionary, developmental, and behavioral timescales, to the signals to which they are exposed.

1193

Because not all signals are equally likely, it is natural to assume that perceptual systems should be able to best process those signals that occur most frequently. Thus, it is the statistical properties of the environment that are relevant for sensory processing. Such concepts are fundamental in engineering disciplines: Source coding, estimation, and decision theories all rely heavily on a statistical "prior" model of the environment.

The establishment of a precise quantitative relationship between environmental statistics and neural processing is important for a number of reasons. In addition to providing a framework for understanding the functional properties of neurons, such a relationship can lead to the derivation of new computational models based on environmental statistics. It can also be used in the design of new forms of stochastic experimental protocols and stimuli for probing biological systems. Finally, it can lead to fundamental improvements in the design of devices that interact with human beings.

Despite widespread agreement that neural processing must be influenced by environmental statistics, it has been surprisingly difficult to make the link quantitatively precise. More than 40 years ago, motivated by developments in information theory, Attneave (1954) suggested that the goal of visual perception is to produce an efficient representation of the incoming signal. In a neurobiological context, Barlow (1961) hypothesized that the role of early sensory neurons is to remove statistical redundancy in the sensory input. Variants of this "efficient coding" hypothesis have been formulated by numerous other authors (e.g. Laughlin 1981, Atick 1992, van Hateren 1992, Field 1994, Riecke et al 1995).

But even given such a link, the hypothesis is not fully specified. One needs also to state which environment shapes the system. Quantitatively, this means specification of a probability distribution over the space of input signals. Because this is a difficult problem in its own right, many authors base their studies on empirical statistics computed from a large set of example images that are representative of the relevant environment. In addition, one must specify a timescale over which the environment should shape the system. Finally, one needs to state which neurons are meant to satisfy the efficiency criterion, and how their responses are to be interpreted.

There are two basic methodologies for testing and refining such hypotheses of sensory processing. The more direct approach is to examine the statistical properties of neural responses under natural stimulation conditions (e.g. Laughlin 1981, Rieke et al 1995, Dan et al 1996, Baddeley et al 1998, Vinje & Gallant 2000). An alternative approach is to "derive" a model for early sensory processing (e.g. Sanger 1989, Foldiak 1990, Atick 1992, Olshausen & Field 1996, Bell & Sejnowski 1997, van Hateren & van der Schaaf 1998, Simoncelli & Schwartz 1999). In such an approach, one examines the statistical properties of environmental signals and shows that a transformation derived according to some statistical optimization criterion provides a good description of the response properties of a set of sensory neurons. In the following sections, we review the basic conceptual framework for linking environmental statistics to neural processing, and we discuss a series of examples in which authors have used one of the two approaches described above to provide evidence for such links.

BASIC CONCEPTS

The theory of information was a fundamental development of the twentieth century. Shannon (1948) developed the theory in order to quantify and solve problems in the transmission signals over communication channels. But his formulation of a quantitative measurement of information transcended any specific application, device, or algorithm and has become the foundation for an incredible wealth of scientific knowledge and engineering developments in acquisition, transmission, manipulation, and storage of information. Indeed, it has essentially become a theory for computing with signals.

As such, the theory of information plays a fundamental role in modeling and understanding neural systems. Researchers in neuroscience had been perplexed by the apparent combinatorial explosion in the number of neurons one would need to uniquely represent each visual (or other sensory) pattern that might be encountered. Barlow (1961) recognized the importance of information theory in this context and proposed that an important constraint on neural processing was informational (or coding) efficiency. That is, a group of neurons should encode as much information as possible in order to most effectively utilize the available computing resources. We will make this more precise shortly, but several points are worth mentioning at the outset.

1. The efficiency of the neural code depends both on the transformation that maps the input to the neural responses and on the statistics of the input. In particular, optimal efficiency of the neural responses for one input ensemble does not imply optimality over other input ensembles!

2. The efficient coding principle should not be confused with optimal compression (i.e. rate-distortion theory) or optimal estimation. In particular, it makes no mention of the accuracy with which the signals are represented and does not require that the transformation from input to neural responses be invertible. This may be viewed as either an advantage (because one does not need to incorporate any assumption regarding the form of representation, or the cost of misrepresenting the input) or a limitation (because such costs are clearly relevant for real organisms).

3. The simplistic efficient coding criterion given above makes no mention of noise that may contaminate the input stimulus. Nor does it mention uncertainty or variability in the neural responses to identical stimuli. That is, it assumes that the neural responses are deterministically related to the input signal. If these sources of external and internal noise are small compared with the stimulus and neural response, respectively, then the criterion described is approximately optimal. But a more complete solution should take noise into account, by maximizing the information that the responses provide about the stimulus (technically, the mutual information between stimulus and response). This quantity is generally difficult to measure, but Bialek et al (1991) and Rieke et al (1995) have recently developed approximate techniques for estimating it.

If the efficient coding hypothesis is correct, what behaviors should we expect to see in the response properties of neurons? The answer to this question may be neatly separated into two relevant pieces: the shape of the distributions of individual neural responses and the statistical dependencies between neurons.

Efficient Coding in Single Neurons

Consider the distribution of activity of a single neuron in response to some natural environment.[1] In order to determine whether the information conveyed by this neuron is maximal, we need to impose a constraint on the response values (if they can take on any real value, then the amount of information that can be encoded is unbounded). Suppose, for example, that we assume that the responses are limited to some maximal value, R_{max}. It is fairly straightforward to show that the distribution of responses that conveys maximal information is uniform over the interval [0, R_{max}]. That is, an efficient neuron should make equal use of all of its available response levels. The optimal distribution depends critically on the neural response constraint. If one chooses, for example, an alternative constraint in which the variance is fixed, the information-maximizing response distribution is a Gaussian. Similarly, if the mean of the response is fixed, the information-maximizing response distribution is an exponential.[2]

Efficient Coding in Multiple Neurons

If a set of neurons is jointly encoding information about a stimulus, then the efficient coding hypothesis requires that the responses of each individual neuron be optimal, as described above. In addition, the code cannot be efficient if the effort of encoding any particular piece of information is duplicated in more than one neuron. Analogous to the intuition behind the single-response case, the joint responses should make equal use of all possible combinations of response levels. Mathematically, this means that the neural responses must be statistically independent. Such a code is often called a factorial code, because the joint probability distribution of neural responses may be factored into the product of the individual response probability distributions. Independence of a set of neural responses also means that one cannot learn anything about the response of any one neuron by observing the responses of others in the set. In other words, the conditional probability distribution of the response of one neuron given the responses of other neurons should be a fixed distribution (i.e. should not depend on the

---

[1]For the time being, we consider the response to be an instantaneous scalar value. For example, this could be a membrane potential, or an instantaneous firing rate.

[2]More generally, consider a constraint of the form $\varepsilon[\phi(x)] = c$, where x is the response, ϕ is a constraint function, ε indicates the expected or average value over the responses to a given input ensemble, and c is a constant. The maximally informative response distribution [also known as the maximum entropy distribution (Jaynes 1978)] is $\mathcal{P}(x) \propto e^{-\lambda\phi(x)}$, where λ is a constant.
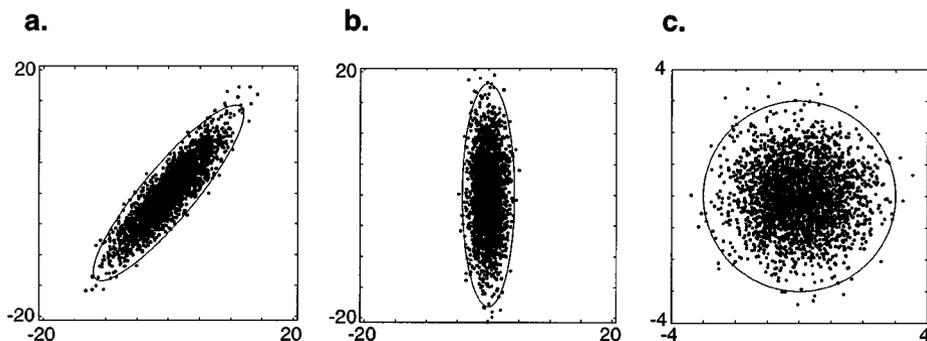
Figure 1: Illustration of principal component analysis on Gaussian-distributed data in two dimensions. (*a*) Original data. Each point corresponds to a sample of data drawn from the source distribution (i.e. a two-pixel image). The ellipse is three standard deviations from the mean in each direction. (*b*) Data rotated to principal component coordinate system. Note that the ellipse is now aligned with the axes of the space. (*c*) Whitened data. When the measurements are represented in this new coordinate system, their components are distributed as uncorrelated (and thus independent) univariate Gaussians.

response levels of the other neurons). The beauty of the independence property is that unlike the result for single neurons, it does not require any auxilliary constraints.

Now consider the problem faced by a "designer" of an optimal sensory system. One wants to decompose input signals into a set of independent responses. The general problem is extremely difficult, because characterizing the joint histogram of the input grows exponentially with the number of dimensions, and thus one typically must restrict the problem by simplifying the description of the input statistics and/or by constraining the form of the decomposition. The most well-known restriction is to consider only linear decompositions, and to consider only the second-order (i.e. covariance or, equivalently, correlation) properties of the input signal. The solution of this problem may be found using an elegant and well-understood technique known as principal components analysis (PCA)[3]. The principal components are a set of orthogonal axes along which the components are decorrelated. Such a set of axes always exists, although it need not be unique. If the data are distributed according to a multi-dimensional Gaussian,[4] then the components of the data as represented in these axes are statistically independent. This is illustrated for a two-dimensional source (e.g. a two-pixel image) in Figure 1.

---

[3]The axes may be computed using standard linear algebraic techniques: They correspond to the eigenvectors of the data covariance matrix.

[4]A multidimensional Gaussian density is simply the extension of the scalar Gaussian density to a vector. Specifically, the density is of the form $\mathcal{P}(\vec{x}) \propto \exp[-\vec{x}^T \Lambda^{-1} \vec{x}/2]$, where Λ is the covariance matrix. All marginal and conditional densities of this density are also Gaussian.

After transforming a data set to the principal component coordinate system, one typically rescales the axes of the space to equalize the variance of each of the components (typically, they are set to one). This rescaling procedure is commonly referred to as "whitening," and is illustrated in Figure 1.

When applying PCA to signals such as images, it is commonly assumed that the statistical properties of the image are translation invariant (also known as stationary). Specifically, one assumes that the correlation of the intensity at two locations in the image depends only on the displacement between the locations, and not on their absolute locations. In this case, the sinusoidal basis functions of the Fourier transform are guaranteed to be a valid set of principal component axes (although, as before, this set need not be unique). The variance along each of these axes is simply the Fourier power spectrum. Whitening may be achieved by computing the Fourier transform, dividing each frequency component by the square root of its variance, and (optionally) computing the inverse Fourier transform. This is further discussed below.

Although PCA can be used to recover a set of statistically independent axes for representing Gaussian data, the technique often fails when the data are non-Gaussian. As a simple illustration, consider data that are drawn from a source that is a linear mixture of two independent non-Gaussian sources (Figure 2). The non-Gaussianity is visually evident in the long tails of data that extend along two oblique axes. Figure 2 also shows the rotation to principal component axes and the whitened data. Note that the axes of the whitened data are not aligned with those of the space. In particular, in the case when the data are a linear mixture of non-Gaussian sources, it can be proven that one needs an additional rotation of the coordinate system to recover the original independent axes.[5] But the appropriate rotation can only be estimated by looking at statistical properties of the data beyond covariance (i.e. order higher than two).

Over the past decade, a number of researchers have developed techniques for estimating this final rotation matrix (e.g. Cardoso 1989, Jutten & Herault 1991, Comon 1994). Rather than directly optimize the independence of the axis components, these algorithms typically maximize higher-order moments (e.g. the kurtosis, or fourth moment divided by the squared second moment). Such decompositions are typically referred to as independent component analysis (ICA), although this is a bit of a misnomer, as there is no guarantee that the resulting components are independent unless the original source actually was a linear mixture of sources with large higher-order moments (e.g. heavy tails). Nevertheless, one can often use such techniques to recover the linear axes along which the data are most independent.[6] Fortuitously, this approach turns out to be quite successful in the case of images (see below).

---

[5]Linear algebraically, the three operations (rotate-scale-rotate) correspond directly to the singular value decomposition of the mixing matrix.

[6]The problem of blind recovery of independent sources from data remains an active area of research (e.g. Hyvarinen & Oja 1997, Attias 1998, Penev et al 2000).
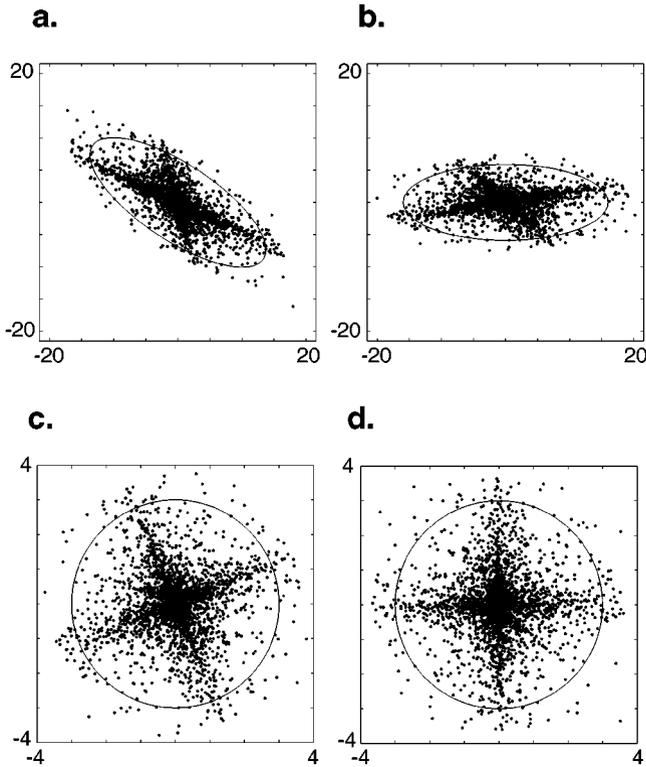
Figure 2 Illustration of principal component analysis and independent component analysis on non-Gaussian data in two dimensions. (*a*) Original data, a linear mixture of two non-Gaussian sources. As in Figure1, each point corresponds to a sample of data drawn from the source distribution, and the ellipse indicates three standard variations of the data in each direction. (*b*) Data rotated to principal component coordinate system. Note that the ellipse is now aligned with the axes of the space. (*c*) Whitened data. Note that the data are not aligned with the coordinate system. But the covariance ellipse is now a circle, indicating that the second-order statistics can give no further information about preferred axes of the data set. (*d*): Data after final rotation to independent component axes.

IMAGE STATISTICS: CASE STUDIES

Natural images are statistically redundant. Many authors have pointed out that of all the visual images possible, we see only a very small fraction (e.g. Attneave 1954, Field 1987, Daugman 1989, Ruderman & Bialek 1994). Kersten (1987) demonstrated this redundancy perceptually by asking human subjects to replace missing pixels in a four-bit digital image. He then used the percentage of correct guesses to estimate that the perceptual information content of a pixel was approximately 1.4 bits [a similar technique was used by Shannon (1948) to estimate the

redundancy of written English]. Modern technology exploits such redundancies every day in order to transmit and store digitized images in compressed formats. In the following sections, we describe a variety of statistical properties of images and their relationship to visual processing.

Intensity Statistics

The simplest statistical image description is the distribution of light intensities in a visual scene. As explained in the previous section, the efficient coding hypothesis predicts that individual neurons should maximize information transmission. In a nice confirmation of this idea, Laughlin (1981) found that the contrast-response function of the large monopolar cell in the fly visual system approximately satisfies the optimal coding criterion. Specifically, he measured the probability distribution of contrasts found in the environment of the fly, and showed that this distribution is approximately transformed to a uniform distribution by the function relating contrast to the membrane potential of the neuron. Baddeley et al (1998) showed that the instantaneous firing rates of spiking neurons in primary and inferior temporal visual cortices of cats and monkeys are exponentially distributed (when visually stimulated with natural scenes), consistent with optimal coding with a constraint on the mean firing rate.

Color Statistics

In addition to its intensity, the light falling on an image at a given location has a spectral (wavelength) distribution. The cones of the human visual system represent this distribution as a three-dimensional quantity. Buchsbaum & Gottshalk (1984) hypothesized that the wavelength spectra experienced in the natural world are well approximated by a three-dimensional subspace that is spanned by cone spectral sensitivities. Maloney (1986) examined the empirical distribution of reflectance functions in the natural world, and showed not only that it was well-represented by a low-dimensional space, but that the problem of surface reflectance estimation was actually aided by filtering with the spectral sensitivities of the cones.

   An alternative approach is to assume the cone spectral sensitivities constitute a fixed front-end decomposition of wavelength, and to ask what processing should be performed on their responses. Ruderman et al (1998), building on previous work by Buchsbaum & Gottschalk (1983), examined the statistical properties of log cone responses to a large set of hyperspectral photographic images of foliage. The use of the logarithm was loosely motivated by psychophysical principles (the Weber-Fechner law) and as a symmetrizing operation for the distributions. They found that the principal component axes of the data set lay along directions corresponding to {L+M+S, L+M−2S, L−M}, where {L,M,S} correspond to the log responses of the long, middle, and short wavelength cones. Although the similarity of these axes to the perceptually and physiologically measured "opponent" mechanisms is intriguing, the precise form of the mechanisms depends on the experiment used to measure them (see Lennie & D'Zmura 1988).
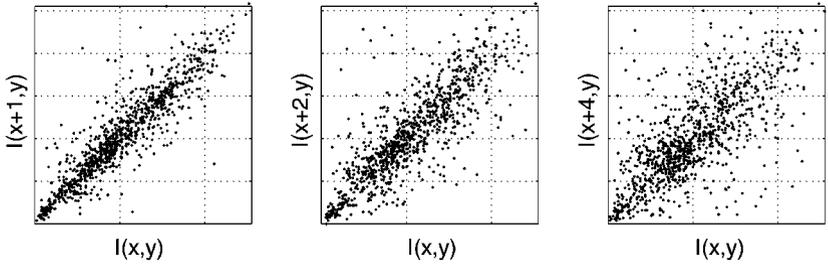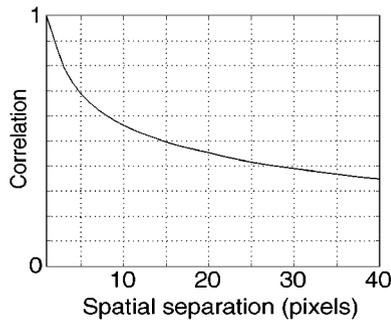
a.

b.

Figure 3 (*a*) Joint distributions of image pixel intensities separated by three different distances. (*b*) Autocorrelation function.

Spatial Correlations

Even from a casual inspection of natural images, one can see that neighboring spatial locations are strongly correlated in intensity. This is demonstrated in Figure 3, which shows scatterplots of pairs of intensity values, separated by three different distances, and averaged over absolute position of several different natural images. The standard measurement for summarizing these dependencies is the autocorrelation function, $C(\Delta x, \Delta y)$, which gives the correlation (average of the product) of the intensity at two locations as a function of relative position. From the examples in Figure 3, one can see that the strength of the correlation falls with distance.[7]

By computing the correlation as a function of relative separation, we are assuming that the spatial statistics in images are translation invariant. As described above,

[7]Reinagel & Zador (1999) recorded eye positions of human observers viewing natural images and found that correlation strength falls faster near these positions than generic positions.

the assumption of translation invariance implies that images may be decorrelated by transforming to the frequency (Fourier) domain. The two-dimensional power spectrum can then be reduced to a one-dimensional function of spatial frequency by performing a rotational average within the two-dimensional Fourier plane. Empirically, many authors have found that the spectral power of natural images falls with frequency, f, according to a power law, $1/f^p$, with estimated values for p typically near 2 [see Tolhurst (1992) or Ruderman & Bialek (1994) for reviews]. An example is shown Figure 4.

The environmental causes of this power law behavior have been the subject of considerable speculation and debate. One of the most commonly held beliefs is that it is due to scale invariance of the visual world. Scale invariance means that the statistical properties of images should not change if one changes the scale at which observations are made. In particular, the power spectrum should not change shape under such rescaling. Spatially rescaling the coordinates of an image by a factor of α leads to a rescaling of the corresponding Fourier domain axes by a factor of $1/\alpha$. Only a Fourier spectrum that falls as a power law will retain its shape under this transformation. Another commonly proposed theory is that the $1/f^2$ power spectrum is due to the presence of edges in images, because edges themselves
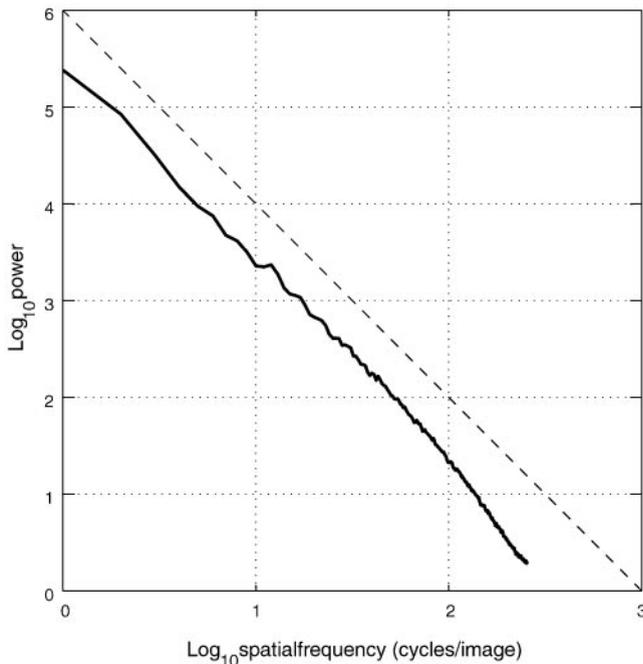
Figure 4 Power spectrum of a natural image (solid line) averaged over all orientations, compared with $1/f^2$ (dashed line).

have a $1/f^2$ power spectrum. Ruderman (1997) and Lee & Mumford (1999) have argued, however, that it is the particular distribution of the sizes and distances of objects in natural images that governs the spectral falloff.

Does the visual system take advantage of the correlational structure of natural images? This issue was first examined quantitatively by Srinivasan et al (1982). They measured the autocorrelation function of natural scenes and then computed the amount of subtractive inhibition that would be required from neighboring photoreceptors in order to effectively cancel out these correlations. They then compared the predicted inhibitory surround fields to those actually measured from first-order interneurons in the compound eye of the fly. The correspondence was surprisingly good and provided the first quantitative evidence for decorrelation in early spatial visual processing.

This type of analysis was carried a step further by Atick & Redlich (1991, 1992), who considered the problem of whitening the power spectrum of natural images (equivalent to decorrelation) in the presence of white photoreceptor noise. They showed that both single-cell physiology and the psychophysically measured contrast sensitivity functions are consistent with the product of a whitening filter and an optimal lowpass filter for noise removal (known as the Wiener filter). Similar predictions and physiological comparisons were made by van Hateren (1992) for the fly visual system. The inclusion of the Wiener filter allows the behavior of the system to change with mean luminance level. Specifically, at lower luminance levels (and thus lower signal-to-noise ratios), the filter becomes more low-pass (intuitively, averaging over larger spatial regions in order to recover the weaker signal). An interesting alternative model for retinal horizontal cells has been proposed by Balboa & Grzywacz (2000). They assume a divisive form of retinal surround inhibition, and show that the changes in effective receptive field size are optimal for representation of intensity edges in the presence of photon-absorption noise.

Higher-Order Statistics

The agreement between the efficient coding hypothesis and neural processing in the retina is encouraging, but what does the efficient coding hypothesis have to say about cortical processing? A number of researchers (e.g. Sanger 1989, Hancock et al 1992, Shonual et al 1997) have used the covariance properties of natural images to derive linear basis functions that are similar to receptive fields found physiologically in primary visual cortex (i.e. oriented band-pass filters). But these required additional constraints, such as spatial locality and/or symmetry, in order to achieve functions approximating cortical receptive fields.

As explained in the introduction, PCA is based only on second-order (covariance) statistics and can fail if the source distribution is non-Gaussian. There are a number of ways to see that the distribution of natural images is non-Gaussian. First, we should be able to draw samples from the distribution of images by generating a set of independent Gaussian Fourier coefficients (i.e. Gaussian white noise), unwhitening these (multiplying by $1/f^2$) and then inverting the Fourier transform.
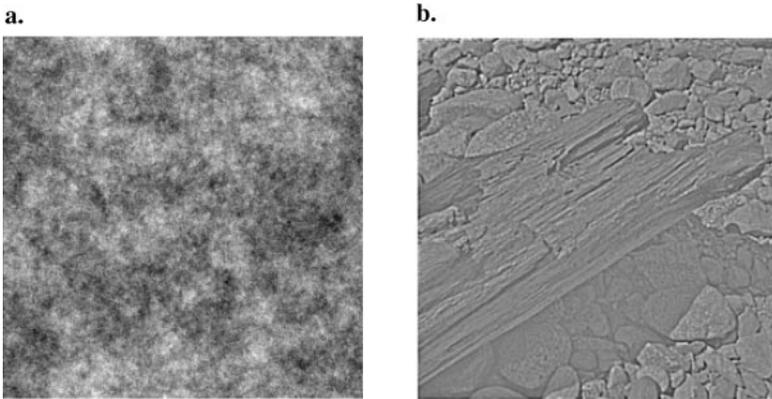
Figure 5   (*a*) Sample of 1/f Gaussian noise; (*b*) whitened natural image.

Such an image is shown in Figure 5*a*. Note that it is devoid of any edges, con-
tours, or many other structures we would expect to find in a natural scene. Second,
if it were Gaussian (and translation invariant), then the Fourier transform should
decorrelate the distribution, and whitening should yield independent Gaussian
coefficients (see Figure 5). But a whitened natural image still contains obvious
structures (i.e. lines, edges, contours, etc), as illustrated in Figure 5*b*. Thus, even if
correlations have been eliminated by whitening in the retina and lateral geniculate
nucleus, there is much work still to be done in efficiently coding natural images.

Field (1987) and Daugman (1989) provided additional direct evidence of the
non-Gaussianity of natural images. They noted that the response distributions of
oriented bandpass filters (e.g. Gabor filters) had sharp peaks at zero, and much
longer tails than a Gaussian density (see Figure 6). Because the density along any
axis of a multidimensional Gaussian must also be Gaussian, this constitutes direct

Figure 6 Histogram of
responses of a Gabor filter
for a natural image, com-
pared with a Gaussian distri-
bution of the same variance.
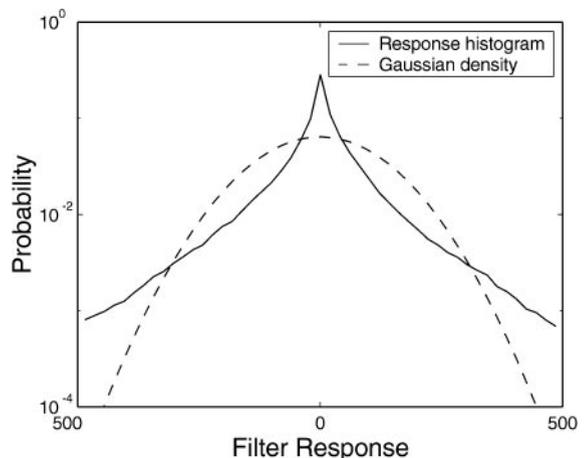
evidence that the overall density cannot be Gaussian. Field (1987) argued that the representation corresponding to these densities, in which most neurons had small amplitude responses, had an important neural coding property, which he termed sparseness. By performing an optimization over the parameters of a Gabor function (spatial-frequency bandwidth and aspect ratio), he showed that the parameters that yield the smallest fraction of significant coefficients are well matched to the range of response properties found among cortical simple cells (i.e. bandwidth of 0.5–1.5 octaves, aspect ratio of 1–2).

Olshausen & Field (1996; 1997) reexamined the relationship between simple-cell receptive fields and sparse coding without imposing a particular functional form on the receptive fields. They created a model of images based on a linear superposition of basis functions and adapted these functions so as to maximize the sparsity of the representation (number of basis functions whose coefficients are zero) while preserving information in the images (by maintaining a bound on the mean squared reconstruction error). The set of functions that emerges after training on hundreds of thousands of image patches randomly extracted from natural scenes, starting from completely random initial conditions, strongly resemble the spatial receptive field properties of simple cells—i.e. they are spatially localized, oriented, and band-pass in different spatial frequency bands (Figure 7). This method may also be recast as a probabilistic model that seeks to explain images in terms of
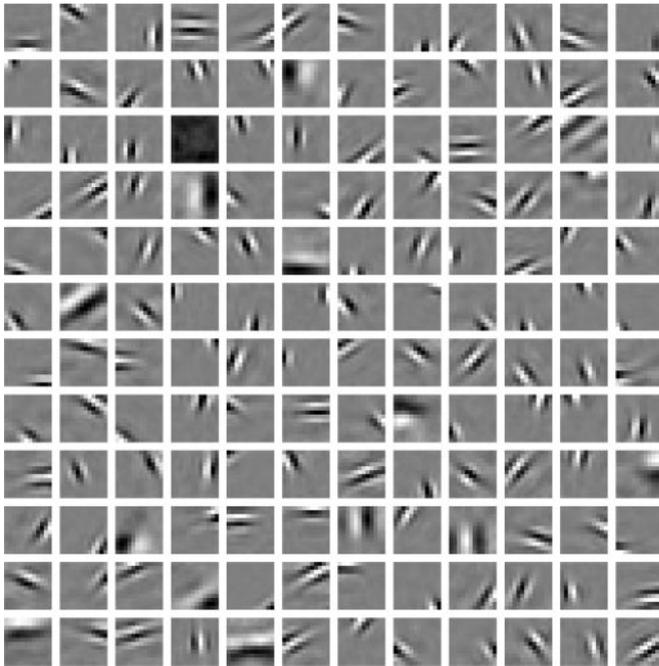
Figure 7 Example basis functions derived using sparseness criterion (see Olshausen & Field 1996).

components that are both sparse and statistically independent (Olshausen & Field 1997) and thus is a member of the broader class of ICA algorithms (see above). Similar results have been obtained using other forms of ICA (Bell & Sejnowski 1997, van Hateren & van der Schaaf 1998, Lewicki & Olshausen 1999), and Hyvärinen & Hoyer (2000) have derived complex cell properties by extending ICA to operate on subspaces. Physiologically Vinje & Gallant (2000) showed that responses of neurons in primary visual cortex were more sparse during presentation of natural scene stimuli.

It should be noted that although these techniques seek statistical independence, the resulting responses are never actually completely independent. The reason is that these models are limited to describing images in terms of linear superposition, but images are not formed as sums of independent components. Consider, for example, the fact that the light coming from different objects is often combined according to the rules of occlusion (rather than addition) in the image formation process. Analysis of the form of these statistical relationships reveals nonlinear dependencies across space as well as across scale and orientation (Wegmann & Zetzsche 1990, Simoncelli 1997, Simoncelli & Schwartz 1999).

Consider the joint histograms formed from the responses of two nonoverlapping linear receptive fields, as shown in Figure 8*a*. The histogram clearly indicates that the data are aligned with the axes, as in the independent components decomposition described above. But one cannot determine from this picture whether the responses are independent. Consider instead the conditional histogram of Figure 8*b*. Each column gives the probability distribution of the ordinate variable r_2, assuming the corresponding value for the abscissa variable, r_1. That is, the data are the
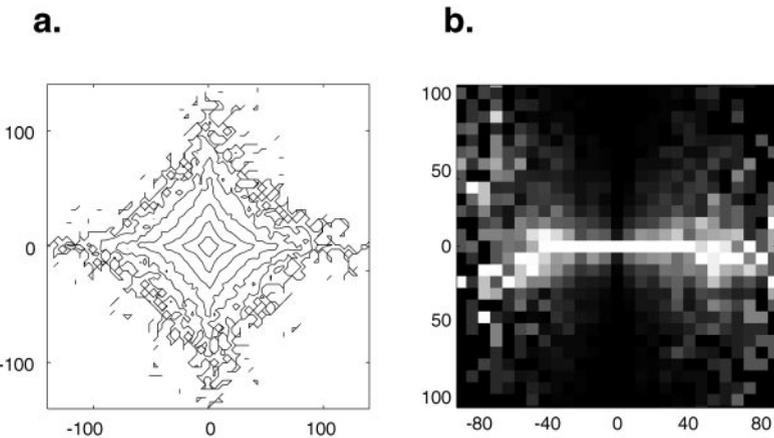
Figure 8 (*a*) Joint histogram of responses of two nonoverlapping receptive fields, depicted as a contour plot. (*b*) Conditional histogram of the same data. Brightness corresponds to probability, except that each column has been independently rescaled to fill the full range of display intensities (see Buccigrossi & Simoncelli 1999, Simoncelli & Schwartz 1999).

same as those in Figure 8*a*, except that each column has been independently normalized. The conditional histogram illustrates several important aspects of the relationship between the two responses. First, they are (approximately) decorrelated: The best-fitting regression line through the data is a zero-slope line through the origin. But they are clearly not independent, because the variance of r_2 exhibits a strong dependence on the value of r_1. Thus, although r_2 and r_1 are uncorrelated, they are still statistically dependent. Furthermore, this dependency cannot be eliminated through further linear transformation.

Simoncelli & Schwartz (1999) showed that these dependencies may be eliminated using a nonlinear form of processing, in which the linear response of each basis function is rectified (and typically squared) and then divided by a weighted sum of the rectified responses of neighboring neurons. Similar "divisive normalization" models have been used by a number of authors to account for nonlinear behaviors in neurons (Reichhardt & Poggio 1973, Bonds 1989, Geisler & Albrecht 1992, Heeger 1992, Carandini et al 1997). Thus, the type of nonlinearity found in cortical processing is well matched to the non-Gaussian statistics of natural images. Furthermore, the weights used in the computation of the normalization signal may be chosen to maximize the independence of the normalized responses. The resulting model is surprisingly good at accounting for a variety of neurophysiological observations in which responses are suppressed by the presence of nonoptimal stimuli, both within and outside of the classical receptive field (Simoncelli & Schwartz 1999, Wainwright et al 2001). The statistical dependency between oriented filter responses is at least partly due to the prevalence of extended contours in natural images. Geisler et al (2001) examined empirical distributions of the dominant orientations at nearby locations and used them to predict psychophysical performance on a contour detection task. Sigman et al (2001) showed that these distributions are consistent with cocircular oriented elements and related this result to the connectivity of neurons in primary visual cortex.

Space-Time Statistics

A full consideration of image statistics and their relation to coding in the visual system must certainly include time. Images falling on the retina have important temporal structure arising from self-motion of the observer, as well as from the motion of objects in the world. In addition, neurons have important temporal response characteristics, and in many cases it is not clear that these can be cleanly separated from their spatial characteristics. The measurement of spatiotemporal statistics in natural images is much more difficult than for spatial statistics, though, because obtaining realistic time-varying retinal images requires the tracking of eye, head, and body movements while an animal interacts with the world. Nevertheless, a few reasonable approximations allow one to arrive at useful insights.

As with static images, a good starting point for characterizing joint space-time statistics is the autocorrelation function. In this case, the spatio-temporal

autocorrelation function $C(\Delta x, \Delta y, \Delta t)$ characterizes the pairwise correlations of image pixels as a function of their relative spatial separation $(\Delta x, \Delta y)$ and temporal separation Δt. Again, assuming spatio-temporal translation invariance, we find that this function is most conveniently characterized in the frequency domain.

The problem of characterizing the spatio-temporal power spectrum was first studied indirectly by van Hateren (1992), who assumed a certain image velocity distribution and a $1/f^2$ spatial power spectrum and inferred from this the joint spatio-temporal spectrum, assuming a $1/f^2$ spatial power spectrum. Based on this inferred power spectrum, van Hateren then computed the optimal neural filter for making the most effective use of the postreceptoral neurons' limited channel capacity (similar to Atick's whitening filter). He showed from this analysis that the optimal neural filter matches remarkably well the temporal response properties of large monopolar cells in different spatial frequency bands. He was also able to extend this analysis to human vision to account for the spatio-temporal contrast sensitivity function (van Hateren 1993).

Dong & Atick (1995a) estimated the spatio-temporal power spectrum of natural images directly by computing the three-dimensional Fourier transform on many short movie segments (each approximately 2–4 seconds in length) and averaging together their power spectra. This was done for an ensemble of commercial films as well as videos made by the authors. Their results, illustrated in Figure 9, show an interesting dependence between spatial and temporal frequency. The slope of the spatial-frequency power spectrum becomes shallower at higher temporal
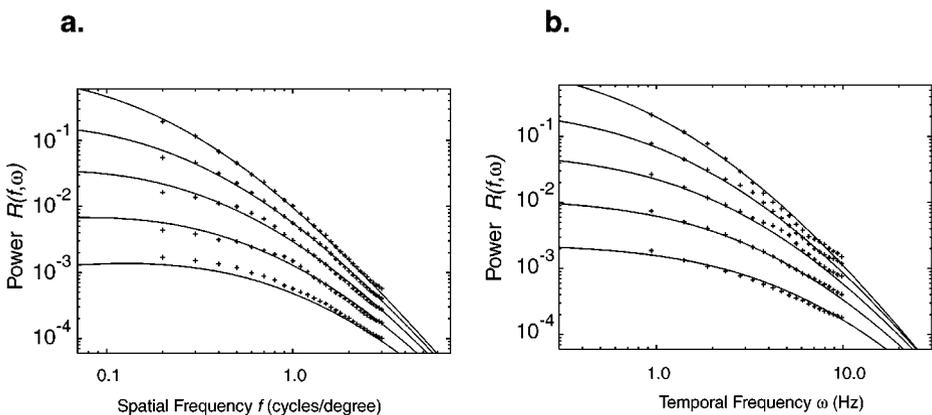
Figure 9 Spatiotemporal power spectrum of natural movies. (*a*) Joint spatiotemporal power spectrum shown as a function of spatial-frequency for different temporal frequencies (1.4, 2.3, 3.8, 6, and 10 Hz, from top to bottom). (*b*) Same data, replotted as a function of temporal frequency for different spatial frequencies (0.3, 0.5, 0.8, 1.3, and 2.1 cy/deg., from top to bottom). Solid lines indicate model fits according to a power-law distribution of object velocities (from Dong & Atick 1995b).

frequencies. The same is true for the temporal-frequency spectrum—i.e. the slope becomes shallower at higher spatial frequencies. Dong & Atick (1995a) showed that this interdependence between spatial and temporal frequency could be explained by assuming a particular distribution of object motions (i.e. a power law distribution), similar in form to van Hateren's assumptions. By again applying the principle of whitening, Dong & Atick (1995b) computed the optimal temporal filter for removing correlations across time and showed that it is closely matched (at low spatial frequencies) to the frequency response functions measured from lateral geniculate neurons in the cat.

Although the match between theory and experiment in the above examples is encouraging, it still does not answer the question of whether or not visual neurons perform as expected when processing natural images. This question was addressed directly by Dan et al (1996) who measured the temporal frequency spectrum of LGN neuron activity in an anaesthetized cat in response to natural movies. Consistent with the concept of whitening, the output power of the cells in response to the movie is fairly flat, as a function of temporal frequency. Conversely, if one plays a movie of Gaussian white noise, in which the input spectrum is flat, the output spectrum from the LGN cells increases linearly with frequency, corresponding to the temporal-frequency response characteristic of the neurons. Thus, LGN neurons do not generically whiten any stimulus, only those exhibiting the same correlational structure as natural images.

The analysis of space-time structure in natural images may also be extended to higher-order statistics (beyond the autocorrelation function), as was previously described for static images. Such an analysis was recently performed by van Hateren & Ruderman (1998) who applied an ICA algorithm to an ensemble of many local image blocks (12×12 pixels by 12 frames in time) extracted from movies. They showed that the components that emerge from this analysis resemble the direction-selective receptive fields of V1 neurons—i.e. they are localized in space and time (within the $12 \times 12 \times 12$ window), spatially oriented, and directionally selective (see Figure 10). In addition, the output signals that result from filtering images with the learned receptive fields have positive kurtosis, which suggests that time-varying natural images may also be efficiently described in terms of a sparse code in which relatively few neurons are active across both space and time. Lewick & Sejnowski (1999) and Olshausen (2001) have shown that these output signals may be highly sparsified so as to produce brief, punctate events similar to neural spike trains.

DISCUSSION

Although the efficient coding hypothesis was first proposed more than forty years ago, it has only recently been explored quantitatively. On the theoretical front, image models are just beginning to have enough power to make interesting predictions. On the experimental front, technologies for stimulus generation and neural
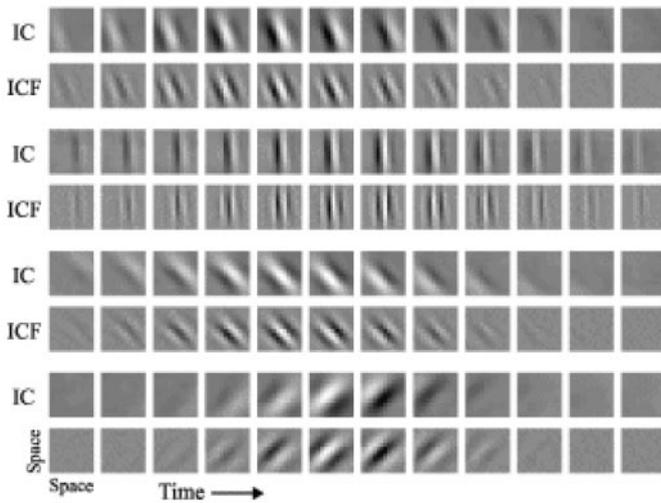
Figure 10 Independent components of natural movies. Shown are four space-time basis functions (rows labeled "IC") with the corresponding analysis functions (rows labeled "ICF"), which would be convolved with a movie to compute a neuron's output (from van Hateren & Ruderman 1998).

recording (especially multiunit recording) have advanced to the point where it is both feasible and practical to test theoretical predictions. Below, we discuss some of the weaknesses and drawbacks of the ideas presented in this review, as well as several exciting new opportunities that arise from our growing knowledge of image statistics.

The most serious weakness of the efficient coding hypothesis is that it ignores the two other primary constraints on the visual system: the implementation and the task. Some authors have successfully blended implementation constraints with environmental constraints (e.g. Baddeley et al 1998). Such constraints are often difficult to specify, but clearly they play important roles throughout the brain. The tasks faced by the organism are likely to be an even more important constraint. In particular, the hypothesis states only that information must be represented efficiently; it does not say anything about what information should be represented. Many authors assume that at the earliest stages of processing (e.g. retina and V1), it is desirable for the system to provide a generic image representation that preserves as much information as possible about the incoming signal. Indeed, the success of efficient coding principles in accounting for response properties of neurons in the retina, LGN, and V1 may be seen as verification of this assumption. Ultimately, however, a richer theoretical framework is required. A commonly proposed example of such a framework is Bayesian decision/estimation theory, which includes both a prior statistical model for the environment and also a loss or reward function that specifies the cost of different errors, or the desirability of different behaviors.

Such concepts have been widely used in perception (e.g. Knill & Richards 1996) and have also been considered for neural representation (e.g. Oram et al 1998).

Another important issue for the efficient coding hypothesis is the timescale over which environmental statistics influence a sensory system. This can range from millenia (evolution), to months (neural development), to minutes or seconds (short-term adaptation). Most of the research discussed in this review assumes the system is fixed, but it seems intuitively sensible that the computations should be matched to various statistical properties on the time scale at which they are relevant. For example, the $1/f^2$ power spectral property is stable and, thus, warrants a solution that is hardwired over evolutionary time scales. On the other hand, several recent results indicate that individual neurons adapt to changes in contrast and spatial scale (Smirnakis et al 1997), orientation (Muller et al 1999), and variance (Brenner et al 2000) on very short time scales. In terms of joint response properties, Barlow & Foldiak (1989) have proposed that short-term adaptation acts to reduce dependencies between neurons, and evidence for this hypothesis has recently been found both psychophysically (e.g. Atick et al 1993, Dong 1995, Webster 1996, Wainwright 1999) and physiologically (e.g. Carandini et al 1998, Dragoi et al 2000, Wainwright et al 2001).

A potential application for efficient coding models, beyond predicting response properties of neurons, lies in generating visual stimuli that adhere to natural image statistics. Historically, visual neurons have been characterized using fairly simple test stimuli (e.g. bars, gratings, or spots) that are simple to parameterize and control, and that are capable of eliciting vigorous responses. But there is no guarantee that the responses measured using such simple test stimuli may be used to predict neural responses to a natural scene. On the other hand, truly naturalistic stimuli are much more difficult to control. An interesting possibility lies in statistical texture modeling, which has been used as a tool for understanding human vision (e.g. Julesz 1962, Bergen & Adelson 1986). Knill et al (1990) and Parraga et al (1999) have shown that human performance on a particular discrimination task is best for textures with natural second-order (i.e. $1/f^2$) statistics, and degraded for images that are less natural. Some recent models for natural texture statistics offer the possibility of generating artificial images that share some of the higher-order statistical structure of natural images (e.g. Heeger & Bergen 1995, Zhu et al 1998, Portilla & Simoncelli 2000).

Most of the models we have discussed in this review can be described in terms of a single-stage neural network. For example, whitening could be implemented by a set of connections between a set of inputs (photoreceptors) and outputs (retinal ganglion cells). Similarly, the sparse coding and ICA models could be implemented by connections between the LGN and cortex. But what comes next? Could we attempt to model the function of neurons in visual areas V2, V4, MT, or MST using multiple stages of efficient coding? In particular, the architecture of visual cortex suggests a hierarchical organization in which neurons become selective to progressively more complex aspects of image structure. In principle, this can allow for the explicit representation of structures, such as curvature, surfaces, or even entire

objects (e.g. Dayan et al 1995, Rao & Ballard 1997), thus providing a principled basis for exploring the response properties of neurons in extra-striate cortex.

Although this review has been largely dedicated to findings in the visual domain, other sensory signals are amenable to statistical analysis. For example, Attias & Schreiner (1997) have shown that many natural sounds obey some degree of self-similarity in their power spectra, similar to natural images. In addition, M S Lewicki (personal communication) finds that the independent components of natural sound are similar to the "Gammatone" filters commonly used to model responses of neurons in the auditory nerve. Schwartz & Simoncelli (2001) have shown that divisive normalization of responses of such filters can serve as a nonlinear whitening operation for natural sounds, analogous to the case for vision. In using natural sounds as experimental stimui, Rieke et al (1995) have shown that neurons at early stages of the frog auditory system are adapted specifically to encode the structure in the natural vocalizations of the animal. Attias & Schreiner (1998) demonstrated that the rate of information transmission in cat auditory midbrain neurons is higher for naturalistic stimuli.

Overall, we feel that recent progress on exploring and testing the relationship between environmental statistics and sensation is encouraging. Results to date have served primarily as post-hoc explanations of neural function, rather than predicting aspects of sensory processing that have not yet been observed. But it is our belief that this line of research will eventually lead to new insights and will serve to guide our thinking in the exploration of higher-level visual areas.

ACKNOWLEDGMENTS

Visit the Annual Reviews home page at www.AnnualReviews.org

LITERATURE CITED

Atick JJ. 1992. Could information theory provide an ecological theory of sensory processing? *Netw. Comput. Neural Syst.* 3:213–51

Atick JJ, Li Z, Redlich AN. 1993. What does post-adaptation color appearance reveal about cortical color representation? *Vis. Res.* 33(1):123–29

Atick JJ, Redlich AN. 1991. *What does the retina know about natural scenes?* Tech.

Rep. IASSNS-HEP-91/40, Inst. Adv. Study, Princeton, NJ

Atick JJ, Redlich AN. 1992. What does the retina know about natural scenes? *Neural Comput.* 4:196–210

Attias H. 1998. Independent factor analysis. *Neural Comput.* 11:803–51

Attias H, Schreiner CE. 1997. Temporal low-order statistics of natural sounds. In

Advances in Neural Information Processing Systems, ed. MC Mozer, M Jordan, M Kearns, S Solla, 9:27–33. Cambridge, MA: MIT Press

Attias H, Schreiner CE. 1998. Coding of naturalistic stimuli by auditory midbrain neurons. In *Advances in Neural Information Processing Systems*, ed. M Jordan, M Kearns, S Solla, 10:103–9. Cambridge, MA: MIT Press.

Attneave F. 1954. Some informational aspects of visual perception. *Psychol. Rev.* 61:183–93

Baddeley R, Abbott LF, Booth MC, Sengpiel F, Freeman T, et al. 1998. Respones of neurons in primary and inferior temporal visual cortices to natural scenes. *Proc. R. Soc. London Ser. B* 264:1775–83

Balboa RM, Grzywacz NM. 2000. The role of early lateral inhibition: more than maximizing luminance information. *Vis. Res.* 17:77–89

Barlow HB. 1961. Possible principles underlying the transformation of sensory messages. In *Sensory Communication*, ed. WA Rosenblith, pp. 217–34. Cambridge, MA: MIT Press

Barlow HB, Foldiak P. 1989. Adaptation and decorrelation in the cortex. In *The Computing Neuron*, ed. R Durbin, C Miall, G Mitchinson, 4:54–72. New York: Addison-Wellesley

Bell AJ, Sejnowski TJ. 1997. The "independent components" of natural scenes are edge filters. *Vis. Res.* 37(23):3327–38

Bergen JR, Adelson EH. 1986. Visual texture segmentation based on energy measures. *J. Opt. Soc. Am. A* 3:99

Bialek W, Rieke F, de Ruyter van Steveninck RR, Warland D. 1991. Reading a neural code. *Science* 252:1854–57

Bonds AB. 1989. Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex. *Vis. Neurosci.* 2:41–55

Brenner N, Bialek W, de Ruyter van Steveninck RR. 2000. Adaptive rescaling maximizes information transmission. *Neuron* 26:695–702

Buccigrossi RW, Simoncelli EP. 1999. Image compression via joint statistical characterization in the wavelet domain. *IEEE Trans. Image Proc.* 8(12):1688–701

Buchsbaum G, Gottschalk A. 1983. Trichromacy, opponent color coding, and optimum colour information transmission in the retina. *Proc. R. Soc. London Ser. B* 220:89–113

Buchsbaum G, Gottschalk A. 1984. Chromaticity coordinates of frequency-limited functions. *J. Opt. Soc. Am. A* 1(8):885–87

Carandini M, Heeger DJ, Movshon JA. 1997. Linearity and normalization in simple cells of the macaque primary visual cortex. *J. Neurosci.* 17:8621–44

Carandini M, Movshon JA, Ferster D. 1998. Pattern adaptation and cross-orientation interactions in the primary visual cortex. *Neuropharmacology* 37:501–11

Cardoso JF. 1989. Source separation using higer order moments. In *Int. Conf. Acoustics Speech Signal Proc.*, pp. 2109–12. IEEE Signal Process. Soc.

Common P. 1994. Independent component analysis, a new concept? *Signal Process* 36:387–14

Dan Y, Atick JJ, Reid RC. 1996. Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. *J. Neurosci.* 16:3351–62

Daugman JG. 1989. Entropy reduction and decorrelation in visual coding by oriented neural receptive fields. *IEEE Trans. Biomed. Eng.* 36(1):107–14

Dayan P, Hinton GE, Neal RM, Zemel RS. 1995. The Helmholtz machine. *Neural Comput.* 7:889–904

Dong DW. 1995. Associative decorrelation dynamics: a theory of self-organization and optimization in feedback networks. In *Advances in Neural Information Processing Systems*, ed. G Tesauro, D Touretzky, T Leen. 7:925–32

Dong DW, Atick JJ. 1995a. Statistics of natural time-varying images. *Netw. Comput. Neural Syst.* 6:345–58

Dong DW, Atick JJ. 1995b. Temporal decorrelation: a theory of lagged and nonlagged

responses in the lateral geniculate nucleus. *Netw. Comput. Neural Syst.* 6:159–78

Dragoi V, Sharma J, Sur M. 2000. Adaptation-induced plasticity of orientation tuning in adult visual cortex. *Neuron* 28:287–88

Field DJ. 1987. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A* 4(12):2379–94

Field DJ. 1994. What is the goal of sensory coding? *Neural Comput.* 6:559–601

Foldiak P. 1990. Forming sparse representations by local anti-hebbian learning. *Biol. Cybernet.* 64:165–70

Geisler WS, Albrecht DG. 1992. Cortical neurons: isolation of contrast gain control. *Vis. Res.* 8:1409–10

Geisler WS, Perry JS, Super BJ, Gallogly DP. 2001. Edge co-occurance in natural images predicts contour grouping performance. *Vis. Res.* 41:711–24

Hancock PJB, Baddeley RJ, Smith LS. 1992. The principal components of natural images. *Network* 3:61–72

Heeger D, Bergen J. 1995. Pyramid-based texture analysis/synthesis. In *Proc. Assoc. Comput. Mach. Special Interest Groups Graph*, pp. 229–38

Heeger DJ. 1992. Normalization of cell responses in cat striate cortex. *Vis. Neurosci.* 9:181–98

Hyvärinen A, Hoyer P. 2000. Emergence of topography and complex cell properties from natural images using extensions of ica. In *Advances in Neural Information Processing Systems*, ed. SA Solla, TK Leen, K-R Müller, 12:827–33, Cambridge, MA: MIT Press

Hyvärinen A, Oja E. 1997. A fast fixed-point algorithm for independent component analysis. *Neural Comput.* 9:1483–92

Jaynes ET. 1978. Where do we stand on maximum entropy? In *The Maximal Entropy Formalism*, ed. RD Levine, M Tribus, pp. 620–30. Cambridge, MA: MIT Press

Julesz B. 1962. Visual pattern discrimination. *IRE Trans. Inf. Theory*, IT-8

Jutten C, Herault J. 1991. Blind separation of sources. Part I: An adaptive algorithm based on neuromimetic architecture. *Signal Process* 24(1):1–10

Kersten D. 1987. Predictability and redundancy of natural images. *J. Opt. Soc. Am. A* 4(12):2395–400

Knill DC, Field D, Kersten D. 1990. Human discrimination of fractal images. *J. Opt. Soc. Am. A* 7:1113–23

Knill DC, Richards W, eds. 1996. *Perception as Bayesian Inference*. Cambridge, UK: Cambridge Univ. Press

Laughlin SB. 1981. A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch.* 36C:910–12

Lee AB, Mumford D. 1999. An occlusion model generating scale-invariant images. In *IEEE Workshop on Statistical and Computational Theories of Vision*, Fort Collins, CO. Also at http://www.cis.ohiostate.edu/~szhu/SCTV99.html

Lennie P, D'Zmura M. 1988. Mechanisms of color vision. *CRC Crit. Rev. Neurobiol.* 3:333–400

Lewicki MS, Olshausen BA. 1999. Probabilistic framework for the adaptation and comparison of image codes. *J. Opt. Soc. Am. A* 16(7):1587–601

Lewicki M, Sejnowski T. 1999. Coding time-varying signals using sparse, shift-invariant representations. In *Advances in Neural Information Processing Systems*, ed. MS Kearns, SA Solla, DA Cohn, 11:815–21. Cambridge, MA: MIT Press

Maloney LT. 1986. Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *J. Opt. Soc. Am. A* 3(10):1673–83

Müller JR, Metha AB, Krauskopf J, Lennie P. 1999. Rapid adaptation in visual cortex to the structure of images. *Science* 285:1405–8

Olshausen BA. 2001. Sparse codes and spikes. In *Statistical Theories of the Brain*, ed. R Rao, B Olshausen, M Lewicki. Cambridge, MA: MIT Press. In press

Olshausen BA, Field DJ. 1996. Emergence of simple-cell receptive field properties by

learning a sparse code for natural images. *Nature* 381:607–9

Olshausen BA, Field DJ. 1997. Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vis. Res.* 37:3311–25

Oram MW, Foldiak P, Perrett DI, Sengpiel F. 1998. The "ideal homunculus": decoding neural population signals. *Trends Neurosci.* 21(6):259–65

Parraga CA, Troscianko T, Tolhurst DJ. 2000. The human visual system is optimised for processing the spatial information in natural visual images. *Curr. Biol.* 10:35–38

Penev P, Gegiu M, Kaplan E. 2000. Fast convergent factorial learning of the low-dimensional independent manifolds in optical imaging data. In *Proc. 2nd Int. Workshop Indep. Comp. Anal. Signal Separation*, pp. 133–38. Helsinki, Finland

Portilla J, Simoncelli EP. 2000. A parametric texture model based on joint statistics of complex wavelet coefficients. *Int. J. Comput. Vis.* 40(1):49–71

Rao RPN, Ballard DH. 1997. Dynamic model of visual recognition predicts neural response properties in the visual cortex. *Neural Comput.* 9:721–63

Reichhardt W, Poggio T. 1979. Figure-ground discrimination by relative movement in the visual system of the fly. *Biol. Cybernet.* 35:81–100

Reinagel P, Zador AM. 1999. Natural scene statistics at the centre of gaze. *Netw. Comput. Neural Syst.* 10:341–50

Rieke F, Bodnar DA, Bialek W. 1995. Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. *Proc. R. Soc. London B* 262:259–65

Ruderman DL. 1997. Origins of scaling in natural images. *Vis. Res.* 37:3385–98

Ruderman DL, Bialek W. 1994. Statistics of natural images: scaling in the woods. *Phys. Rev. Lett.* 73(6):814–17

Ruderman DL, Cronin TW, Chiao CC. 1998. Statistics of cone responses to natural images: implications for visual coding. *J. Opt. Soc. Am. A* 15(8):2036–45

Sanger TD. 1989. Optimal unsupervised learning in a single-layer network. *Neural Netw.* 2:459–73

Schwartz O, Simoncelli E. 2001. Natural sound statistics and divisive normalization in the auditory system. In *Advances in Neural Information Processing Systems*, ed. TK Leen, TG Dietterich, V Tresp, Vol. 13. Cambridge, MA: MIT Press. In Press

Shannon C. 1948. The mathematical theory of communication. *Bell Syst. Tech. J.* 27:379–423

Shouval H, Intrator N, Cooper LN. 1997. BCM Network develops orientation selectivity and ocular dominance in natural scene environment. *Vis. Res.* 37(23):3339–42

Sigman M, Cecchi GA, Gilbert CD, Magnasco MO. 2001. On a common circle: natural scenes and gestalt rules. *Proc. Natl. Acad. Sci.* 98(4):1935–40

Simoncelli EP. 1997. *Statistical Models for Images: Compression, Restoration and Synthesis*. Asilomar Conf. Signals, Systems, Comput. 673–78. Los Alamitos, CA: IEEE Comput. Soc. http://www.cns.nyu.edu/~eero/publications.html

Simoncelli EP, Schwartz O. 1999. Image statistics and cortical normalization models. In *Advances in Neural Information Processing Systems*, ed. MS Kearns, SA Solla, DA Cohn. 11:153–59

Smirnakis SM, Berry MJ, Warland DK, Bialek W, Meister M. 1997. Adaptation of retinal processing to image contrast and spatial scale. *Nature* 386:69–73

Srinivasan MV, Laughlin SB, Dubs A. 1982. Predictive coding: A fresh view of inhibition in the retina. *J. R. Soc. London Ser. B* 216:427–59

van Hateren JH. 1992. A theory of maximizing sensory information. *Biol. Cybern.* 68:23–29

van Hateren JH. 1993. Spatiotemporal contrast sensitivity of early vision. *Vis. Res.* 33:257–67

van Hateren JH, van der Schaaf A. 1998. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. R. Soc. London Ser. B* 265:359–66

Vinje WE, Gallant JL. 2000. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* 287:1273–76

Wainwright MJ. 1999. Visual adaptation as optimal information transmission. *Vis. Res.* 39:3960–74

Wainwright MJ, Schwartz O, Simoncelli EP. 2001. Natural image statistics and divisive normalization: modeling nonlinearity and adaptation in cortical neurons. In *Statistical Theories of the Brain*, ed. R Rao, B Olshausen, M Lewicki. Cambridge, MA: MIT Press. In press

Webster MA. 1996. Human colour perception and its adaptation. *Netw. Comput. Neural Syst.* 7:587–634

Wegmann B, Zetzsche C. 1990. Statistical dependence between orientation filter outputs used in an human vision based image code. In *Proc. SPIE Vis. Commun. Image Processing*, 1360:909–22. Lausanne, Switzerland: Soc. Photo-Opt. Instrum. Eng.

Zhu SC, Wu YN, Mumford D. 1998. FRAME: Filters, random fields and maximum entropy—towards a unified theory for texture modeling. *Int. J. Comp. Vis.* 27(2):1–20

Annual Review of Neuroscience
Volume 24, 2001

CONTENTS