Optimal stimulus encoders for natural tasks

Wilson S. Geisler Center for Perceptual Systems and Department of Psychology, University of Texas at Austin, Austin, TX, USA

Jiri Najemnik Center for Perceptual Systems and Department of Psychology, University of Texas at Austin, Austin, TX, USA

Almon D. Ing Center for Perceptual Systems and Department of Psychology, University of Texas at Austin, Austin, TX, USA

Determining the features of natural stimuli that are most useful for specific natural tasks is critical for understanding perceptual systems. A new approach is described that involves finding the optimal encoder for the natural task of interest, given a relatively small population of noisy "neurons" between the encoder and decoder. The optimal encoder, which necessarily specifies the most useful features, is found by maximizing accuracy in the natural task, where the decoder is the Bayesian ideal observer operating on the population responses. The approach is illustrated for a patch identification task, where the goal is to identify patches of natural image, and for a foreground identification task, where the goal is to identify which side of a natural surface boundary belongs to the foreground object. The optimal features (receptive fields) are intuitive and perform well in the two tasks. The approach also provides insight into general principles of neural encoding and decoding.

Keywords: structure of natural images, computational modeling, receptive fields, perceptual organization, categorization

Citation: Geisler, W. S., Najemnik, J., & Ing, A. D. (2009). Optimal stimulus encoders for natural tasks. *Journal of Vision, 9*(13):17, 1–16, http://journalofvision.org/9/13/17/, doi:10.1167/9.13.17.

Introduction

Evolution tends to select perceptual systems that encode, or learn to encode, those properties of the environment that are relevant for successful performance of the organism's natural tasks or behaviors. Thus, the systematic study of a perceptual system requires characterizing the task-relevant properties of environments and sensory stimuli, as well as determining how those properties are exploited by the nervous system to perform natural tasks (for reviews see Geisler, 2008; Simoncelli & Olshausen, 2001).

The crucial first step of this enterprise, characterizing natural stimuli, involves measuring and analyzing natural scene statistics. There have been two general approaches, which have both been productive. One involves collecting natural stimuli, determining some of their statistical structure using mathematical tools such as principle components analysis, independent components analysis and information theory, and then interpreting that structure with respect to the principle of efficient coding (Bell & Sejnowski, 1997; Laughlin, 1981; Lee, Pedersen, & Mumford, 2003; Olshausen & Field, 1997; Ruderman & Bialek, 1994; Simoncelli & Olshausen, 2001; Smith & Lewicki, 2006; van Hateren & van der Schaaf, 1998). The other approach is similar but focuses on specific natural tasks by attempting to characterize the statistical relationship between specific properties of sensory stimuli and specific environmental (scene) properties relevant for efficient performance in a given task (Brunswik & Kamiya, 1953; Elder & Goldberg, 2002; Fowlkes, Martin, & Malik, 2007; Geisler, 2008; Geisler & Perry, 2009; Geisler, Perry, Super, & Gallogly, 2001; Konishi, Yuille, Coughlan, & Zhu, 2003; Martin, Fowlkes, & Malik, 2004; Motoyoshi, Nishida1, Sharan, & Adelson, 2007; Ullman, 2007; Ullman, Vidal-Naquet, & Sali, 2002). A weakness of the former approach is that it provides little insight into which specific statistical properties of natural stimuli are relevant to which specific natural tasks; in fact, some stimulus properties may not be relevant to any task performed by a given organism. A weakness of the latter approach has been that the analyzed stimulus properties are often selected on the basis of intuition, historical precedence, and trial-and-error, rather than on the basis of a principled and unbiased procedure (but see Ullman et al., 2002). Here we describe a principled and unbiased procedure for determining the most relevant stimulus properties for specific natural tasks and for quantifying the usefulness of those properties.

The proposed analysis builds on previous approaches for measuring natural scene statistics and on applications of Bayesian ideal observer theory and information theory to neural coding. The central idea is to determine the optimal encoder for the natural task of interest, assuming a small number of "neurons" (a limited channel) with given

doi: 10.1167/9.13.17 Received December 22, 2008; published December 16, 2009 ISSN 1534-7362

constraints between the encoder and its matched Bayesian optimal decoder: The optimal encoder is the one that yields the best performance when combined with its matched optimal decoder (which generally will be different for each candidate encoder). There are three reasons for considering optimal encoding for a limited neural channel. First, if the optimal encoder can be determined, then the encoding properties of the small neural population necessarily represent the most relevant stimulus properties for the given task. Second, there are always limitations in neural resources and the optimal encoder shows how to best use those limited resources for the given task. Third, in addition to providing a rigorous characterization of the natural scene statistics for the given task, determining the optimal encoder provides a principled starting point for proposing and testing hypotheses for the actual neural encoding and decoding.

Many natural tasks involve making categorical judgments about the environment given the proximal stimuli encoded by the sensory receptors (e.g., identifying physical objects or materials and their locations in space and time). This paper focuses on such identification tasks where the goal is to maximize identification accuracy. In this case, the output of the proposed analysis is a set of optimal receptive fields that capture and hence represent the stimulus properties most relevant for performing the specific natural identification task.

In addition to incorporating the natural scene statistics, the proposed analysis explicitly incorporates constraints on the dynamic range and noise properties of the neurons making up the limited channel between the encoder and the decoder. Here, the neurons making up the limited channel have a maximum response that is typical of cortical neurons, and like cortical neurons, they have a response variance proportional to the mean response.

After describing the proposed analysis in general terms, it is illustrated by finding optimal linear receptive fields for two natural identification tasks: an image-patch identification task and a foreground identification task.

Methods

Accuracy maximization analysis (AMA)

In a given natural identification task the organism receives a particular sensory stimulus and based upon the responses of a neural population to the stimulus makes a decision about which specific category of object is present in the environment (Figure 1). To be concrete, we represent the specific category of object by a vector $\boldsymbol{\omega}$ that can take one of a discrete number of possible values, indexed by the integer variable k, and we represent the received stimulus by a vector $\mathbf{s}$ (e.g., a patch of image or a sample of sound) that also can take one of a discrete number of possible values, indexed by the pair of integer variables (k, l), where l is the specific exemplar from category k. Thus, the natural scene statistics for the identification task can be represented by a joint probability distribution, $p_0(k, l)$, and any given randomly sampled stimulus in the task can be regarded as a random sample (K, L) from this distribution. It is the task-relevant structure of this unknown joint probability distribution that we wish to characterize.

To characterize the structure of $p_0(k, l)$, we suppose stimuli sampled from this distribution are encoded in the responses of a population of q neurons. The responses to a stimulus $\mathbf{s}(k, l)$ can be represented by a random vector, $\mathbf{R}_q(k, l) = [R_1(k, l),\ldots, R_q(k, l)]$, and the observer's guess of the category (i.e., the observer's response) based on these random responses can be represented by $\hat{\boldsymbol{\omega}}[\mathbf{R}_q(k, l)]$. The mean response functions, $\mathbf{r}_q(k, l) = [r_1(k, l),\ldots, r_q(k, l)]$, describe the mapping between the stimulus and the mean response of each neuron in the population and can be regarded as the encoding functions. For example, each encoding function might be defined by a unique receptive field or tuning function. The aim of AMA is to find encoding functions that maximize identification performance in a specific categorization task.

To find the optimal encoding functions, it is both essential and useful to explicitly represent neural noise. First, if the mean response functions are specified as real-valued functions, then some noise must be included to obtain meaningful answers (solutions degenerate if responses have arbitrarily high precision because each stimulus produces a unique number). Second, all real systems include some form of noise; for example, even in a perfect digital visual system the quantization of gray level effectively introduces uniform noise within each quantization step. Third, an explicit representation of the neural noise allows one to investigate the effect of realistic neural noise properties on optimal encoding. We assume that the variability of each neuron (to repeated presentations of the same stimulus) is determined by a random sample of neural noise $\mathbf{N}_q = [N_1,\ldots, N_q]$ that may be correlated across neurons (Gawne & Richmond, 1993; Zohary, Shadlen, & Newsome, 1994) and may depend on the mean response of the neuron (e.g., in cortical neurons the variance of the response is proportional to the mean response, Geisler & Albrecht, 1997; Tolhurst, Movshon, & Dean, 1983).

A complete description of an identification task must include a specification of the costs and benefits (utility) of different task outcomes. Here we consider tasks where the utility of all errors is zero and the utility of all corrects is one (i.e., the goal is to maximize accuracy), but it is possible to generalize to arbitrary utility functions (see Discussion).

The central question addressed here is this: Given a natural identification task, and a set of constraints on the

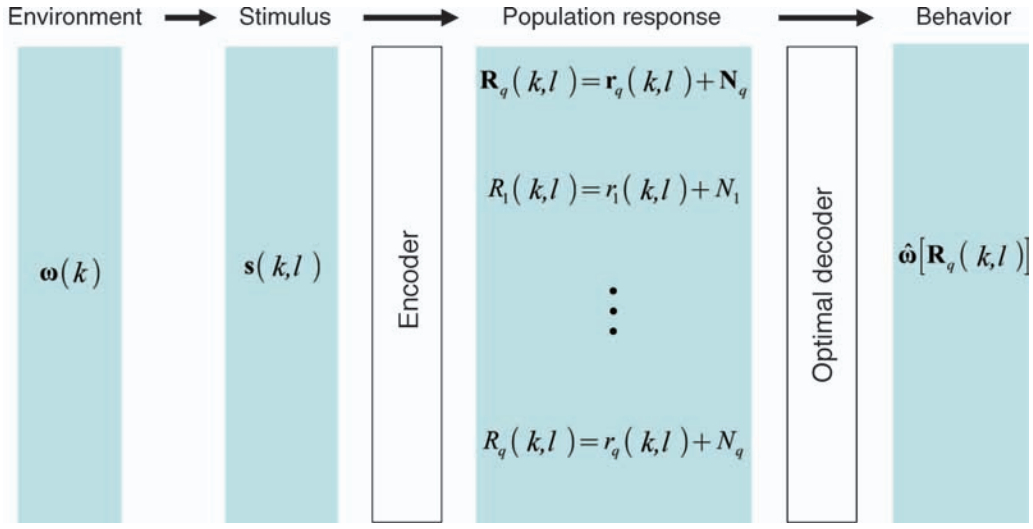

Figure 1. Framework for characterizing natural scene statistics for specific identification tasks. The category of object in the environment is represented by a vector $\boldsymbol{\omega}(k)$, indexed by category number k. The proximal stimulus is represented by a vector $\mathbf{s}(k, l)$, indexed by category number k and exemplar number l. Thus, a randomly sampled natural stimulus in the natural task can be regarded as random sample (K, L) from an unknown joint probability distribution, $p_0(k, l)$. The proximal stimulus is encoded by a population of q neurons where the mean response of the population to a particular stimulus $\mathbf{s}(k, l)$ is $\mathbf{r}_q(k, l) = [r_1(k, l),\ldots, r_q(k, l)]$, and the variability of each neuron's response (to repeated presentations of the same stimulus) is represented by an additive sample of noise $\mathbf{N}_q = [N_1,\ldots, N_q]$ whose variance may depend upon the mean response and that may be correlated across neurons. The population response is optimally decoded into an estimate $\hat{\boldsymbol{\omega}}$ of the object category in the environment (the distal stimulus). The goal of accuracy maximization analysis is to determine the encoding functions $[r_1(k, l),\ldots, r_q(k, l)]$ that maximize accuracy in the identification task. (Bold letters represent vector quantities, capital letters represent random variables.)

population responses, what is the optimal mapping between the stimuli and the mean responses of the neurons in the population? In other words, we wish to determine the encoding functions, $[r_1(k, l),\ldots, r_q(k, l)]$, that maximize accuracy in the identification task. To do this, we must consider the ideal observer whose input is the neural population response. The decision rule of the ideal observer is the optimal decoder. If the goal is to maximize accuracy, then the ideal decision rule is to pick the category that is most likely given the observed neural population response (i.e., the category with the greatest posterior probability):

$$\text{Pick category } i \text{ if } p\left(i|\mathbf{R}_q\right) > p\left(j|\mathbf{R}_q\right) \text{ for } j \neq i. \qquad (1)$$

The encoding functions that maximize the performance of this ideal observer are the optimal encoding functions.

The direct procedure for determining the optimal encoding functions would be to search the space of possible encoding functions by simulating the performance of the ideal observer for each set of candidate functions. Unfortunately, this direct procedure is generally impractical. The more practical heuristic approach taken here (which we verify with Monte Carlo simulation) is to consider the shape of the posterior probability distribution across categories computed by the ideal observer. Consider an arbitrary stimulus $\mathbf{s}(k, l)$. The posterior probability distribution computed by the ideal observer from the population response to this stimulus is $p(i\,|\,\mathbf{R}_q(k, l))$. This posterior probability distribution varies randomly because of the randomness of the neural population responses to the same stimulus, and thus the ideal observer will be accurate if this posterior probability distribution is on average as close as possible to a probability distribution $f(i)$ that is 1 at the correct category k and is 0 elsewhere. A principled measure of the difference between two probability distributions $f(x)$ and $g(x)$ is the *relative entropy, D,* also known as the Kullback–Leibler divergence (Cover & Thomas, 2006):

$$D = \sum_x f(x)\log\frac{f(x)}{g(x)}.$$

Relative entropy plays a special role in information theory by providing a precise measure of the uncertainty difference (in bits) between two probability distributions. In our case, the average relative entropy reduces to the simple formula:

$$D_q(k,l) = -E\left[\log p\left(k|\mathbf{R}_q(k,l)\right)\right]. \qquad (2)$$

In agreement with what one would expect of an appropriate measure we note that $D_q(k, l)$ decreases toward zero monotonically as the posterior probability at the correct category approaches 1.0. It follows intuitively that the overall accuracy of the ideal observer will be maximized when the average difference $\bar{D}_q$ over all possible stimuli is minimized, where

$$\bar{D}_q = -\sum_{k,l} p_0(k,l) E\left[\log p\left(k|\mathbf{R}_q(k,l)\right)\right]. \quad (3)$$

An obstacle to directly minimizing Equation 3 is that the expected log posterior probability can generally be computed only by Monte Carlo simulation, which can be prohibitively slow and noisy when searching a large space of possible encoding functions. Therefore, an approximation is needed. Here, we assume that the average relative entropy, given a random response vector, is approximately equal to the relative entropy, given the average value of the response vector:

$$\begin{aligned} -E\left[\log p\left(k|\mathbf{R}_q(k,l)\right)\right] &\cong -\log p\left(k|E\left[\mathbf{R}_q(k,l)\right]\right) \\ &= -\log p\left(k|\mathbf{r}_q(k,l)\right). \end{aligned} \quad (4)$$

There are surely better approximations of expected relative entropy than Equation 4, but we show in the examples that this approximation appears to be sufficiently accurate for the present purpose of determining optimal encoding functions.

Finally, note that Equation 3 can be regarded as the mean value of the expected relative entropy over the prior probability distribution of categories and stimuli, and thus given a large enough training set of random samples (K_i, L_i) of natural stimuli, the optimal encoding functions can be obtained by minimizing the sample mean:

$$\bar{D}_q = -\frac{1}{n}\sum_{i=1}^{n} \log p\left(K_i|\mathbf{r}_q(K_i,L_i)\right). \quad (5)$$

Thus, a practical procedure for estimating optimal encoding functions is to minimize Equation 5 over the space of encoding functions, for a large number of random samples of the natural stimuli that arise in the natural identification task. Note that minimizing Equation 5 is equivalent to maximizing the geometric mean of the posterior probability at the correct category across the training samples. An obvious alternative is to maximize the arithmetic mean (i.e., drop the logarithm from Equation 5). This behaves similarly, but we have found slightly better convergence behavior in some cases using the geometric mean. This completes the general derivation of accuracy maximization analysis.

Optimal linear receptive fields and Gaussian neural noise

In any specific application of accuracy maximization analysis the task and the procedure of selecting training stimuli must be specified, and some constraint must be placed on the family of possible neural encoding functions. For the examples described here, the input stimuli are normalized gray-scale image patches (12 × 12 pixels), and the observer's task is to indicate the category to which the image patch belongs. In other words, the input stimulus is given by

$$\mathbf{s}(k,l) = \frac{\mathbf{x}(k,l) - \bar{x}(k,l)}{sd(k,l)\sqrt{n_{pixels}}}, \quad (6)$$

where $\mathbf{x}(k, l)$ is the un-normalized image patch, $\bar{x}(k, l)$ is the mean gray level of the patch, $sd(k, l)$ is standard deviation gray level of the patch, and n_{pixels} is the number of pixels in the patch.

The family of possible encoding functions is constrained in three ways: (a) we consider only linear weighting functions (linear receptive fields), (b) the response of each neuron cannot exceed a value of $r_{\max}$, and (c) the neural noise is independent Gaussian with variance proportional to the mean response. The first two constraints are implemented by the following equation for the mean response of the t^{th} neuron:

$$r_t(k,l) = r_{\max}\mathbf{s}(k,l) \cdot \mathbf{w}_t, \quad (7)$$

where $\mathbf{w}_t$ is a vector of weights (normalized to a length of 1.0) that defines the 12 × 12 receptive field, and $\mathbf{s}(k, l) \cdot \mathbf{w}_t$ is the dot product of the stimulus with the receptive field. Because the stimulus and receptive field (RF) are both normalized to a vector length of 1.0, their maximum dot product is 1.0 (when the receptive field matches the stimulus) and hence the maximum possible response is $r_{\max}$. The third constraint is implemented by requiring that the probability of response r from the t^{th} neuron in the population is given by:

$$p(r|k,l) = \frac{1}{\sqrt{2\pi}\sigma_t(k,l)} \exp\left[-\frac{1}{2}\frac{[r - r_t(k,l)]^2}{\sigma_t^2(k,l)}\right], \quad (8)$$

$$\sigma_t^2(k,l) = \alpha|r_t(k,l)| + \sigma_0^2, \quad (9)$$

where α is the Fano factor and σ_0^2 is the small baseline variability.[1] The dynamic range and noise parameters of the neurons were set to the mean values for neurons in monkey V1 reported in Geisler and Albrecht (1997), for 200 ms (fixation-like) stimulus presentations: $r_{max} = 5.7$ spks, $\alpha = 1.36$, $\sigma_0^2 = 0.23$.

Under the above assumptions, and expanding $p(k|\mathbf{r}_q(k, l))$ into a recursive formula using Bayes' rule (see Appendix A), we have:

$$p\left(k|\mathbf{r}_q(k,l)\right) = \frac{1}{Z} p\left(k|\mathbf{r}_{q-1}(k,l)\right) \frac{1}{n_k} \sum_{j=1}^{n_k} \frac{1}{\sigma_q(k,j)} \exp\left[-\frac{1}{2} \frac{\left[r_q(k,l) - r_q(k,j)\right]^2}{\sigma_q(k,j)^2}\right], \quad (10)$$

where n_k is the number of training samples from category k, $p(k|\mathbf{r}_0(k, l)) = n_k/n$, n is the total number of training stimuli, and Z is a normalization factor. In keeping with the approximation in Equation 4, the logarithm of this formula gives the average relative entropy when the stimulus is $\mathbf{s}(k, l)$. Thus, Equations 5–10 provide a closed-form expression for the average relative entropy of the posterior probability distribution (that the ideal observer computes) for arbitrary samples from the joint probability distribution of environmental categories and associated stimuli, $p_0(k, l)$.

To estimate the optimal linear receptive fields we use a 'greedy' procedure. In other words, neurons are added to the population one at a time, with each neuron's receptive field being selected to produce the biggest decrease in decoding error. Specifically, we proceed sequentially by first finding the encoding function $r_1(k, l)$ that minimizes $\bar{D}_1$ (see Equation 5); then we substitute the resulting expected posterior probability distribution $p(i|\mathbf{r}_1(k, l))$ into Equation 10 and find the encoding function $r_2(k, l)$ that minimizes $\bar{D}_2$; then we substitute the resulting expected posterior probability distribution $p(i|\mathbf{r}_2(k, l))$ into Equation 10, and so on. A consequence of this procedure is that the neural encoding functions tend to be rank ordered in how much they reduce the relative entropy, with the first function producing the largest decrease. (It is possible that a simultaneous rather than greedy procedure could lead to better performance, but we have not yet explored this more computationally intense approach.)

When each neuron is added we first initialize its weights, and then perform gradient descent. Sometimes we initialize the weights randomly. Alternatively, to reduce the likelihood of getting trapped in local minima, we also use an initializing step similar to that of Ullman et al. (2002). Specifically, we first try out each normalized image patch in the training set as an RF, and then set the initial weights to be the normalized image patch that yields the greatest reduction in relative entropy. The advantage of this "stimulus-manifold-sampling" approach is that it randomly samples starting weights that lie in the relatively small-dimensional manifold of the natural stimuli.

Results

Image patch identification task

In this task the goal is to accurately identify specific image patches randomly sampled from gray-scale natural images. In other words, we ask what encoding functions (receptive fields) would be optimal for identifying arbitrary specific patches of natural image. The image patches were 12 × 12 pixels in size and randomly sampled from calibrated gray-scale images containing no human-made objects (van Hateren & van der Schaaf, 1998). Example image patches are shown in Figure 2a.

Figure 2b shows the estimated receptive fields of the first six neurons (starting with random weights), for the training set of 200 randomly selected natural image patches shown in Figure 2a. As can be seen, the receptive fields are relatively low spatial-frequency (smooth) patterns that are similar to oriented edges and bars. It is interesting to note that these first six receptive fields are nearly orthogonal, even though no orthogonality constraint was imposed; Table 1 shows all the pair-wise spatial correlations between the receptive fields. In general, one expects a mixture of orthogonal and non-orthogonal receptive fields depending on the properties of the stimuli and on the reliability of the individual neurons (see second example).

We have estimated optimal receptive fields for a number of different random samples of 200 image patches and they are generally similar in appearance to those in Figure 2, but sometimes with a slightly different ordering. Similarly the ordering of the receptive fields varies somewhat depending on the initial weights, and on whether they are selected at random or by stimulus-manifold-sampling. This is to be expected in this case, because each of these optimal receptive fields reduces the expected relative entropy by a similar amount (especially the first four receptive fields), and hence each receptive field is about equally useful. As shown in Figure 3a, the relative entropy declines approximately linearly with a slope of 0.92 bits per neuron, from its initial value of 7.64 bits ($\log_2 200$). The value of the slope depends in part upon the dynamic range and reliability of the neurons; the larger dynamic range and the lower the neural noise level, the steeper the slope.

An important issue is whether these estimated receptive fields are actually the ones that optimally reduce the average relative entropy of the posterior probability distributions computed by the ideal observer. Recall that

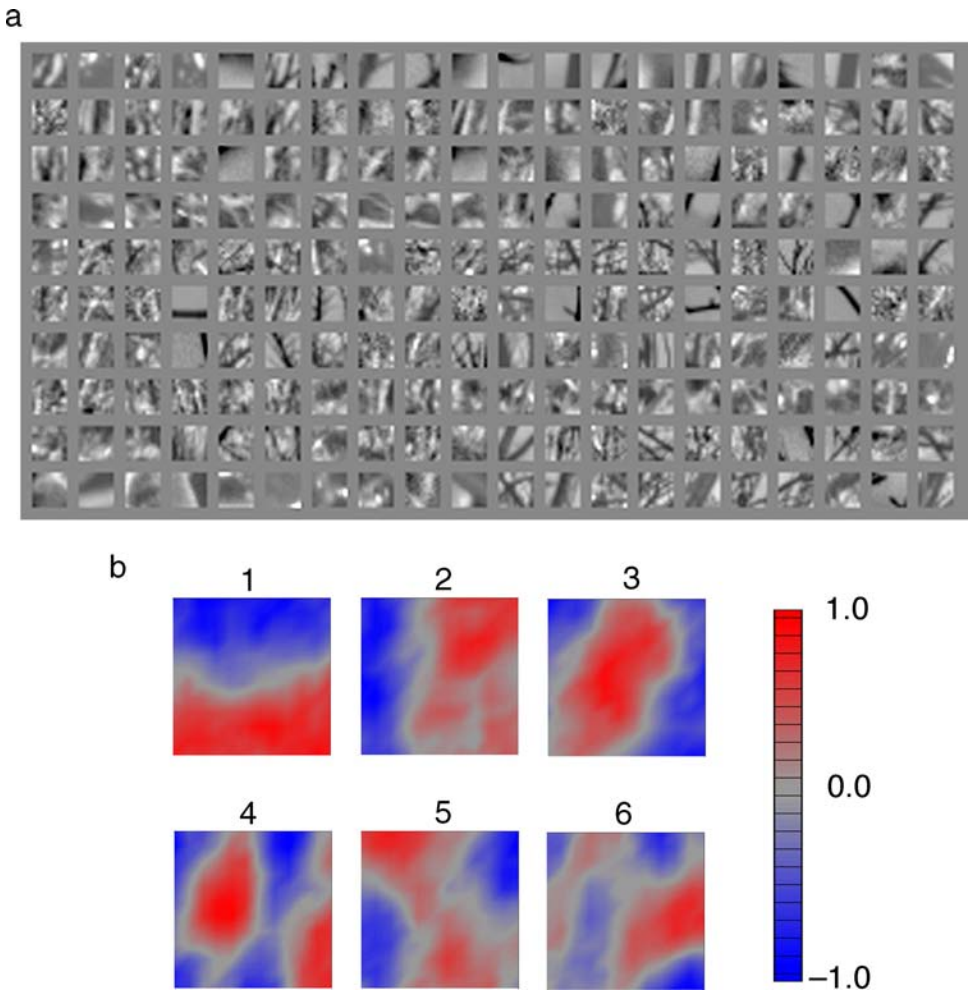

Figure 2. Optimal linear receptive fields for a natural image patch identification task. a. Example set of 200 training patches randomly sampled from calibrated natural images. b. The final weights for the first six optimal linear RFs obtained by gradient descent from random weights. (For display purposes the receptive fields have been scaled so that the maximum absolute value is 1.0, and then interpolated by the plotting software.)

to make the estimation procedure tractable we used the potentially questionable approximation in Equation 4. To address this issue we directly computed the average relative entropy in Equation 3 using Monte Carlo techniques, for the optimal receptive fields in Figure 2 (see Equation A3 in the Appendix). Specifically, for each image patch we determined the actual relative entropy by Monte Carlo simulation (1000 trials) and compared it with the estimated relative entropy based on Equation 4. Figure 3b plots the correlation between the actual and estimated relative entropy as a function of the number of receptive fields; the correlations are relatively high, suggesting that the approximation in Equation 4 is adequate for estimating optimal neural encoding functions. Importantly, as shown in Figure 3c, there is an approximately linear relationship between the average actual relative entropy and the average estimated relative entropy predicted by Equation 4, although the slopes decrease and the intercepts increase as each new receptive field is added. If the relationship were nonlinear, then the approximation would be biased depending on how difficult the image patches (training samples) are to identify. Notice that this figure suggests that the approximation in Equation 4 systematically underestimates the relative entropy.

How well does the optimal decoder, using the optimal receptive fields, actually perform in the patch identification

Spatial Correlations of Receptive Fields

RF2	RF3	RF4	RF5	RF6	
−0.04	−0.10	0.17	−0.09	0.08	**RF1**
	0.10	0.03	−0.04	0.00	**RF2**
		0.08	−0.01	−0.02	**RF3**
			0.01	0.04	**RF4**
				−0.04	**RF5**

Table 1. Spatial correlations between the six optimal receptive fields in Figure 2b.

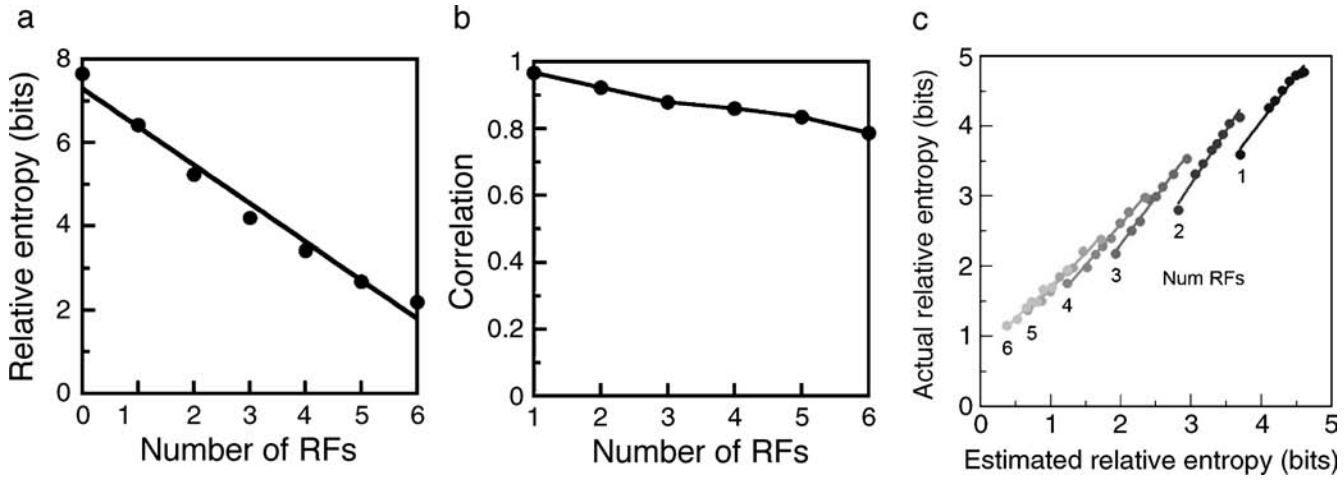

Figure 3. Evaluation of the approximation of the relative entropy of the average posterior probability distribution across categories computed by the ideal observer (the optimal decoder). a. The average relative entropy (across patches), determined by Monte Carlo simulation, as function of the number of receptive fields. b. The correlation between actual (simulated) and estimated relative entropy (using Equation 4) as a function of the number of receptive fields in the population. The correlations were computed over image patches in the training set. c. Average actual relative entropy as a function of average estimated relative entropy (using Equation 4) for different numbers of receptive fields in the population. The points show averages for the image patches in 8 quantiles.

task? Figure 4a plots the accuracy of the optimal decoder as each optimal receptive field is added in. Each data point was obtained by Monte Carlo simulation of 200 trials for each of the 200 image patches. As can be seen, six neurons with the noise characteristics of typical cortical neurons are sufficient to identify image patches with 37% accuracy where chance is 0.5%. Similar performance (well over 30% correct) is obtained for randomly selected test patches not in the training set, suggesting that the receptive fields in Figure 2 are robust and not the result of over-fitting. Figure 4b shows that the actual accuracy (for all 6 RFs and for 200 test patches not from the training set) is strongly correlated with the estimated relative entropy, thus providing further evidence for the adequacy of Equation 4.

Foreground identification task

Natural images are generally filled with occlusions where the surface of one object blocks those of other objects. Thus, a common natural task is to determine which side of an object boundary contour corresponds to the foreground surface and which side to the background.

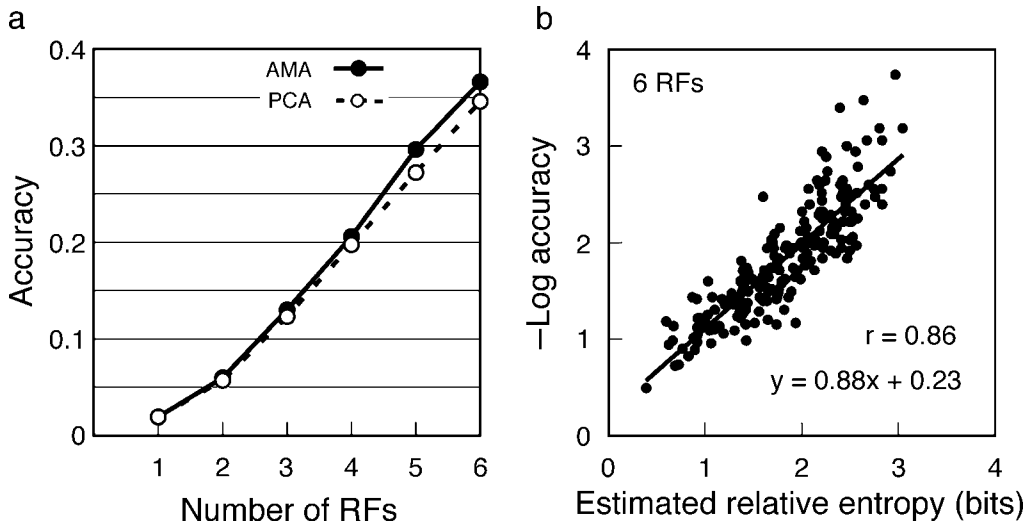

Figure 4. Actual accuracy of the optimal decoder in the patch identification task, as determined by Monte Carlo simulation (i.e., applying Equation 1, trial by trial). a. Actual accuracy as a function of the number of optimal receptive fields. b. Actual accuracy vs. estimated relative entropy, for 200 test patches not in the training set.

In this example, we consider the task of identifying which side of a contour in 12 × 12 image patches of natural foliage is the foreground surface. Unlike the previous example, there are just two categories and a large number of sample stimuli from each category.

To determine the optimal linear receptive fields for this task we make use of our database of calibrated images of close-up foliage (Geisler, Perry, & Ing, 2008). We chose to analyze close-up foliage because that is the dominant natural environment for macaque monkeys (the primary animal model for human vision) and a major component of the natural environment for many other species. Briefly, these images were obtained with a 36-bit (12 bits per color) camera calibrated to give the 12-bit luminance at each pixel location. The images were hand-segmented by human observers into objects (i.e., leafs and branches), which provided a large number of sample surface boundary contours. An example of a segmented image is shown Figure 5. The database contains over 1,600 segmented objects from a wide range of foliage images collected under various lighting conditions. Training image patches were obtained by randomly selecting surface boundary points from the 1,600 segmented objects. For each image patch the segmentation gives the orientation of the surface boundary contour at the center of the patch and the side of the contour that corresponds to the foreground object surface.

Figure 6a shows 200 randomly sampled 12 × 12 pixel training patches. In order to focus on statistical properties that distinguish foreground from background, each image patch in the training set was rotated to a canonical vertical orientation. Figure 6b shows the first six optimal linear receptive fields, given the same constraints on the neurons as in the patch identification task and using stimulus-manifold-sampling for initializing the RFs. As can be seen, there are of two general types of optimal RF. One type (RFs 1 & 3) is similar to an edge selective (sine phase) simple cell found in primary visual cortex. This shape is consistent with the observation that the foreground surface in close-up foliage tends on average to be more intense than the background side and to have a highlight along the occluding surface side of the boundary (see Figure 6a). The other type (RFs 2, 4 & 5) has a relatively uniform slightly excitatory region on one side and excitatory and inhibitory sub-regions on the other side. This shape is consistent with the observation that occluding surfaces tend to cut across contours in the background creating "t-junctions" and other patterns of contrast modulation that are more or less perpendicular to the occluding surface boundary (see Figure 6a).

Figure 5. One of 96 hand-segmented close-up images of foliage used to obtain random samples of surface boundary; segmented leaves (blue, brown); segmented branches (yellow). The brown leaf illustrates a single segmented object.

Interestingly, there are correlations among some of the optimal receptive fields: RFs 1 & 3 and RFs 4 & 6 are moderately correlated, and RFs 2 & 5 strongly correlated (see Table 2). The relatively redundant receptive fields provide useful additional information because they reduce the effects of the neural noise and because the task-relevant information is concentrated within a narrow range of spatial patterns.

The solid symbols in Figure 7a show the actual accuracy of the optimal decoder as each optimal receptive field is added in. Each data point was obtained by Monte Carlo simulation of 200 trials for each of the 200 image patches. Performance jumps up with the first receptive field and then climbs steadily so that six neurons, with the noise characteristics of typical cortical neurons, are sufficient to identify the foreground side of the image patch with an accuracy of 83%, where chance is 50%. Similar performance is obtained for randomly selected test patches not in the training set, suggesting that the receptive fields in Figure 6 are robust and not the result of over-fitting. Note also that despite the correlations between some of the receptive fields, the increments in accuracy are fairly constant after the first receptive field. Figure 7b shows that the actual accuracy found by Monte Carlo simulation (for all 6 RFs and 200 test patches not in the training set) is highly correlated with the estimated relative entropy based on Equation 4, again providing evidence for the adequacy of Equation 4.

Discussion

This paper describes a new approach, accuracy maximization analysis (AMA), for determining the stimulus features that are most relevant for performing specific identification tasks. The analysis takes as input random samples of stimuli that occur in the identification task and returns encoding functions that represent the properties of

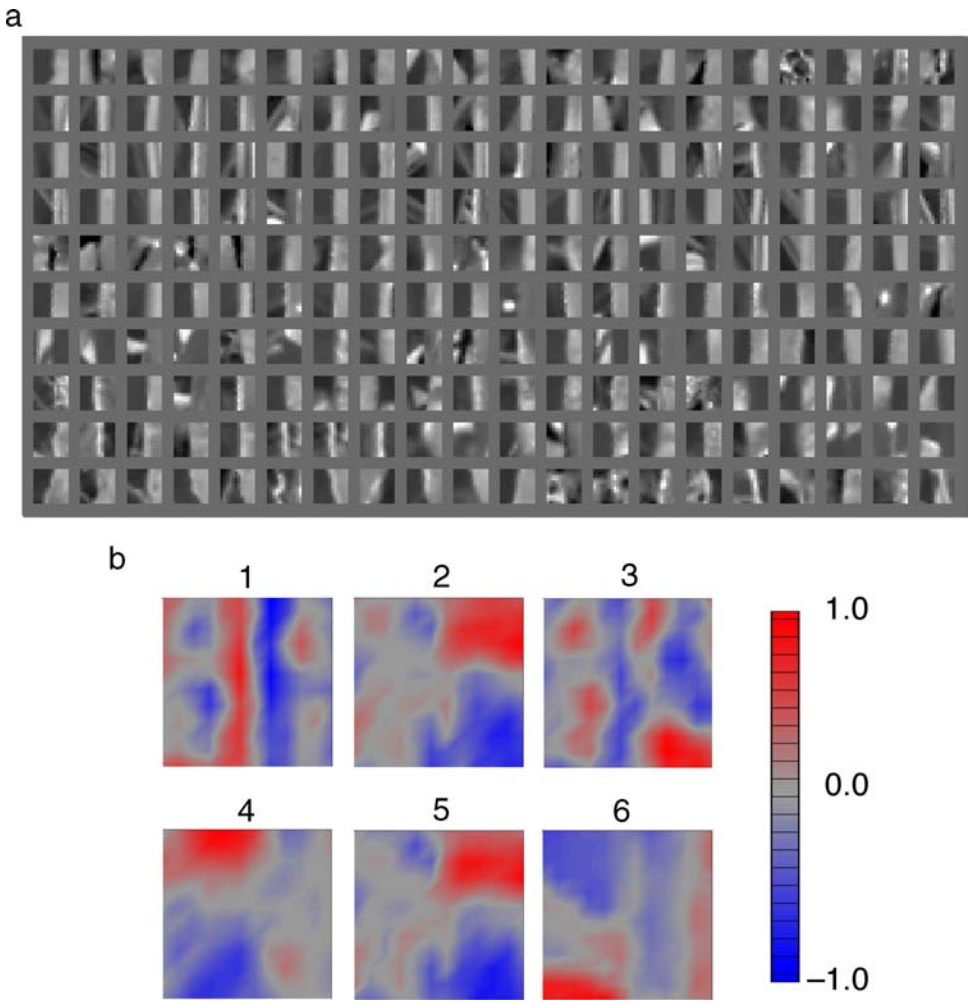

Figure 6. Foreground identification task. a. Example 12 × 12 pixel training patches. Each patch is centered on a randomly selected point along a surface boundary contour. In estimating the receptive fields all training patches were rotated to a canonical vertical orientation; when k = 1 the foreground was on the left when k = 2 the foreground was on the right. These patches have the foreground on the right. b. Final weights for first six optimal receptive fields. (For display purposes the receptive fields have been scaled so that the maximum absolute value is 1.0, and then interpolated by the plotting software.)

the stimuli that are optimal for performing the task. The optimal encoding functions are defined to be those that maximize performance accuracy when combined with the optimal decoder (Bayesian ideal observer) for the task.

AMA was illustrated by estimating optimal linear receptive fields for an image patch identification task and a foreground identification task. We found that stable and intuitive results were obtained with relatively few training samples (in the hundreds). Although the two examples were picked primarily as demonstrations of the general approach, they provide important new insights. The optimal linear receptive fields for identifying arbitrary natural image patches are relatively smooth (low-frequency) shapes similar to receptive fields found in primary visual cortex. Applying AMA at different spatial scales yielded similar receptive field shapes (not shown here), consistent with the approximate scale invariance of rural outdoor images. The optimal linear receptive fields for the foreground identification task were of two major types, an edge-selective shape parallel to the surface boundary, and edge-selective shapes perpendicular to the surface boundary.

Spatial Correlations of Receptive Fields

RF2	RF3	RF4	RF5	RF6	
−0.10	−0.60	−0.04	−0.11	0.18	**RF1**
	−0.15	−0.02	0.96	−0.08	**RF2**
		0.10	−0.21	−0.08	**RF3**
			−0.04	−0.60	**RF4**
				−0.07	**RF5**

Table 2. Spatial correlations between the six optimal receptive fields in Figure 6b.

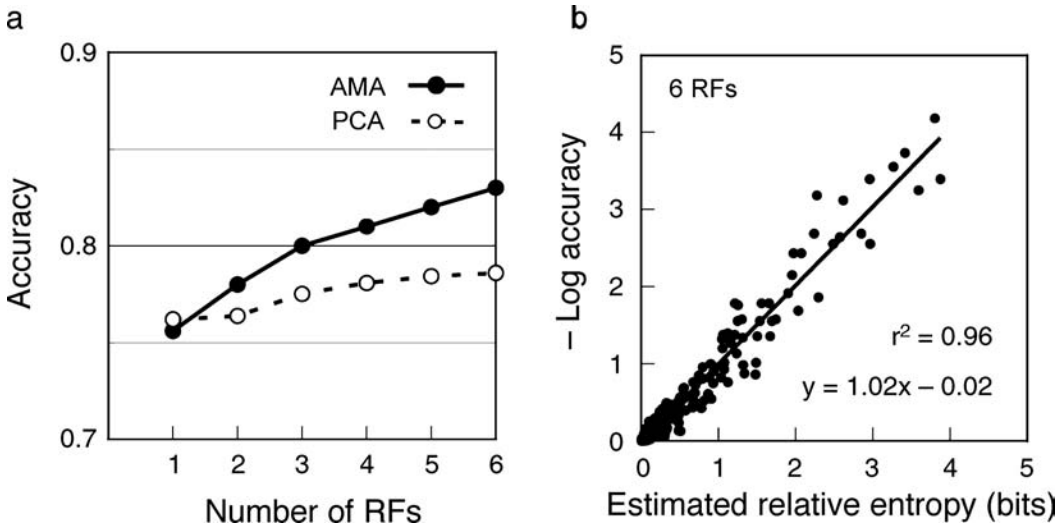

Figure 7. Actual accuracy of the optimal decoder in the foreground identification task, as determined by Monte Carlo simulation. a. Actual accuracy as a function of the number of optimal AMA receptive fields (solid symbols) and as a function of the number of PCA receptive fields (open symbols). b. Actual accuracy vs. estimated relative entropy, for 200 test patches not in the training set.

It is important to recognize that, like all ideal observer analyses, AMA is not a model of perceptual processing but a tool for understanding the stimulus information available to perform a task. Whether or not the information is used by an organism, and if it is, how the organism extracts and uses that information are separate issues. The primary purpose of AMA is to characterize the properties of natural stimuli that are most relevant to a given task, a crucial step for understanding perceptual systems.

Finding optimal features

Many methods have been proposed for finding optimal features from natural stimuli. One class of methods, which includes principle components analysis (PCA) and independent components analysis (ICA), focuses on characterizing the general statistical properties of natural stimuli, often for the purpose of understanding efficient coding of natural stimuli. An obvious question is how the feature dimensions found with such methods compare to those obtained with AMA. Figure 8a shows the first six principle components (receptive fields) of the training images (Figure 2a) for the patch identification task. In PCA, all the receptive fields are required to be orthogonal and are rank-ordered in the percentage of variance accounted for in the training set, with the first principle component accounting for the most variance. Comparison of Figures 8a and 2b shows that AMA and PCA find very similar receptive fields in this task. Furthermore, the open symbols in Figure 4a show that the identification performance of the optimal decoder that uses the PCA receptive fields is nearly as good as the performance of the optimal decoder that uses the AMA receptive fields.

This is an intuitive result that has several implications. In the patch identification task the goal is to uniquely identify each image patch, and thus it is intuitive that one would want an encoding where the representation of the patches is as spread out as possible. The directions of maximum variation in the space of image patches (the first n principle components) might provide a near optimal encoding given a limited number of feature dimensions. Indeed the similarity of the PCA and AMA receptive fields provides evidence that PCA, which is computationally much simpler and faster than AMA, produces receptive fields that are very effective for uniquely identifying image patches. This result also suggests that AMA is finding approximately the global optimum for this task.

A rather different result is obtained for the foreground identification task. Figure 8b shows the first six principle components of the training images (Figure 6a) for the foreground identification task. Comparison of Figures 6b and 8b shows that AMA and PCA find rather different receptive fields in this task. The open symbols in Figure 7a show that the PCA receptive fields are less efficient than the AMA receptive fields, except perhaps for the first receptive field. For example, the second PCA component captures the dimension in image-patch space with the second largest variance, yet it provides essentially no improvement in performance accuracy, even with its matched optimal decoder for the task. Indeed, none of the PCA receptive fields after the first provides as much improvement in performance as the corresponding AMA receptive field. Note that even when an added AMA receptive field is substantially correlated with one or more previously added receptive fields it increases accuracy more than adding an orthogonal PCA receptive field,

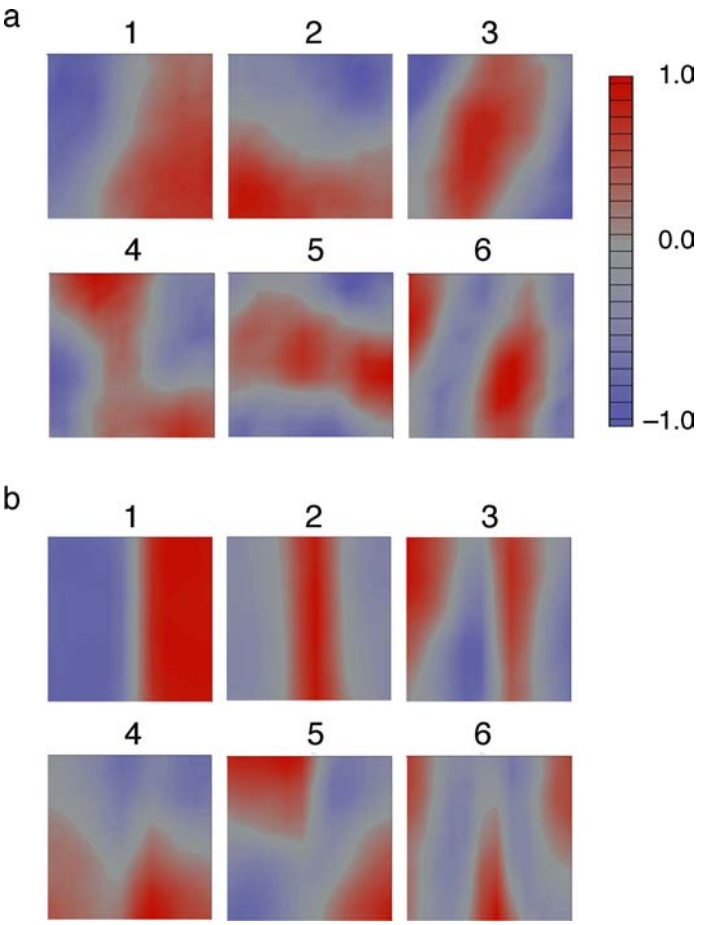

Figure 8. Principle components. a. First six principle components for the training data in the patch identification task (cf., Figure 2b). b. First six principle components for the training data in the foreground identification task (cf., Figure 6b). To maximize comparisons with the AMA receptive fields, all image patches were normalized to a mean of 0.0 and a standard deviation of 1.0 before computing the principle components (see Equation 6).

which one would normally expect to capture more independent information. These results are perhaps not surprising because for many identification tasks only certain stimulus dimensions are relevant to the task. It is the patch identification task that may be unique because all the dimensions of variation between patches can contribute to task performance.

In Figures 4a and 7a the PCA components were added in rank order (i.e., in the order of the percentage of variance accounted for in the training set). Another approach is to select successive components that maximally increase identification accuracy, allowing repeated use of the same component. For the patch identification task (Figure 4a) the results are identical; each component is only picked once and the order corresponds to the PCA rank order. For the foreground identification task the optimal ordering of PCA components is 1, 3, 3, 6, 4, 6 (see Figure 8b). However, the overall accuracy improves only slightly; like Figure 4a, as components are added the accuracy increases only about half as much as it does for the AMA components.

AMA belongs to the second major class of methods for finding optimal features for natural stimuli. These methods focus on finding optimal features for specific tasks. The most popular are based on multilayer neural networks and include back-propagation (Rumelhart, Hinton, & Williams, 1986) and related methods (e.g., see Duda, Hart, & Stork, 2001). Back-propagation methods are theoretically capable of reaching optimal classification performance. The learned weights in the early hidden layer might be interpreted as the optimized feature dimensions of the encoder and learned weights of the later layers as the optimized decoder. However, both the encoding and decoding weights must be learned simultaneously and there is no guarantee that the specific architecture of the network selected (i.e., the numbers of units per layer and activation functions) will support optimal performance. An advantage of AMA is that the decoder is guaranteed to be optimal and does not have to be learned, and thus the training is entirely focused on the encoding functions. Other than the potential for getting trapped in local optima during training and the sampling-noise effect of a finite training set, AMA should find the optimal encoding functions within the family of possible encoding functions under consideration (e.g., linear weighting functions).

AMA is, perhaps, most related to the method of Ullman et al. (2002), which attempts to find optimal features by maximizing the mutual information between image fragments (which serve as features) and the categories. However, the Ullman et al. method restricts the encoding functions to the set of image fragments in the training set, does not represent noise in the encoder, and does not easily generalize to larger numbers of categories.

In sum, AMA has a unique combination of desirable properties. First, if the approximation in Equation 4 is sufficiently accurate (which appears to be the case for our examples), then the approach is entirely principled, and should provide near optimal encoding functions, given sufficient training stimuli. Second, the approach allows explicit incorporation of biophysical constraints in the representation of the encoded stimuli, such as neural noise and a limited dynamic range, which are unavoidable minimal constraints for any real perceptual system. Third, as discussed below, the approach easily generalizes to other kinds of tasks and to nonlinear families of encoding functions. Fourth, it applies directly to arbitrary numbers of categories (e.g., from 2 to 200 in our examples).

Efficiency and redundancy

The efficient coding hypothesis holds that evolution together with learning over the lifespan push perceptual systems toward encoding sensory information as compactly as possible (i.e., with the fewest numbers of

neurons and/or action potentials), or equivalently, that they push any given population of neurons in a perceptual system toward encoding as much sensory information as possible (Attneave, 1954; Barlow, 1961, 2001; Olshausen, 2003; Simoncelli, 2003; Simoncelli & Olshausen, 2001). The simplest forms of the efficient coding hypothesis ignore neural noise, and thus maximizing efficiency generally involves maximally reducing redundancy in the neural code (Simoncelli, 2003). However, substantial neural noise exists in all perceptual systems and thus efficient coding in real perceptual systems may require substantial redundancy. Furthermore, in some tasks redundancy is optimal because the relevant information tends to be concentrated in certain stimulus patterns (e.g., the foreground identification task).

One advantage of AMA is that neural noise is explicitly represented. Another advantage is that AMA (unlike PCA and ICA) requires no arbitrary assumptions such as orthogonality of the encoding functions or statistically independent sources in the natural stimuli. Thus, like natural evolution and learning, AMA is free to select orthogonal, sparse or redundant encoding functions depending on the task, the natural scene statistics, the neural noise, and/or other biophysical constraints. For example, in the patch identification task the optimal encoding functions tend to be orthogonal, whereas in the foreground identification task the optimal encoding functions tend to be more redundant. More generally, the framework illustrated in Figure 1 may provide both an intuitive and formal understanding the relationship between efficiency and redundancy.

Neural noise

In the examples presented here the neural noise was set to mimic the noise properties of individual neurons in primary visual cortex: variance proportional to the mean response with proportionality constant (Fano factor) of 1.3, and low spontaneous activity. Thus, the estimated linear receptive fields should be close to optimal for neurons with noise characteristics like those in primary visual cortex (neurons in other cortical areas generally have similar noise characteristics to those in V1). Nonetheless, an important question is the degree to which the optimal encoding functions depend upon the amount and structure of the neural noise. Intuition suggests that the level of neural noise should have relatively little effect on the form of the optimal encoding functions, but will have an effect on the amount of redundancy in the optimal encoding functions; the greater the level of neural noise the more performance can be improved by adding copies of the same optimal encoding function. In agreement with these intuitions we estimated RFs for Fano factors varying from 0.4 to 1.4 and found them to be quite similar. Similarly, simulations show (in agreement with intuition) that switching from multiplicative to additive noise has a relatively modest effect on the form of the optimal encoding functions.

Unlike the present assumption of statistical independence, the noise of neurons in primary visual cortex (and other cortical areas) tends to be correlated over space and time (Gawne & Richmond, 1993; Lee, Port, Kruse, & Georgopoulos, 1998; Romo, Hernandez, Zainos, & Salinas, 2003; Zohary et al., 1994). The correlations are generally small (on the order of 0.2 or less). Such correlations have relatively little effect on optimal encoding and decoding for small populations of neurons (such as the small populations considered in the present examples), but can have substantial effects in large populations (Chen, Geisler, & Seidemann, 2006, 2008; Seidemann, Chen, & Geisler, 2009). It is straightforward to apply AMA in the case of correlated Gaussian neural noise, because the optimal decoder for correlated Gaussian noise is well understood. Specifically, applying a 'whitening' filter (the inverse of the noise correlation matrix) to the population response removes the noise correlation, allowing AMA to then be applied in the same way it is applied in the statistically independent case. However, the noise correlation can affect the number of redundant optimal encoding functions needed for optimal performance and can also have some effect on the form (shape) of the optimal encoding functions, because of the whitening operation.

Computational issues

There are two substantial computational limitations to the version of AMA described here. First, for a fixed number of categories, the number of arithmetic operations increases in proportion to the square of the number of training samples. Thus, applying the algorithm to very large training sets can be prohibitively slow on standard desktop computers; our experience is that estimating a half dozen optimal linear (12×12 pixel) receptive fields, for several hundred training stimuli, can take several hours. However, AMA is amenable to large scale parallel computing, which could be used to obtain estimates for large sample sizes. An alternative procedure we employ is to obtain estimates for different sets of training samples. For the current examples, the optimal receptive fields were very similar with different sets of samples, suggesting that the receptive fields would not change greatly (just be smoother) with a very large sample size.

A second limitation is that simple gradient descent is used to estimate the optimal encoding functions, and thus there is the usual potential problem of getting trapped in local optima. Two versions of gradient descent were used here. In one version the encoding functions were initialized with random values. For the patch identification task a similar family of final encoding functions was obtained for different random starting values, but the ordering of the different function shapes varied somewhat; for the

foreground identification task, the encoding functions were less stable. In the other version, the encoding functions were initialized with stimulus-manifold-sampling. For the patch identification task the final encoding functions were very similar to those obtained with random sampling; for the foreground identification task, the final encoding functions were superior (lower relative entropy and more stable across training sets). A future direction is to explore more sophisticated parameter estimation procedures.

Cost–benefit (utility) functions

In some natural identification tasks, correct responses will not all have equal benefit and errors will not all have equal costs. In such cases it is necessary to include a cost–benefit (utility) function $\gamma(i, j)$ that gives the utility of picking category i when the correct category is j. This utility function is specified in the definition of the identification task. In the standard Bayesian approach, a rational observer is defined to be one that picks the category with the maximum expected utility (i.e., minimum risk):

$$\bar{\gamma}\left(i|\mathbf{R}_q\right) = \sum_{j=1}^{m} \gamma\left(i,j\right) p\left(j|\mathbf{R}_q\right).$$

A simple way to generalize the current algorithm is to restrict (without loss of generality) the utility function to values greater than zero. Then, the expected utility of each category is always greater than or equal to zero, and thus we can obtain a simple generalization of Equation 5:

$$\bar{D}_q = -\frac{1}{n}\sum_{i=1}^{n} \log\left[\max_x \bar{\gamma}\left(x|\mathbf{r}_q(K_i, L_i)\right)\right].$$

This function reduces to Equation 5 when the utility of corrects is 1.0 and of errors is 0.0.

How much does the utility function, $\gamma(i, j)$, affect the optimal encoding functions? We have not yet systematically explored this question, but it is possible that the optimal encoding functions are relatively insensitive to modest variations of the utility function. For example, the primary effect of variations of the utility function in many identification tasks is to change decision boundaries and such changes presumably do not change the stimulus dimensions (features) that are optimal for performance of the task, but this remains to be explored.

Finally, note that estimation tasks can be described as identification tasks where the categories are ordered and dense. Thus, by defining utility functions that appropriately reward picking categories near the correct category it may be possible to apply AMA to natural estimation tasks.

Nonlinear encoding functions

In the present examples the encoding functions were restricted to the family of linear weighting functions. However, any parameterized family of encoding functions could be used. For example, the encoding functions could include a parameterized expansive, compressive or divisive nonlinearity. The parameters of the nonlinearity could be estimated simultaneously with the linear weights, all by gradient descent, with little additional computational cost. Thus, with AMA there is no particular limitation on the family of encoding functions considered.

Optimal decoding

The purpose of accuracy maximization analysis is to find those stimulus properties (represented by the optimal encoding functions) that are most useful for performing a given perceptual task. To do this we use, as the decoder, a closed-form approximation of the Bayesian ideal observer that knows the mean response and noise characteristics of each neuron (i.e., of each encoding function output) to each stimulus in the training set. Because we are only interested here in finding the optimal encoding functions, we do not require the optimal decoder to generalize beyond the training set. We do however check the generality of the optimal encoding functions by comparing the optimal encoding functions obtained with different training sets, and by checking whether the optimal encoding functions estimated from a given set of training stimuli yield similar performance accuracy on other random sets of stimuli.

Once the optimal encoding functions are nailed down with AMA, a logical next step is to determine the joint distribution of environmental states and optimal feature values by measuring the responses of the optimal encoding functions to large numbers of stimuli (e.g., see Geisler, 2008). Determining this distribution is relatively easy because the natural stimuli are mapped into the relatively low-dimensional space where each axis represents the response of one of the optimal encoding functions. Standard pattern classification techniques (Duda et al., 2001) can then be used to find the decision functions of the generalized optimal decoder, which might serve as a principled hypothesis for decoding in the brain (e.g., see Geisler & Perry, 2009).

Task dependent versus task independent encoding

The purpose of AMA and related methods is to determine the stimulus properties that are most relevant for performing specific tasks. This is useful for gaining an

understanding of the computational requirements of natural tasks, providing a benchmark against which to evaluate an organism's performance, suggesting hypotheses for neural encoding, and designing artificial vision systems for specific tasks.

A more general issue is whether the receptive fields (encoding functions) observed in the early levels of the visual system (e.g., V1) are a composite of specialized receptive fields for various specific tasks or whether they are better described as an efficient coding of the general structure of natural images (e.g., see Simoncelli & Olshausen, 2001). An argument for the latter case is that primates (and many other mammals) perform such a wide variety of specific visual tasks, each requiring different kinds of information, that the best one can do is encode the local image information in as compact (low redundancy) and as separable/accessible (sparse) a representation as possible. However, it is perhaps even more plausible that certain common low-level tasks, such as local contour detection/grouping, local texture discrimination/grouping, and foreground identification, are components of most visual tasks and hence that these low-level tasks have driven the selection/learning of specific receptive field types in early visual areas.

These seemingly different views of early visual coding may not be so different. For example, in the current pattern identification task the goal seems to boil down to finding a small set of feature dimensions that spread out the representation of natural image patches as much as possible. Thus, the optimal feature dimensions for this task at various spatial scales (sizes of image patches) may represent an efficient coding of the general structure of natural images. Another reason that the two views may overlap is that there may be substantial statistical dependence between some of the common low-level tasks, which may lead to features optimized for more than one low-level task (e.g., there is similarity between some of the optimal features for the patch identification and foreground identification tasks).

Conclusion

Perceptual systems must reflect those statistical properties of natural stimuli that enable performance of the tasks the organism normally performs to survive and reproduce. Thus, a critical step in systems and behavioral neuroscience is to gain a rigorous understanding of task-relevant natural scene statistics. A rigorous understanding of these statistics is not only important in its own right, but can provide principled hypotheses for what stimulus properties are coded by perceptual systems and for how the brain might exploit those properties in performing its natural tasks. The method for measuring these statistics described here (accuracy maximization analysis) is computationally intensive, but has the potential for rigorously determining optimal stimulus properties for a wide range of specific natural tasks.

Appendix A

Here we derive formulas for the posterior probability distribution that is computed by the ideal Bayesian observer when receiving a population response $\mathbf{R}_q(k, l)$ to a presentation of stimulus $\mathbf{s}(k, l)$. (Keep in mind that the ideal observer does not know that the stimulus is $\mathbf{s}(k, l)$, but does know the mean response of each neuron in the population to each stimulus in the training set.) According to Bayes' rule:

$$p\left(x|\mathbf{R}_q(k,l)\right) = \frac{p\left(\mathbf{R}_q(k,l)|x\right)p(x)}{\sum_{i=1}^{m} p\left(\mathbf{R}_q(k,l)|i\right)p(i)}.$$

Given the assumed statistical independence of the neural noise we have,

$$p\left(x|\mathbf{R}_q(k,l)\right) = \frac{p(x)\prod_{t=1}^{q} p(R_t(k,l)|x)}{\sum_{i=1}^{m} p(i)\prod_{t=1}^{q} p(R_t(k,l)|i)}. \quad \text{(A1)}$$

To derive the recursive formula in text Equation 10 we rewrite the above equation,

$$p\left(x|\mathbf{R}_q(k,l)\right) = \frac{p\left(R_q(k,l)|x\right)p\left(x\right)\prod_{t=1}^{q-1} p(R_t(k,l)|x)}{\sum_{i=1}^{m} p\left(R_q(k,l)|i\right)p\left(i\right)\prod_{t=1}^{q-1} p(R_t(k,l)|i)}$$

$$p\left(x|\mathbf{R}_q(k,l)\right) = \frac{p\left(R_q(k,l)|x\right)p\left(x|\mathbf{R}_{q-1}(k,l)\right)}{\sum_{i=1}^{m} p\left(R_q(k,l)|i\right)p\left(i|\mathbf{R}_{q-1}(k,l)\right)},$$

$$p\left(x|\mathbf{R}_q(k,l)\right) = \frac{p\left(x|\mathbf{R}_{q-1}(k,l)\right)\sum_{j=1}^{n_x} p\left(R_q(k,l)|x,j\right)p(j|x)}{\sum_{i=1}^{m} p\left(i|\mathbf{R}_{q-1}(k,l)\right)\sum_{j=1}^{n_i} p\left(R_q(k,l)|i,j\right)p(j|i)},$$

$$p\left(x|\mathbf{R}_q(k,l)\right) = \frac{p\left(x|\mathbf{R}_{q-1}(k,l)\right)\frac{1}{n_x}\sum_{j=1}^{n_x} p\left(R_q(k,l)|x,j\right)}{\sum_{i=1}^{m} p\left(i|\mathbf{R}_{q-1}(k,l)\right)\frac{1}{n_i}\sum_{j=1}^{n_i} p\left(R_q(k,l)|i,j\right)}.$$

By substitution of text Equation 8 we have:

$$p\left(x|\mathbf{R}_q(k,l)\right) = \frac{p\left(x|\mathbf{R}_{q-1}(k,l)\right)\frac{1}{n_x}\sum_{j=1}^{n_x}\frac{1}{\sigma_q(x,j)}\exp\left[-\frac{1}{2}\frac{\left[R_q(k,l)-r_q(x,j)\right]^2}{\sigma_q(x,j)^2}\right]}{\sum_{i=1}^{m}p\left(i|\mathbf{R}_{q-1}(k,l)\right)\frac{1}{n_i}\sum_{j=1}^{n_i}\frac{1}{\sigma_q(i,j)}\exp\left[-\frac{1}{2}\frac{\left[R_q(k,l)-r_q(i,j)\right]^2}{\sigma_q(i,j)^2}\right]}. \quad \text{(A2)}$$

Equation 10 then follows by substitution from Equation 4. Note, Z in Equation 10 corresponds to the denominator in Equation A2.

The following non-recursive version of this equation, which follows directly from Equation A1, was used in the Monte Carlo simulations underlying Figures 3, 4 and 7:

$$p\left(x|\mathbf{R}_q(k,l)\right) = \frac{\left(\frac{1}{n_x}\right)^{q-1}\prod_{t=1}^{q}\sum_{j=1}^{n_x}\frac{1}{\sigma_t(x,j)}\exp\left[-\frac{1}{2}\frac{\left[R_t(k,l)-r_t(x,j)\right]^2}{\sigma_t(x,j)^2}\right]}{\sum_{i=1}^{m}\left(\frac{1}{n_i}\right)^{q-1}\prod_{t=1}^{q}\sum_{j=1}^{n_i}\frac{1}{\sigma_t(i,j)}\exp\left[-\frac{1}{2}\frac{\left[R_t(k,l)-r_t(i,j)\right]^2}{\sigma_t(i,j)^2}\right]}. \quad \text{(A3)}$$

Acknowledgments

We thank James Elder, Jonathan Pillow and Eyal Seidemann for helpful comments and discussion. Supported by NIH grant EY11747.

Commercial relationships: none.
Corresponding author: Wilson S. Geisler.
Email: geisler@psy.utexas.edu.
Address: Center for Perceptual Systems and Department of Psychology, University of Texas at Austin, Austin, TX 78712, USA.

Footnote

[1]Note that making the variance proportional to the absolute value of the mean response allows negative responses. We could easily have half-wave rectified the responses to be more consistent with real neurons, but allowing negative responses reduces the number of optimal encoding functions that need to be estimated.

References

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review, 61,* 183–193. [PubMed]

Barlow, H. B. (1961). Possible principles underlying the transformations of sensory messages. In W. A. Rosenblith (Ed.), *Sensory communication* (pp. 217–234). Cambridge, MA: MIT Press.

Barlow, H. B. (2001). Redundancy reduction revisited. *Network, 12,* 241–253. [PubMed]

Bell, A. J., & Sejnowski, T. J. (1997). The "independent components" of natural scenes are edge filters. *Vision Research, 37,* 3327–3338. [PubMed]

Brunswik, E., & Kamiya, J. (1953). Ecological cue-validity of "proximity" and other Gestalt factors. *American Journal of Psychology, 66,* 20–32.

Chen, Y., Geisler, W. S., & Seidemann, E. (2006). Optimal decoding of correlated neural population responses in the primate visual cortex. *Nature Neuroscience, 9,* 1412–1420. [PubMed] [Article]

Chen, Y., Geisler, W. S., & Seidemann, E. (2008). Optimal temporal decoding of V1 population responses in a reaction-time detection task. *Journal of Neurophysiology, 99,* 1366–1379. [PubMed] [Article]

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory, 2nd edition*. New York: John Wiley & Sons, Inc.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification*. New York: Wiley.

Elder, J. H., & Goldberg, R. M. (2002). Ecological statistics of Gestalt laws for the perceptual organization of contours. *Journal of Vision, 2*(4):5, 324–353, http://journalofvision.org/2/4/5/, doi:10.1167/2.4.5. [PubMed] [Article]

Fowlkes, C. C., Martin, D. R., & Malik, J. (2007). Local figure–ground cues are valid for natural images. *Journal of Vision, 7*(8):2, 1–9, http://journalofvision.org/7/8/2/, doi:10.1167/7.8.2. [PubMed] [Article]

Gawne, T. J., & Richmond, B. J. (1993). How independent are the messages carried by adjacent inferior temporal cortical-neurons. *Journal of Neuroscience, 13,* 2758–2771. [PubMed] [Article]

Geisler, W. S. (2008). Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology, 59,* 167–192. [PubMed]

Geisler, W. S., & Perry, J. S. (2009). Contour statistics in natural images: Grouping across occlusions. *Visual Neuroscience, 26,* 109–121.

Geisler, W. S., & Albrecht, D. G. (1997). Visual cortex neurons in monkeys and cats: Detection, discrimination and identification. *Visual Neuroscience, 14,* 897–919. [PubMed]

Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts

contour grouping performance. *Vision Research, 41,* 711–724. [PubMed]

Geisler, W. S., Perry, J. S. & Ing, A. D. (2008). Natural systems analysis. In B. Rogowitz & T. Pappas (Eds.), *Human vision and electronic imaging, SPIE Proceedings,* Vol. 6806.

Konishi, S., Yuille, A. L., Coughlan, J. M., & Zhu, S. C. (2003). Statistical edge detection: Learning and evaluating edge cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 25,* 57–74.

Laughlin, S. B. (1981). A simple coding procedure enhances a neuron's information capacity. *Zeitschrift für Naturforschung. C, 36,* 910–912. [PubMed]

Lee, A. B., Pedersen, K. S., & Mumford, D. (2003). Nonlinear statistics of high-contrast patches in natural images. *International Journal of Computer Vision, 54,* 83–103.

Lee, D., Port, N. L., Kruse, W., & Georgopoulos, A. P. (1998). Variability and correlated noise in the discharge of neurons in motor and parietal areas of the primate cortex. *Journal of Neuroscience, 18,* 1161–1170. [PubMed] [Article]

Martin, D. R., Fowlkes, C. C., & Malik, J. (2004). Learning to detect natural image boundaries using local brightness, color and texture cues. *IEEE Transactions Pattern Analysis Machine Intelligence, 26,* 530–549. [PubMed]

Motoyoshi, I., Nishida1, S., Sharan, L., & Adelson, E. H. (2007). Image statistics and the perception of surface qualities. *Nature, 447,* 206–209. [PubMed]

Olshausen, B. A. (2003). Principles of image representation in visual cortex. In L. M. Chalupa & J. S. Werner (Eds.), *The visual neurosciences* (pp. 1603–1615). Cambridge, MA: MIT Press.

Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy by V1? *Visual Research, 37,* 3311–3325. [PubMed]

Romo, R., Hernandez, A., Zainos, A., & Salinas, E. (2003). Correlated neuronal discharges that increase coding efficiency during perceptual discrimination. *Neuron, 38,* 649–657. [PubMed]

Ruderman, D. L., & Bialek, W. (1994). Statistics of natural images: Scaling in the woods. *Physiological Review Letters, 73,* 814–817. [PubMed]

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by back-propagating errors. *Nature, 323,* 533–536.

Seidemann, E., Chen, Y., & Geisler, W. S. (2009). Encoding and decoding with neural populations in the primate cortex. In M. S. Gazzaniga (Ed.), *The cognitive neuroscience* IV. Cambridge: MIT Press.

Simoncelli, E. P. (2003). Vision and the statistics of the visual environment. *Current Opinion Neurobiology, 13,* 144–149. [PubMed]

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Reviews Neuroscience, 24,* 1193–1216. [PubMed]

Smith, E. W., & Lewicki, M. S. (2006). Efficient auditory coding. *Nature, 432,* 978–982. [PubMed]

Tolhurst, D. J., Movshon, J. A., & Dean, A. F. (1983). The statistical reliability of signals in single neurons in the cat and monkey visual cortex. *Vision Research, 23,* 775–785. [PubMed]

Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. *Trends in Cognitive Science, 11,* 58–64. [PubMed]

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience, 5,* 682–687. [PubMed]

van Hateren, J. H., & van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society London B: Biological Science, 265,* 359–366. [PubMed] [Article]

Zohary, E., Shadlen, M. N., & Newsome, W. T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature, 370,* 140–143. [PubMed]