Communicated by Graeme Mitchison

What Does the Retina Know about Natural Scenes?

Joseph J. Atick*
A. Norman Redlich
School of Natural Sciences, Institute for Advanced Study, Princeton, NJ 08540 USA

By examining the experimental data on the statistical properties of natural scenes together with (retinal) contrast sensitivity data, we arrive at a first principles, theoretical hypothesis for the purpose of retinal processing and its relationship to an animal's environment. We argue that the retinal goal is to transform the visual input as much as possible into a statistically independent basis as the first step in creating a redundancy reduced representation in the cortex, as suggested by Barlow. The extent of this whitening of the input is limited, however, by the need to suppress input noise. Our explicit theoretical solutions for the retinal filters also show a simple dependence on mean stimulus luminance: they predict an approximate Weber law at low spatial frequencies and a De Vries–Rose law at high frequencies. Assuming that the dominant source of noise is quantum, we generate a family of contrast sensitivity curves as a function of mean luminance. This family is compared to psychophysical data.

1 The Retina and the Visual Environment

An animal must have knowledge of its environment. As Barlow (1989) has emphasized, one important type of knowledge that needs to be stored in the brain is knowledge of the statistical properties of sensory messages. This provides an animal with data about the regular structures or features in its environment. New sensory messages can then be compared to expectations based on this background data; for example, the background data can be subtracted. In this way, one can argue, the brain is able to discover unexpected events and new associations. Here we explicitly explore the possibility that even the retina knows some of the statistical properties of visual messages. Our prejudice is that discovering how this information is used in the retina will not only help explain retinal processing but will be invaluable in applying this idea to the cortex.

To discover what the retina knows about the statistics of its environment, it is first necessary to find out just what characterizes the ensemble

*Address after July 1, 1992: The Rockefeller University, 1230 York Ave., New York, NY 10021, USA.

of visual messages in a natural environment. An important step in this direction has been taken by Field (1987), who has been analyzing pictures of "natural" scenes, such as landscapes without human-made objects as well as pictures of human faces. As Field has argued, these represent a very small subset of all possible images: all possible arrangements and values of a set of pixels. What he found is that natural images have unique and clearly defined statistical properties.

The first statistical measure Field calculated is the two-dimensional spatial autocorrelator

$$R(\mathbf{x}, \mathbf{y}) \equiv \langle L(\mathbf{x})L(\mathbf{y})\rangle \tag{1.1}$$

which is defined as the average over many scenes (or the average over one large scene assuming ergodicity) of the product of luminance levels $L(\mathbf{x})$ and $L(\mathbf{y})$ at two spatial points $\mathbf{x}$ and $\mathbf{y}$. Actually, by homogeneity of natural scenes the autocorrelator is only a function of the relative distance: $R(\mathbf{x}-\mathbf{y})$. One can thus define the *spatial power spectrum*, which is the Fourier transform of the autocorrelator $R(\mathbf{f}) = \int d\mathbf{x} e^{i\mathbf{f}\cdot\mathbf{x}} R(\mathbf{x})$. This is the quantity that Field directly measured. What he found is

$$R(\mathbf{f}) \sim \frac{1}{|\mathbf{f}|^2}$$

which corresponds to a scale invariant autocorrelator: under a global rescaling of the spatial coordinates $x \to \alpha x$ the autocorrelator $R(\alpha\mathbf{x}) \to R(\mathbf{x})$. Although this scale invariant spatial power spectrum is by no means a complete characterization of natural scenes, it is the simplest regularity they possess. The retina, being the first major stage in visual processing, is not expected to have knowledge beyond the simplest aspects of natural scenes and hence for understanding the retina the power spectrum may be sufficient.

The question at this stage is what is the relationship between this property of the visual environment and the observed visual processing by the retina? To answer this, let us explore what happens to the spatial power spectrum of the visual signal after it is processed by the retina. The output of one major class of retinal ganglion cells[2] is known to be related to the light input approximately through a linear filter:

$$O(\mathbf{x}_j) = \int d\mathbf{x}\ \ K(\mathbf{x}_j - \mathbf{x})\ L(\mathbf{x}) \equiv K \cdot L \tag{1.2}$$

where $L(\mathbf{x})$ is the light intensity at point $\mathbf{x}$, $O(\mathbf{x}_j)$ is the output of the jth ganglion cell, and $K(\mathbf{x}_j - \mathbf{x})$ is the linear ganglion cell kernel ($\mathbf{x}_j$ is the center of the cell's receptive field. Here we assume translation invariance of the kernel K, which means that all ganglion cell kernels are the same function, but translated on the retina). Once adapted to bright light, this ganglion cell kernel, in spatial frequency space, is a bandpass filter.

[2]X-cells in cat, P-pathway cells in monkey.

Typical retinal filters at high luminosity are shown in Figure 1*A* and *C*,[3] where the experimental responses $K(\mathbf{f})$ [actually the contrast sensitivity which is $K(\mathbf{f})$ times the mean luminance I_0] are plotted against stimulus frequency. The data shown in Figure 1A are from De Valois *et al.* (1974), while the data in Figure 1C are from Kelly (1972).

Now to see how the power spectrum is modified by the retina, we need only multiply the input spectrum $R(\mathbf{f})$ by $K(\mathbf{f})K^*(\mathbf{f})$ since the average output spectrum is $\langle O(\mathbf{f})O^*(\mathbf{f})\rangle = \langle (K(\mathbf{f})L(\mathbf{f}))(K(\mathbf{f})L(\mathbf{f}))^*\rangle$. We can also plot the square root of this output spectrum — the *amplitude* spectrum — simply by multiplying the experimentally measured kernels $K(\mathbf{f})$ in Figure 1A and C by the input amplitude spectrum

$$\sqrt{R(\mathbf{f})} = |\mathbf{f}|^{-1}$$

This has been done in Figure 1B and D, which shows an intriguing result. At low frequencies, the input spectrum $|\mathbf{f}|^{-2}$ is converted into a flat spectrum at the retinal output: $\langle O(\mathbf{f})O^*(\mathbf{f})\rangle$ = constant. This *whitening* of the input by the retina continues up to the frequency where the kernels in Figure 1A and C peak. Had this whitening continued up to the system's cutoff frequency, this would have meant the ganglion cell outputs would be completely *decorrelated* in space. This is because a white or flat spectrum in frequency space Fourier transforms into a delta function in space, giving $\langle O(\mathbf{x}_i)O(\mathbf{x}_j)\rangle \sim \delta_{ij}$. In other words, the signals on different ganglion cell nerve fibers would be statistically independent. So it appears that the retina is attempting to decorrelate its input, at least down to the scale of the peak frequency.

The idea that the brain is attempting to transform its sensory input to a statistically independent basis has been suggested by Goodall (1960) and Barlow (1989) (see also Barlow and Foldiak 1989), and has been discussed by many others. Barlow has emphasized that one advantage of having a statistically independent set of outputs O_i is that all of their joint probabilities $P_{ijk\ldots}$ can be obtained directly from knowledge of the relatively small set of individual probabilities P_i. The values of the individual P_i can also be represented by taking the output strengths O_i to be proportional to their improbability, $-\log(P_i)$, that is, to the amount of information in each output. This then gives a very compact representation of not only the signals, but also their probabilities. In such a statistically

[3] Actually, what is plotted in Figure 1A and C are the results of psychophysical contrast sensitivity measurements, rather than of single ganglion cell responses. The single-cell results, however, are qualitatively similar, and in this short paper for conciseness we compare theory exclusively to psychophysical results (all figures). In general, we believe that the psychophysical data represent an envelope of the collection of single-cell contrast sensitivities. Then, given our assumption of translation invariance, the psychophysical envelope and the single-cell results should coincide. However, we do not exclude the possibility of a more complicated relationship between psychophysical and single-cell contrast sensitivities.

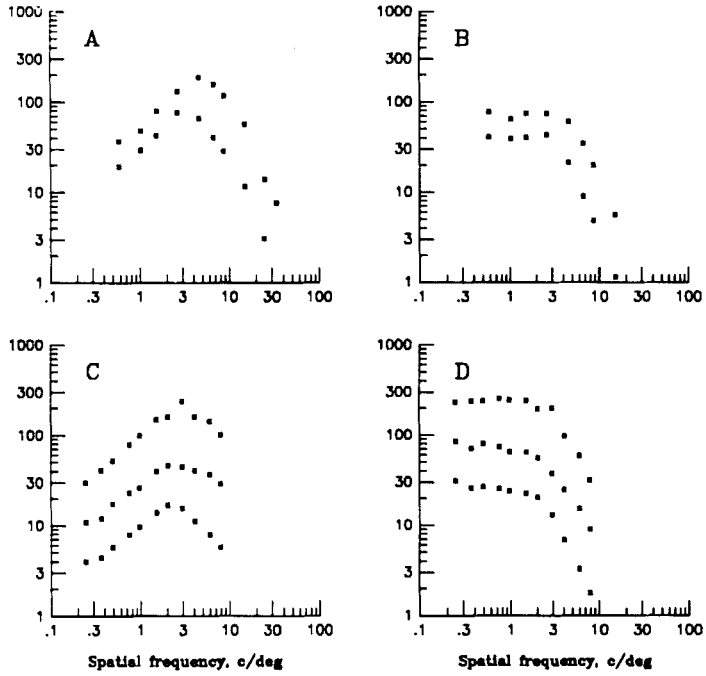

Figure 1: Retinal filters (A, C) in Fourier space at high mean luminosities, taken from the contrast sensitivity data of De Valois *et al.* (1974) (A) and Kelly (1972) (C). B (D) is the data in A (C) multiplied by $1/|\mathbf{f}|$, which is the amplitude spectrum of natural scenes. This gives the retinal ganglion cells' output amplitude spectrum. Notice the whitening of the output at low frequencies. The ordinate units are arbitrary.

independent basis, the outputs O_i represent "features," for example, in English text they would correspond roughly to "words"; they are the statistical structures that carry useful information. Finding these features effectively reduces the redundancy in the original sensory messages, leaving only the so-called "textual" (not predictable) information. One may therefore state this goal of statistical independence in information theory language as a type of redundancy reduction.

Based on the experimental evidence in Figure 1B and D, one might advance the hypothesis that the goal of the retinal processing is to produce a decorrelated representation of an image. However, this cannot be the only goal in the presence of input noise such as photon noise or biochemical transduction noise. In that case, decorrelation alone would

be a very dangerous computational strategy as we now illustrate: If the retina were to whiten all the way up to the cutoff frequency or resolution limit, the kernel $K(\mathbf{f})$ would be proportional to $|\mathbf{f}|$ up to that limit. This would imply a constant average squared response KRK^* to "natural" signals $L(\mathbf{x})$, which for $R \sim |\mathbf{f}|^{-2}$ have large spatial power at low frequencies and low power at high frequencies. But this same $K(\mathbf{f}) \sim |\mathbf{f}|$ acting on input noise whose spatial power spectrum is approximately flat (noise is usually already decorrelated) has a very undesirable effect, since it amplifies the noise at high frequencies where noise power, unlike signal power, is not becoming small. Therefore, even if input noise were not a major problem without decorrelation, after complete decorrelation (or whitening up to cutoff) it would become a problem. Also, if both noise and signal are decorrelated at the output, it is no longer possible to distinguish them. Thus, if decorrelation is a strategy, there must be some guarantee that no significant input noise is passed through the retina to the next stage.

Further evidence that the retina is concerned about not passing significant amounts of input noise is found in experiments in which the mean stimulus luminance is lowered. In response to this change, the ganglion cell kernel $K(\mathbf{f})$ makes a transition from bandpass to lowpass filtering. This is just the type of transition expected if the kernel is adapting to a lower signal to noise ratio, since lowpass filtering is a standard signal processing technique for smoothing away noise. Such a bandpass to lowpass transition also occurs when the temporal modulation frequency of the stimulus is increased (the retinal kernel is actually a function of both the spatial frequency $\mathbf{f}$ and the temporal frequency w, which has up to now been suppressed). In this case too there is an effective decrease in the *spatial* signal to noise ratio, so it is also evidence for noise suppression.

In a previous paper (Atick and Redlich 1990) we found an information theoretic formalism that unifies redundancy reduction and noise suppression. That formalism predicts all the qualitative aspects of the experimental data. However, it is highly technical and uses parameters that do not seem to have clear physical roles. This makes it more difficult to do quantitative comparisons with experiments, since the necessary dependence of these parameters on, for example, mean luminance is not intuitive. In this paper we adopt a modular approach where noise suppression and redundancy reduction are done in separate stages. This has two advantages: first it produces parameters with more direct physical meaning, and second it gives a clearer theoretical understanding of the purpose of retinal processing.

In the next section we formulate our theory mathematically making more concrete the heuristic notions of decorrelation and noise suppression. We then derive a simple theoretical retinal transfer function, and compare it to experiments.

2 Decorrelation as a Computational Strategy in Retina

2.1 Decorrelation in the Absence of Noise. In the previous section, we gave some experimental evidence leading to the hypothesis that the goal of retinal processing is to produce a representation with reduced redundancy. This implies a representation where the ganglion cell activities are as decorrelated as possible (more generally, statistically independent), given the inherent problem of input noise in the retina. In this section, we formulate this notion as a mathematical theory of the retina. We first set up the decorrelation problem ignoring noise, and later introduce the simple but important modification needed for noise suppression.

The outputs $\{O(\mathbf{x}_i)\}$ of the array of ganglion cells are completely decorrelated iff $\langle O(\mathbf{x}_i)O(\mathbf{x}_j)\rangle \sim \delta_{ij}$, where the brackets denote an ensemble average over natural stimuli. In general, due to the presence of noise, the retina will not decorrelate completely. Instead the filter K will only tend to decorrelate (or decorrelate up to a given scale). For this reason it is most natural to formulate the problem in terms of a variational principle with an "energy" or cost functional, $E\{K\}$, that grades different kernels according to how well they decorrelate the output. Any constraints on this process are easily incorporated as penalty terms in the energy functional. To find the correct energy functional for decorrelation one may use Wegner's theorem (Bodewig 1956), which states that

$$\det\langle O(\mathbf{x}_i)O(\mathbf{x}_j)\rangle \leq \prod_i \langle O^2(\mathbf{x}_i)\rangle \tag{2.1}$$

with equality if and only if the matrix $\langle O(\mathbf{x}_i)O(\mathbf{x}_j)\rangle$ is diagonal. This means that decorrelation can be achieved by keeping $\det\langle O(\mathbf{x}_i)O(\mathbf{x}_j)\rangle$ fixed and minimizing $\prod_i\langle O^2(\mathbf{x}_i)\rangle$. One reason for keeping $\det\langle O(\mathbf{x}_i)O(\mathbf{x}_j)\rangle = \det(K^TRK)$ fixed is that this ensures a reversible transformation, since it is the same as requiring $\det(K^TK) > 0$. [Here we are treating the kernel as a matrix $K_{ij} \equiv K(\mathbf{x}_i - \mathbf{x}_j)$.]

Actually, there are a couple of mathematical steps that lead to a simpler energy functional. First, with the assumption of translation invariance we can minimize $\langle O^2(\mathbf{x}_0)\rangle$ for one ganglion cell at location x_0 instead of $\prod_i\langle O^2(\mathbf{x}_i)\rangle$. Again by translation invariance, this is equivalent to minimizing the explicitly invariant expression $\sum_i\langle O^2(\mathbf{x}_i)\rangle = \mathrm{Tr}(KRK^T)$. Finally, it is more convenient to hold fixed $\log\det(K^TK)$ rather than $\det(K^TK)$. Thus[4]

$$E\{K\} = \mathrm{Tr}(KRK^T) - \rho\ \log\ \det(K^TK) \tag{2.2}$$

ρ is a lagrange multiplier used to fix $\det(K^TK)$ to some value, but since we do not know this value we will subsequently treat ρ as a parameter penalizing small $\det(K^TK)$.

[4]We should point out that the decorrelating filter K that minimizes 2.2 is not the usual Karhunen–Loeve transform which would be the Fourier transform for translationally invariant R. This KL transform gives a nonlocal, nontranslationally invariant K.

To find the kernel K that minimizes equation 2.2, it is best to work in frequency space, where traces such as $\mathrm{Tr}(KRK^T)$ become integrals over frequencies. Also, the second term in equation 2.2 can be converted to an integral, by first using the matrix identity $\log\det(K^TK) = \mathrm{Tr}\log(K^TK)$. The equivalent energy functional becomes

$$E\{K\} = \int d\mathbf{f}\, |K(\mathbf{f})|^2\, R(\mathbf{f}) - \rho \int d\mathbf{f} \log |K(\mathbf{f})|^2 \tag{2.3}$$

which when varied with respect to $K(\mathbf{f})$ gives

$$|K(\mathbf{f})| = \sqrt{\frac{\rho}{R(\mathbf{f})}} \tag{2.4}$$

With Field's $R(\mathbf{f}) \sim 1/|\mathbf{f}|^2$, this gives the whitening filter $K(\mathbf{f}) \sim \sqrt{\rho}|f|$.

Having arrived at the energy functional [equation 2.2 (or 2.3)] as the one that produces decorrelation, it is now straightforward to explain its information theoretic interpretation. Minimizing the first term in equation 2.2 is equivalent (see Atick and Redlich 1990) to minimizing the sum of bit entropies $\sum_i H_i = -\sum_i \int dO_i P(O_i) \log[P(O_i)]$, where $P(O_i)$ is the probability density for the ith ganglion cell output $O_i \equiv O(\mathbf{x}_i)$. The second term in equation 2.2 is the change in entropy H (including correlations, not just bit entropy) due to the retinal transformation, so requiring this term to vanish would impose the constraint that no information is lost — this is related to requiring reversibility, although it is stronger. Therefore minimizing E in equation 2.2 has the effect of reducing the ratio of bit entropy to true entropy: $\sum_i H_i/H$, which is what we mean here by redundancy. Minimizing this ratio reduces the number of bits carrying the information H; technically, it reduces all but the first order redundancy. Also, one can prove that $\sum_i H_i \leq H$ with equality only when the $O(\mathbf{x}_i)$ are statistically independent, so minimizing this ratio produces statistically independent outputs.

2.2 Introducing the Noise. Since here we are primarily interested in testing redundancy reduction, we take a somewhat simplified approach to the problem with noise. As discussed earlier, instead of doing a full-fledged information theoretic analysis (as in Atick and Redlich 1990), we work in a formalism where the signal is first low-pass filtered to eliminate noise. The resulting signal is then decorrelated as before. Actually, since we will be comparing with real data, we have now to be more explicit about the stages of processing that we believe precede the decorrelation stage.

In Figure 2 we show a schematic of the signal processing stages that we assume take place in the retina. First, images from natural scenes pass through the optical medium of the eye and in doing so their image quality is lowered. It is well known that this effect can be taken

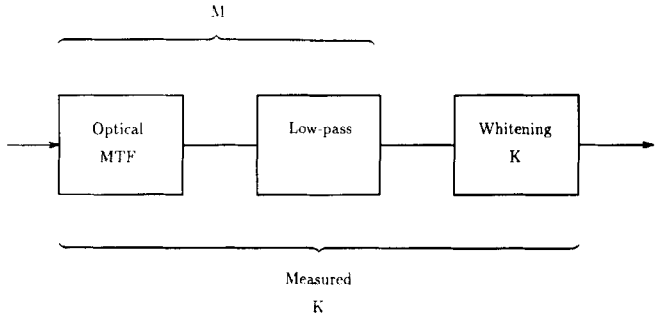

Figure 2: Schematic of the signal processing stages assumed to take place in the retina.

into account by multiplying the images by the optical *modulation transfer function* or MTF of the eye, a function of spatial frequency that is measurable in purely non-neural experiments. In fact, an exponential of the form $\exp[-(|f|/f_c)^\alpha]$, for some scale f_c characteristic of the animal (in primates $f_c \sim 22$ c/deg and $\alpha \sim 1.4$) is a good approximation to the optical MTF. The resulting image is then transduced by the photoreceptors and is low-pass filtered to eliminate input noise. Finally, we assume that it is decorrelated. In this model, the output–input relation takes the form

$$O = K \cdot [M \cdot (L + n) + n_0] \tag{2.5}$$

where the dot denotes a convolution as defined in equation 1.2. $n(\mathbf{x})$ is the input noise (such as quantum noise) while $n_0(\mathbf{x}_i)$ is some intrinsic noise that models postreceptor synaptic noise. Finally, M is the filter that takes into account both the optical MTF as well as the low-pass filtering needed to eliminate noise. An explicit expression for M will be derived below.

With this model, the energy functional determining the decorrelation filter K is

$$E\{K\} = \int d\mathbf{f}\ |K(\mathbf{f})|^2\ \left\{M^2(\mathbf{f})[R(\mathbf{f}) + N^2] + N_0^2\right\} - \rho \int d\mathbf{f} \log |K(\mathbf{f})|^2 \tag{2.6}$$

where $N^2(\mathbf{f}) \equiv \langle |n(\mathbf{f})|^2 \rangle$ and $N_0^2(\mathbf{f}) \equiv \langle |n_0(\mathbf{f})|^2 \rangle$ are the input and synaptic noise powers, respectively. This energy functional is the same as that in equation 2.3 but with the variance $R(\mathbf{f})$ replaced by the output variance of O in equation 2.5.

As before, the variational equations $\delta E/\delta K = 0$ are easy to solve for K. The experimentally measured filter K_{exp} is then this variational solution, K, times the filter M:

$$|K_{\text{exp}}(\mathbf{f})| = |K(\mathbf{f})| \quad M(\mathbf{f}) = \frac{M(\mathbf{f})\sqrt{\rho}}{\left\{M^2(\mathbf{f})\,[R(\mathbf{f}) + N^2] + N_0^2\right\}^{1/2}} \tag{2.7}$$

An identical result can be obtained in space–time trivially by replacing the autocorrelator $R(\mathbf{f})$ and the filter $M(\mathbf{f})$ by their space–time analogs $R(\mathbf{f}, w)$ and $M(\mathbf{f}, w)$, respectively, with w the temporal frequency. However, we focus here on the purely spatial problem where we have Field's (1987) measurement of the spatial autocorrelator $R(\mathbf{f})$ of natural scenes: $R(\mathbf{f}) = I_0^2/|\mathbf{f}|^2$.

2.3 Deriving the Low-Pass Filter. In our explicit expression for K_{exp}, below, we shall use the following low-pass filter

$$M(\mathbf{f}) = \frac{1}{N}\left[\frac{1}{I_0}\,\frac{R(\mathbf{f})}{R(\mathbf{f}) + N^2}\right]^{1/2} e^{-(|\mathbf{f}|/f_c)^\alpha} \tag{2.8}$$

The exponential term is the optical MTF while the first term is a low-pass filter that we derive next. The reader who is not interested in the details of the derivation can skip this section without loss of continuity.

It is not clear in the retina what principle dictates the choice of the low-pass filter or how much of the details of the low-pass filter influence the final result. In the absence of any strong experimental hints, of the type that imply redundancy reduction, we shall try a simple information theoretic principle to derive an M: We will insist that the filter M should be chosen such that the filtered signal $O' = M \cdot (L + n)$ carries as much information as possible about the *ideal* signal L subject to some constraint. To be more explicit, the amount of information carried by O', about L, is the mutual information $I(O', L)$. However, as is well known (for L and n statistically independent gaussian variables, see Shannon and Weaver 1949) $I(O', L) = [H(O') - \text{Noise Entropy}]$, and thus if we maximize $I(O', L)$ keeping fixed the entropy $H(O')$ we achieve a form of noise suppression.

We can now formulate this as a variational principle. To simplify the calculation we assume gaussian statistics for all the stochastic variables involved. The output–input relation including quantization units, n_q, takes the form $O' = M \cdot (L + n) + n_q$. A standard calculation leads to

$$I(O', L) = \int d\mathbf{f} \log\left[\frac{M^2(R + N^2) + N_q^2}{M^2 N^2 + N_q^2}\right]$$

Similarly, one finds for the entropy $H(O') = -\int d\mathbf{f} \log[M^2(R + N^2) + N_q^2]$. The variational functional or energy for smoothing can then be written

as $E\{M\} = -I(O', L) - \eta H(O')$. It is not difficult to show that the optimal noise suppressing solution $\delta E/\delta M = 0$ takes the form

$$M = \left(\frac{N_q}{N}\right)\left(\frac{1}{\eta}\frac{R}{R+N^2} - 1\right)^{1/2}$$

with the parameter $\eta \sim N_q^2 I_0$ in order to hold $H(O')$ fixed with mean luminance. Actually, below we will be working in the regime where the quantization units are much smaller than the signal and noise powers and hence we can safely drop the -1 term in M_1 since the 1η term dominates for small N_q^2. We can also ignore any overall factors in M that are independent of $\mathbf{f}$. This then is the form that we exhibit in the first term in equation 2.8.

2.4 Analyzing the Solution. Let us now analyze the form of the complete solution 2.7, with M given in equation 2.8. In Figure 3 we have plotted $K_{\text{exp}}(\mathbf{f})$ (curve a) for a typical set of parameters. We have also plotted the filter without noise $R(\mathbf{f})^{-1/2}$ (equation 2.4) (curve b) and $M(\mathbf{f})$ (equation 2.8) (curve c). There are two points to note: at low frequency the kernel $K_{\text{exp}}(\mathbf{f})$ (curve a) is identically performing decorrelation, and thus its shape in that regime is completely determined by the statistics of natural scenes: the physiological functions M and N drop out. At high frequencies, on the other hand, the kernel coincides with the function M, and the power spectrum of natural scenes R drops out.

We can also study the behavior of the kernel in (equation 2.7) as a function of mean luminosity I_0. If one assumes that the dominant source of noise is quantum noise, then the dependence of the noise parameter on I_0 is simply $N^2 = I_0 N'^2$ where N' is a constant independent of I_0 and independent of frequency (flat spectrum). This gives an interesting result. At low frequency where K_{exp} goes like $1/\sqrt{R}$ its I_0 dependence will be $K_{\text{exp}} \sim 1/I_0$ (recall $R \sim I_0^2$) and the system exhibits a Weber law behavior, that is, its contrast sensitivity $I_0 K_{\text{exp}}$ is independent of I_0. While in the other regime — at high frequency — where the kernel asymptotes M with $N^2 > R$ then $K_{\text{exp}} \sim 1/I_0^{1/2}$ which is a De Vries–Rose behavior $I_0 K_{\text{exp}} \sim I_0^{1/2}$. This predicted transition from Weber to De Vries–Rose with increasing frequency is in agreement with what is generally found (see Kelly 1972, Fig. 3).

Given the explicit expression in equation 2.7 and the choice of quantum noise for N we can generate a set of kernels as a function of I_0. The resulting family is shown for primates in Figure 4. We need to emphasize that there are no free parameters here which depend on I_0. The only variables that needed to be fixed were the numbers f_c, α, ρ, and N' and they are independent of I_0. Also we work in units of synaptic noise n_0, so the synaptic noise power N_0^2 is set to one. We have superimposed on this family the data from the experiments of Van Ness and Bouman (1967) on human psychophysical contrast sensitivity. It does not take

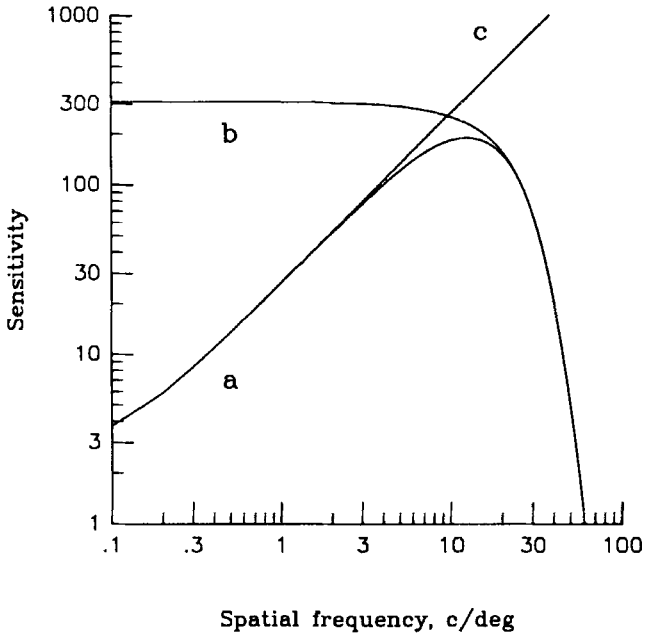

Figure 3: Curve a is the predicted retinal filter from equation 2.7 for a typical set of parameters, while curve b is $R(\mathbf{f})^{-1/2}$, which is the pure whitening filter. Finally, curve c is the low-pass filter M. The figure shows that at low frequencies curves a and b coincide and thus the system is whitening, while at high frequencies curves a and c coincide and thus the retinal filter is determined by the low-pass filter.

much imagination to see that the agreement is very reasonable especially keeping in mind that this is not a fit but a *parameter free* prediction.

3 Discussion

One major aim of this paper has been to answer the question, what does the retina know about its visual environment? Our initial answer comes from noting that the experimental ganglion cell kernel whitens the $|\mathbf{f}|^{-2}$ spatial power spectrum of natural scenes found in completely independent experiments by Field (1987). This shows that the retinal code has been optimized — assuming whitening as a design principle — for an environment with a $|\mathbf{f}|^{-2}$ spectrum. In other words, the retina knows at least one statistical property of natural scenes: the spatial autocorrelator.

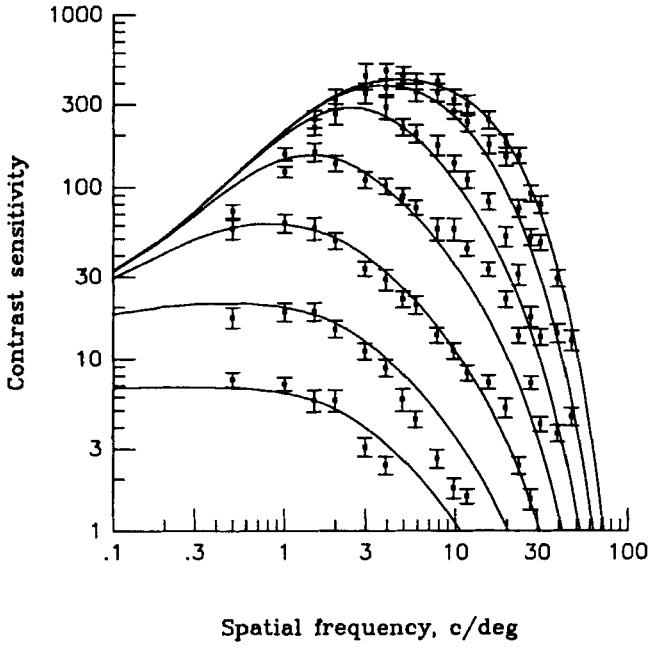

Figure 4: The family of solid curves are the predicted retinal filters (equation 2.7) at different I_0 separated by one log units, assuming that the dominant source of input noise is quantum noise ($N^2 \sim I_0$). No other parameters depend on I_0. The fixed parameters are $f_c = 22$ c/deg, $\alpha = 1.4$, $\rho = 2.7 \times 10^5$, $N' = 1.0$. The data are from human psychophysical contrast sensitivity measurements of Van Ness and Bouman (1967).

But what is useful about whitening the input signal? One possible answer is that whitening compresses the (photoreceptor) input signal so that it can fit into a channel with a more limited dynamical range, or capacity. Such a limitation may be a physical one in the retina such as at the bipolar cell input synapses or it may be in the ganglion cell output cable, the optic nerve (see also Srinivisan et al. 1982). Another possible explanation for the whitening is Barlow's idea that a statistically independent, or redundancy reduced representation is desirable as a *cortical* strategy for processing sensory data. From this point of view, the retinal filter is only performing the first step in reducing redundancy, by reducing second-order statistics (correlation). With this explanation, the capacity limitation is located further back in the brain, and may be best understood as an effective capacity limit, which is due to a computational

bottleneck, for example, the attentional bottleneck of $\sim$ 40 bits/sec. Of course, since redundancy reduction usually allows compression of a signal, there is no reason both explanations for whitening — physical bottleneck in the retina or computational bottleneck in cortex — must be mutually exclusive. Also, to paraphrase Linsker (1989), the brain may create physiological capacity limitations at one stage in order to force an encoding whose true utility is in its use as part of a larger strategy, such as Barlow's redundancy reduction.

There is, however, some evidence favoring the cortical redundancy reduction hypothesis: First, assuming a physiological bottleneck in the retina implies that the output code has a fixed and limited number of states available, and these are fewer than the number of states at the input. If one assumes that all of these outputs states are being used maximally at all luminosities, this produces a dependence on I_0 that does not match experiment. One finds that such a capacity limitation constraint predicts a Weber ($K \sim I_0$) type scaling with I_0 at all frequencies so long as the kernel is bandpass; this is contradicted by experiments that show a significant decrease in contrast sensitivity (Derrington and Lennie 1982), for example, at peak frequency, even while there is little change in the shape of the kernel. Second, some animals show bandpass (whitening) filtering even at very low luminosities where the input signal to noise is such that no capacity limitation is likely. Third, the ganglion cell bandpass characteristic is sharpened at later stages, such as in the LGN, and in monkeys some cortical cells have receptive fields very much like those of ganglion cells (Hubel and Wiesel 1974). Finally, some animals have orientation selective cells already in their retinas. This, together with the third point, suggests that whitening (giving bandpass filtering) is likely to be a first stage in a strategy of visual processing which is continued in the cortex, and which may also explain, for example, orientation selectivity.

To finally decide on the true purpose of the retinal whitening of natural scenes will require more experiments. In particular, to avoid some assumptions, it would be best to experimentally measure the correlation between ganglion cell outputs (also cortical cells) for an animal in its natural environment. Because of the need to suppress noise, as shown here, we would predict some correlation for nearby ganglion cells, but a much smaller correlation length for ganglion cells than for the natural luminance signal. Also, the stimulus must be the animal's natural environment, or at least have a $|\mathbf{f}|^{-2}$ spectrum, because of course any other type of input correlation will show up as output correlation.

Beyond such questions about the purpose or presence of decorrelation, we should stress that without considering the problem of noise one cannot fully explain the form of the experimental ganglion cell kernel. In fact, too much whitening of a signal that includes noise can be dangerous. This is an obvious point that has not always been appreciated. We find consideration of this need to suppress noise is the only other ingredient needed in order to explain an abundance of experimental data. It gives

an explanation of the relatively low peak frequency of the retinal filter in bright light. It also leads to the prediction of a bandpass to lowpass transition with decreasing mean stimulus luminance. In fact, our solutions predict an approximately Weber behavior at low frequencies, and assuming quantum noise, an approximately De Vries–Rose behavior at high frequencies.

The same property of our solutions that leads to the observed behavior with changing luminance also explains another set of experiments: a similar bandpass to lowpass transition is observed when the temporal frequency of the stimulus is increased. That is, the effect of lowering I_0 is predicted to be very close to the effect of raising temporal frequency. A more complicated relationship between color processing and changes in stimulus frequency is also predicted by our theory, as is the cone to rod transition. So a very large class of experimental observations can all be explained as the consequence of a single principle. They also, as mentioned, probe more specific properties of an animal's environment, so they further test the dependence of retinal processing on environment. All of these space–time–color–luminance interactions are explored in a separate paper (Atick et al. 1992).

Acknowledgment

Work supported in part by a grant from the Seaver Institute.

References

Atick, J. J., and Redlich, A. N. 1990. Towards a theory of early visual processing. *Neural Comp.* **2**, 308–320; and 1990. Quantitative tests of a theory of retinal processing: Contrast sensitivity curves. Report no. IASSNS-HEP-90/51.

Atick, J. J., Li, Z., and Redlich, A. N. 1992. Understanding retinal color coding from first principles. To appear in *Neural Comp.* 1992.

Barlow, H. B. 1989. Unsupervised learning. *Neural Comp.* **1**, 295–311.

Barlow, H. B., and Foldiak, P. 1989. *The Computing Neuron.* Addison-Wesley, New York.

Bodewig, E. 1956. *Matrix Calculus.* North-Holland, Amsterdam.

Derrington, A. M., and Lennie, P. 1982. The influence of temporal frequency and adaptation level on receptive field organization of retinal ganglion cells in cat. *J. Physiol.* **333**, 343–366.

De Valois, R. L., Morgan, H., and Snodderly, D. M. 1974. Psychophysical studies of monkey vision-III. spatial luminance contrast sensitivity tests of macaque and human observers. *Vision Res.* **14**, 75–81.

Field, D. J. 1987. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A* **4**, 2379–2394.

Goodall, M. C. 1960. Performance of a stochastic net. *Nature (London)* **185**, 557–558.

Hubel, D. H., and Wiesel, T. N. 1974. Sequence regularity and geometry of orientation columns in the monkey striate cortex. *J. Comp. Neurol.* **158**, 267–294.

Kelly, D. H. 1972. Adaptation effects on spatio-temporal sine-wave thresholds. *Vis. Res.* **12**, 89–101.

Linsker, R. 1989. An application of the principle of maximum information preservation to linear systems. In *Advances in Neural Information Processing Systems*, Vol. 1, D. S. Touretzky, ed., pp. 186–194. Morgan Kaufmann, San Mateo, CA.

Shannon, C. E., and Weaver, W. 1949. *The Mathematical Theory of Communication*. The University of Illinois Press, Urbana.

Srinivisan, M. V., Laughlin, S. B., and Dubs, A. 1982. Predictive coding: A fresh view of inhibition in the retina. *Proc. R. Soc. London Ser. B* **216**, 427–459.

Van Ness, F. L., and Bouman, M. A. 1967. Spatial modulation transfer in the human eye. *J. Opt. Soc. Am.* **57**, 401–406.

Received 15 July 1991; accepted 3 October 1991.