INSTITUTE OF PHYSICS PUBLISHING NETWORK: COMPUTATION IN NEURAL SYSTEMS

Network: Comput. Neural Syst. **12** (2001) 395–407 www.iop.org/Journals/ne PII: S0954-898X(01)26724-4

Optimal nonlinear codes for the perception of natural colours

Tassilo von der Twer[1] **and Donald I A MacLeod**[2]

[1] Bergische Universitat Wuppertal, D-42097 Wuppertal, Germany
[2] University of California at San Diego, La Jolla, CA 92093-0109, USA

E-mail: twer@wppc16.physik.uni-wuppertal.de and dmacleod@ucsd.edu

Received 1 December 2000

Abstract

We discuss how visual nonlinearity can be optimized for the precise representation of environmental inputs. Such optimization leads to neural signals with a compressively nonlinear input–output function the gradient of which is matched to the cube root of the probability density function (PDF) of the environmental input values (and not to the PDF directly as in histogram equalization). Comparisons between theory and psychophysical and electrophysiological data are roughly consistent with the idea that parvocellular (P) cells are optimized for precision representation of colour: their contrast-response functions span a range appropriately matched to the environmental distribution of natural colours along each dimension of colour space. Thus P cell codes for colour may have been selected to minimize error in the perceptual estimation of stimulus parameters for natural colours. But magnocellular (M) cells have a much stronger than expected saturating nonlinearity; this supports the view that the function of M cells is mainly to detect boundaries rather than to specify contrast or lightness.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

What nonlinear input–output function for a graded neural signal allows that signal to represent environmental stimuli best, under the constraint of a definitely restricted neural response range? This is the general problem we consider. For the sake of concreteness, we consider it in the particular context of colour vision. In colour as in sensory processing in general, a quantitative input has to be represented quantitatively, and different system designs may differ in the precision with which they represent the input. The essence of the problem can be quickly appreciated with the help of a crude analogy (which ultimately proves to be too crude): a *slicing* of the input space. Imagine that successive, reliably different, levels of each output

0954-898X/01/030395+13$30.00

signal define bins or slices into which the system can reliably divide its inputs. The total number of these slices is fixed for a given output signal by the range of output signal strength and by the degree of intrinsic randomness ('noise') present in the output signal. However, by suitable choice of the mapping from input to output these slices can be arranged to cover part or all of the input range consecutively in any desired manner, with any desired variation in slice thickness across the input range.

The visual system's problem, which we take up in section 4 below, is: what is the best way to slice the input space (e.g. the space of colour stimuli) using slices that can vary in thickness (and also, in multi-dimensional cases such as that of colour, in direction)? Intuitively it makes sense to make the slices thinner in parts of the input space where inputs are in practice densely concentrated, even though this must be done at the expense of fine discrimination within ranges of stimulus value that occur only seldom. For example, natural colours cluster around white (as discussed in section 2 below), so it makes sense to slice colour space more finely around white than in the outlying regions. Laughlin (1981, 1983) investigated this problem. His answer is that the thickness of the slices should be chosen by the infomax criterion, maximizing the mutual information between input and output. The infomax criterion leads to the arrangement of slice thickness known as histogram equalization, in which slice thickness (or histogram bin width) varies in inverse proportion to the input probability density (so that all bins are equally populated). This arrangement minimizes the probability that pairs of randomly chosen inputs will share the same bin and be confused. If we consider the bins or slices to correspond to equal intervals in the output signal, histogram equalization is achieved by making the input–output function $y(x)$ linear with the integral of the probability density $p(x)$ of the inputs. Thus near white, the 'output' signals representing colour should have a steep gradient with the 'input' photoreceptor excitations, because the probability density of environmental colours is high in this region of the input space. And for saturated colours, the gradients of the input–output functions should be correspondingly shallow.

Although histogram equalization minimizes confusion rates, different considerations arise when the output signal is contaminated by random variation (noise). In the presence of noise, *any* stimulus is confused with a slightly different one each time it is presented, and it is important to minimize the magnitude of these errors. Here the slicing metaphor becomes inexact, the infomax criterion becomes too simple, and histogram equalization is no longer optimal. The subject of this paper is the optimization of nonlinearity for precise representation of inputs in the presence of noise. We assume initially that the noise originates at the output, after the nonlinear encoding that is to be optimized, but we discuss how this assumption can be relaxed.

In section 5, we extend the treatment to systems in which multiple output signals are available for coding a single input dimension. Electrophysiological studies of sensory neurons commonly reveal an opponent code. The clearest instance of this is in the representation of colour, where individual neurons at stages following the photoreceptor stage (Derrington *et al* 1984, DeValois and DeValois 1975) are excited by certain parts of the spectrum and inhibited by others. This happens because these neurons are excited by one or two types of spectrally selective photoreceptor and inhibited by others. A somewhat paradoxical consequence of this encoding scheme is that the postreceptoral neurons are poorly responsive to the physiologically and phenomenally neutral stimuli that are most abundant in the environment. Likewise in the case of motion detection, directionally selective neurons respond poorly to static or nearly static stimuli, with inhibitory or zero response for motion in directions opposed to the preferred direction. Motion and colour thus adopt what we will call a '*split range*' code, where an input continuum such as colour, or (signed) input velocity, is divided at a physiological null point (near-white, or zero velocity), and where separate neurons respond to inputs on opposite sides of that null point. We show that by adopting a split range code the visual system improves

the average precision in its representation of natural inputs in the presence of neural noise introduced at the output. Finally (in section 6) we compare the optimal form of the split-range nonlinearity with experimental estimates from psychophysics and from neurophysiology, finding fair agreement in some comparisons and instructive disagreement in others.

2. The distribution of natural surface colours

We begin by considering the probability distribution of chromatic stimulus values. Our analysis will depend particularly on the dispersion of these distributions. The stimulus values we consider depend on the photoreceptor spectral sensitivities. Because the cone spectral sensitivities are broad, with substantial overlap, the ratios of their excitations, on which perception of colour depends, are generally not very large and the excitations of different cone types are strongly correlated. This is especially striking for natural stimuli, which themselves have broad spectral distributions. The correlation between the cone excitations is for that reason higher for natural colours than for spectral lights. We have mainly relied on two sets of measurements on natural colours (Brown 1994, Ruderman *et al* 1998). We derived human L, M and S (short-wavelength sensitive) cone excitations from the measured surface reflectances by integrating their crossproducts with the cone sensitivities of Stockman *et al* (1993). For 574 haphazard samples of natural colours measured in San Diego by Richard O Brown (in preparation), the correlation between L and M cone excitation is 0.985. These samples were not selected in such a way as to be representative in any sense, and they may have included a disproportionate number of highly chromatic surfaces since these would more readily have caught the investigator's eye. Ruderman, Cronin and Chiao (Ruderman *et al* 1998) obtained spectral reflectance estimates, pixel by 3 min arc pixel, for 12 entire views of natural environments; for this data set comprising nearly 200 000 pixels, the correlation between L and M cone excitation is even higher at 0.9983.

Surface luminance is given simply by the summed excitations of L and M cones, which we denote here simply by L and M (Eisner and MacLeod 1980, Lennie *et al* 1993). The standard deviation of $\log 10(L + M)$ in Brown's data set is 0.46, which corresponds to a factor of three in luminance; for Ruderman *et al* data the value is 0.24, a little less than a factor of two. The purely chromatic variations are conveniently indexed by two axes that represent colours without regard for their intensity, each chromatic axis representing the excitation of a specific photoreceptor per unit luminance. First, $r = L/(L+M)$ forms the (roughly speaking) 'red/green' axis of a photoreceptor-based chromaticity diagram (Luther 1927, MacLeod and Boynton 1979) and is proportional to L cone excitation per unit luminance. As the high correlation between L and M implies, the standard deviation of r is far smaller than that for luminance: only 7.5%, or 0.03 in the decimal logarithm for Brown's data, and only 1% for the entire scenes of Ruderman *et al*. For the remaining chromatic axis we adopt the luminance-normalized S cone excitation, $b = S/(L + M)$, which is low for yellows and very high for violets. The standard deviation of b is an order of magnitude greater than for r, though still lower than the one for luminance: 0.39 in $\log 10(b)$ or a factor of about 2.5 in b for Brown's data, and 0.114 in $\log 10(b)$ or 30% in b for the data of Ruderman *et al* (crosses in figure 1 show the histogram for $\log 10(b)$). In both these data sets, the average of the natural stimuli is near to, but slightly yellower than, the equal energy white, which plots at a horizontal coordinate of zero in figure 1.

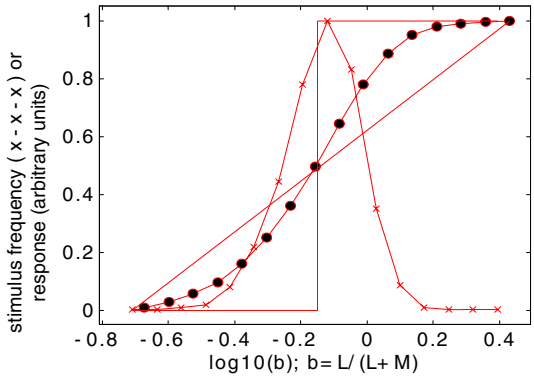

Figure 1. Crosses: frequency distribution of log 10(b) for Ruderman *et al* 's set of natural colours; b specifies S cone excitation per unit luminance, i.e. $b = S/(L + M)$. Whites and greens are near the middle of the distribution, with equal energy white represented at log 10(b) $= 0$. To the right lie bluish colours; to the left, generally yellowish or reddish ones. Candidate input–output functions: pleistochrome (circles), compared with linear and stepwise alternatives (curves).

3. Best directions for slicing colour space

It is obviously advantageous to limit the range of neurally represented stimulus values to the range that naturally occurs. Because the dispersion of cone excitations from natural colours is so much narrower in the red–green direction than in the luminance direction, this calls for a recoding of colour into luminance and chromatic signals, using signals that (roughly speaking) slice cone excitation space along its diagonals.

Others (Atick *et al* 1992, Buchsbaum and Gottschalk 1983, Fukurotani 1982) have justified such a recoding with the claim that the optimal directions for slicing colour space are those of the principal components of the distribution of photoreceptor excitations: neural signals encoding these principal components will be uncorrelated. Note, however, that in the context of a strictly linear analysis, where the neural signals are not assumed to be limited in range, a recoding in terms of the principal components of the distribution of photoreceptor excitations offers no advantage for colour discrimination: recoding does not change the effect of input noise, and the influence of noise originating at the output can be made negligible simply by making the gain (the slope of the function relating output to input) sufficiently large. In terms of the slicing analogy, a given pattern of sufficiently fine slices with strictly linear spacing (with different spacing, but unlimited range for each axis), can be rotated in any direction without much affecting the number of boxes that are contained within a given ellipsoidal distribution of environmental stimuli. Moreover, while principal components analysis can specify a set of *directions* for slicing the input space, it cannot indicate a preferred origin of the coordinate system in cone excitation space, and for that reason can give no rationale for opponent codes.

In this paper we suggest that the benefits of adopting a diagonal slicing of cone excitation space result from the limited range of neural signals; this nonlinearity limits the number (and thickness) of the slices that are created by each type of neuron. We present a rationale for opponent neural codes which takes as its starting point the idea that the *thickness*, and not the direction, of the slices should satisfy a principled requirement: specifically, the arrangement of slice thicknesses should provide the most precise representation of natural colours. This criterion allows determination of an optimum nonlinear code, subject to the constraint of a limited output range. The optimal code with a single neuron per stimulus dimension is a sigmoid. For two neurons per dimension it is a split range code employing rectifying opponent cells.

4. Best choice of slice thickness: the pleistochrome

Initially we assume, for simplicity, that variability originating at the retinal output predominates over sources of error at earlier stages. A rationale for this is that the optic nerve constitutes an informational bottleneck for vision, where the number of nerve fibres and of nerve impulses is relatively limited: much less, for instance, than the number of absorbed photons at daylight light levels (Barlow 1965). Hence relatively large errors are introduced by random fluctuations in the optic nerve fibre impulse counts (Bialek and Rieke 1992, Lee and others 1993).

When discrimination is limited by noise introduced into an output signal, good discrimination around any input value can be achieved by making the gradient of the input–output function at the relevant point as steep as necessary. But the limited total available range in firing rate means that (for monotonic response functions) an increase in gradient at one point has to be paid for by a decrease at other points within the input range, and hence by reduced discrimination at those points. By suitable choice of a nonlinear response function, relative discriminative precision can be distributed in any desired way over the range of input values. So which choice is best? For example, what input–output function gives the smallest RMS error in the estimated input, averaged over all naturally occurring cases?

Clearly it would be inefficient to make the code linear (with constant gradient and constant discrimination) over an input range greater than what is naturally encountered. An example is the straight line in figure 1, where the crosses represent the natural distribution of $\log 10(b)$ from the data of Ruderman *et al*: this sacrifices discrimination among frequently occurring stimuli in order to preserve discrimination in ranges where it is never needed. Also inefficient is the opposite extreme: a response function that steps abruptly from minimum to maximum firing rate at the peak of the distribution of natural colours (step function, figure 1). This provides exquisite discrimination between bluish and yellowish colours, but completely fails to distinguish among blues or among yellows. The best choice will be an intermediate one (circles, figure 1): a gently curving sigmoid, which retains some discrimination in the tails while slicing colour space most finely at the peak. More specifically, the optimal response function has a gradient matched to the cube root of the probability density function of the input distribution, as we now show.

Denote the input by x (which might be, for instance, some weighted sum of cone excitations), and the noisy output (e.g. retinal ganglion cell firing rate) by the random variable $y = g(x) + n$, where $g(x)$ is a nonlinear function of x, and n is a random noise term with zero mean and standard deviation $\sigma(n)$. The noise impairs discrimination because the system will confuse different values of x when they happen to elicit the same noisy output y. The range of confusion (the slice thickness at the input) can be measured by the standard deviation of the equivalent input noise, or by its square, the mean squared error at the input. It depends both on the range of the output error and on the gradient of the input–output response function $g(x)$. As we will show, under a simplifying assumption the mean squared error in the represented input value is proportional to the output variance $\sigma^2(n)$, and inversely to the square of $g'(x)$, the derivative of the function relating input to output.

The input that would elicit output y in the absence of noise is given by $g^{-1}(y)$. Treating this as the perceptual estimate of the input[3], a perceptual error equal to $g^{-1}(y) - x$ ensues

[3] This estimate is not necessarily the optimal one: for instance, knowledge of the environmental probability density function (PDF) of x might dictate a different choice. But the decoding of from y^* to x^*y is not important for our argument. The confusions that occur between different inputs, and that determine MSE(x) below, are invariant under strictly monotonic transformations of the noise-contaminated signal y^*. Our conclusions argument could therefore be developed without introducing the idea that perception involves an estimate of the input, by considering only instead the distributions of stimuli that are confused, and attaching a cost to variance in that distribution.

when a noise perturbation at the output causes an input x to elicit the output y. We consider first the problem of minimizing the mean square of this error (other optimization criteria such as minimization of absolute error are discussed below). With the simplifying assumption that $g(x)$ can be treated as linear over the limited range of confusion, we have

$$g^{-1}(y) = g^{-1}(g(x) + n) = x + n/g'(x).$$

Since the mean of n is zero, the mean of $g^{-1}(y)$ is simply the true input x, and the mean squared error due to noise for a given input x, which we call MSE(x), is the variance of $g^{-1}(y)$. This is proportional to the output variance $\sigma^2(n)$, which we initially assume to be fixed and independent of x:

$$\text{MSE}(x) = \frac{\sigma^2(n)}{(g'(x))^2}. \tag{1}$$

The constant of proportionality is the inverse square of $g'(x)$, the gradient of the input–output function $g(x)$ at x: where $g(x)$ is steep, a smaller range of inputs suffices to span the range of the noise that is added at the output.

Now denote the environmental probability distribution of x (for all stimuli encountered, or of interest) by $p(x)$. The mean squared error to be minimized, MSE, is the error due to noise, MSE(x), averaged for all environmental inputs x, that is, the probability-weighted integral of MSE(x) over x:

$$\text{MSE} = \int p(x)\text{MSE}(x)\,\mathrm{d}x.$$

To see how this can be minimized, consider the effect of small variations in the response gradient or incremental gain $g'(x)$ around its optimal value. An increase in $g'(x)$ at one value of x has to be paid for with an equal decrease at other values of x, and in the optimal condition the effects of such complementary changes must cancel, that is $p(x)\,\mathrm{d}(\text{MSE}(x))/\mathrm{d}g'(x)$ must be independent of x. Thus

$$\mathrm{d}(\text{MSE}(x))/\mathrm{d}(g'(x)) = \alpha/p(x) \tag{2}$$

with a constant α.

Since from (1)

$$\mathrm{d}(\text{MSE}(x))/\mathrm{d}(g'(x)) = -2(\sigma^2(n))/(g'(x))^3$$

the optimal condition occurs when the gradient is matched to the cube root of the PDF:

$$g'(x) = \beta p^{1/3}(x). \tag{3}$$

The scaling factor β absorbs the output variance factor $\sigma^2(n)$; it serves only to define units of measurement for y and can be set to 1 if those units are not defined independently[4]. Hence

$$g(x) = \int_{-\infty}^{x} \beta p^{1/3}(u)\,\mathrm{d}u. \tag{4}$$

This result (4) parallels the solution to the vector quantization result from the engineering literature (Panter and Dite 1951, Pratt 1978). Moreover, the exponent 1/3 occurs in the slicing patterns of self-organized Kohonen networks (Ritter *et al* 1989).

But the general form of our problem, in which the noise variance at the output is dependent on the mean output $g(x)$, is not accounted for in this stream of literature. The more general problem is easily dealt with, however: if $\sigma(n)$ is regarded as a function of the mean output $g(x)$, let us say $h(g(x))$, then $g'(x)$is replaced in the left-hand side of equation (3) by $g'(x)/h(g(x))$,

[4] Convergence of the integral above now requires that $p(x)$ decrease more rapidly than $1/x^3$ for large $|x|$.

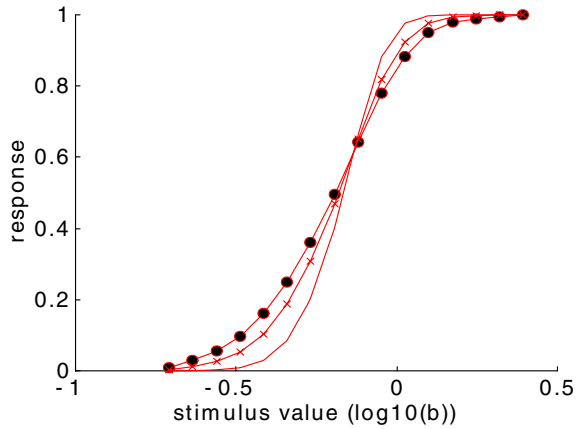

Figure 2. Pleistochromes based on minimization of mean squared error (circles), or of mean absolute error (crosses), compared with the function that maximizes mutual information through histogram equalization (plain curve).

yielding a differential equation for g which may be solved by separation of variables. The underlying idea is simple: some monotonic function of the output will have a standard deviation independent of its mean (in the case of Poisson noise for example, this will be the square root of the output). The optimal condition occurs when that function of the output conforms to equation (4). With the input–output nonlinearity thus optimized, h has no effect at all (except for an overall sensitivity factor) on the variation of equivalent input noise with input x: the range of confusion around x varies as $p^{1/3}(x)$ in all cases.

We refer to the optimal response function of equation (4) (illustrated by the circles of figure 1) as the *pleistochrome*, from the Greek *pleistos* meaning 'most'. As figure 2 shows, the pleistochrome is roughly similar to the cumulative distribution of x prescribed by histogram equalization (Laughlin 1981, 1983). But it is wider than that function by a factor of about the square root of three, allowing much better discrimination in the margins of colour space. The infomax criterion that leads to histogram equalization is not derived from a noise-based theoretical framework, and it lacks an acceptable rationale in terms of minimization of random error or the associated cost. A straightforward generalization of the above analysis shows that to minimize the nth power of error, the gradient should match the $(n + 1)$th root of the environmental PDF $p(x)$; histogram equalization is therefore the limit approached by letting the exponent on error decline from 2 (minimization of mean squared error) toward zero (minimization of errors without regard for their size).

With input noise present, equation (3) still guarantees least mean squared error, because the input noise (if uncorrelated with the output noise) simply adds an optimization-irrelevant constant to the mean squared error that must be minimized. A more awkward complication is that if the input noise is non-uniform, the cost of added input error also becomes non-uniform. Such situations can be dealt with by suitably weighting the errors for different values of x. A procedure like the one advocated for non-uniform output noise suffices to do this: some monotonic function $f(x)$ of the input x will exist such that estimation errors of equal absolute magnitude in $f(x)$ are equally undesirable. For instance, $f(x)$ would be a logarithmic function of x, if Weber's law described the error cost. Equation (3) specifies the optimum dependence of output firing rate on this function $f(x)$, and thence on x. Figure 1 illustrates this: the histogram shown is for the logarithm of the b chromaticity coordinate, and the continuous

curve is the response function that minimizes error in the estimate of $\log(b)$ rather than in the absolute value of b.

Indeed, the greater the input noise, the more general the validity of equation (3) as a prescription for optimal nonlinearity. If noise sources inherent in the input predominate over errors originating at the output, as happens in the visual system at low light levels (Barlow *et al* 1971, Baylor 1987, Donner 1992), a relatively small output-derived mean squared error is simply added to the relatively large mean squared error inherent in the input itself to give the total mean squared error in the perceptually estimated input. Then, *any* useful measure of average error or of the associated cost—be it mean square, mean absolute error or any other continuous function of mean squared error—will be approximately linear with the *mean square* error contributed from sources of variation at the output. Denote the mean 'error' or the associated mean cost that prevails for a particular stimulus value x—the quantity whose integral over x is to be minimized—by $\mathrm{MC}(x)$, which we take to be some fixed continuous function of the mean squared error $\mathrm{MSE}(x)$; we also suppose that that function always has a linear term in its Taylor series expansion about $\mathrm{MSE}(x)$ and that the contribution of output noise to MSE is always small enough to allow neglect of higher derivatives. Then $\mathrm{MC}(x)$ is always approximately linear with $\mathrm{MSE}(x)$ within the range of influence of the output noise (and is also linear with the output-noise-derived component of $\mathrm{MSE}(x)$).

When input noise is uniform, both $\mathrm{MSE}(x)$ and the gradient of $\mathrm{MC}(x)$ with respect to $\mathrm{MSE}(x)$ are independent of x, and to minimize the integral of $\mathrm{MSE}(x)$ over x is also to minimize the integral of $\mathrm{MC}(x)$. When input noise is not uniform, the gradient of $\mathrm{MC}(x)$ with respect to $\mathrm{MSE}(x)$ may depend on x. But this complication can again be dealt with by replacing x with a suitable monotonic function of x. (That is, for suitable x', the error cost $\mathrm{MC}(x')$ will have the same approximately linear dependence on $\mathrm{MSE}(x')$ for all x'.)

In this situation, therefore, the optimum input–output nonlinearity is determined for *any* cost function (quadratic or otherwise) by first transforming the input so that input-derived error (or more precisely, the cost per unit of added error in the estimated input) is independent of mean input, and then applying the MSE pleistochrome—the cube root construction of equation (3)—to derive the best dependence of output on that input. Thus the least-squared-error pleistochrome is uniquely robust to input noise.

5. Split range coding with two neurons: something for nothing

Our discussion has not yet yielded any rationale for opponent codes. In contrast, the sigmoidal nature of the pleistochrome (circles, figure 1) is incompatible with a null response to white. But when more than one neuron is available to represent a single stimulus dimension, new opportunities for coding are introduced. The non-opponent pleistochrome that optimizes encoding by a single neuron has an opponent counterpart when encoding is done by a pair of neurons. Two rectifying neurons, one red-excitatory and the other green-excitatory, and each with a purely compressive nonlinearity, can represent opposite halves of the red–green stimulus continuum with positive firing rates, but with no response to greenish or to reddish stimuli respectively. Such a representation is almost equivalent to that produced by a single neuron with sigmoidal nonlinearity (Marr 1974). As figure 3 shows, the responses of two such neurons (upward and downward triangles) correspond to the two halves of the single-neuron pleistochrome sigmoid (plain curve, figure 3), but with the left half flipped up so that the response gradients for the neuron responding in the left half of the stimulus range are simply reversed. This, however, uses only the upper half of the output range of each neuron. The optimal implementation of such a 'split range' code using two rectifying neurons is instead to set the cross point, x_0, at the 50% point

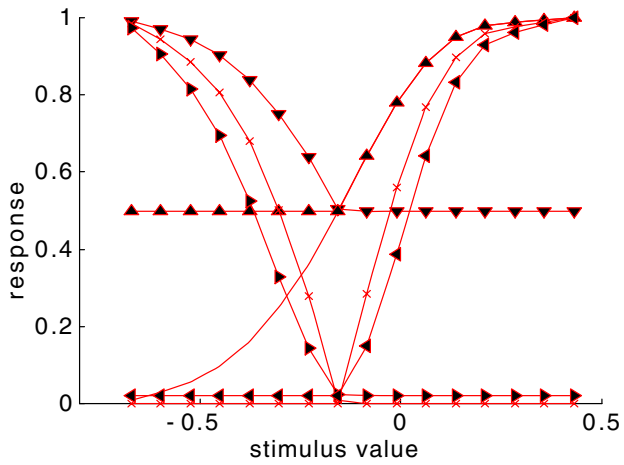

Figure 3. From single-neuron pleistochrome (continuous curve, from figure 1, circles) to 2-neuron split range opponent code (up and downward pointing triangles for two rectifying neurons). Crosses show how using the full output range for each neuron to cover only half the full input range ('split range' code) allows doubled response function gradients. Left and rightward pointing arrows show how the optimum response function (crosses) is modified when Poisson noise and spontaneous activity are assumed.

on the sigmoid, and then replace the integral from $-\infty$ in equations (4) or (5) with *twice* the integral upward or downward from x_0. In this way the appropriate segment of the input range can elicit the maximum possible response range from each neuron, and the gradients of the response functions are doubled everywhere (left and rightward triangles, figure 3).

By using two neurons in this way the visual system can therefore double the precision in its representation of the input in the presence of output noise. If, alternatively, the two neurons had each been endowed with the same sigmoidal nonlinearity that is optimal for single neurons, then averaging of their signals (on the generous assumption of independent noise) would have reduced average error by only the square root of two. Thus the net benefit of adopting the 'split range' code (as opposed to the alternative of similar neurons operating in parallel with optimal nonlinearity) is a *square root of two* reduction of average error.

The split range code may be viewed as a step from a purely analogue representation (the sigmoidal single-neuron pleistochrome) to a hybrid, analogue–digital one. In a still more efficient encoding of a stimulus dimension, a set of N neurons, each with m reliably distinct outputs, can represent m^N different stimulus levels by allocating successive sets of digits to the different neurons. But unlike the split range code this more fully digital representation requires individual neural outputs to depend discontinuously (and non-monotonically) on the input value. This greatly complicates the problem of readout (besides creating a risk of large errors). A split range code with $N > 2$, where the input range is divided into *many* consecutive segments each spanned by the graded firing range of one neuron, would avoid these problems and would still allow a simple centre-of-gravity or weighted sum of neural firing rates to represent the stimulus value. But this would have the drawback that some neurons must have high firing rates for frequent (e.g. near-white) stimuli. By reserving high firing rates for unusual stimuli the two-neuron split-range code can facilitate selective response to unusual inputs (Barlow 1972, Field 1987), and can reduce average firing rate and neurotransmitter release so as to lower the metabolic cost of perception.

6. Comparison of the pleistochrome with psychophysical and electrophysiological data

Our proposal that visual nonlinearity is optimized for discrimination among natural colours in the presence of output noise leads to a simple prediction: *the average error in visual discrimination or matching should be inversely proportional to $g'(x)$ in equation (4), and hence to the cube root of the natural probability density function $p(x)$*. As noted above, this prediction holds even for non-uniform output noise, provided that the nonlinearity $g(x)$ is appropriately optimized in each case.

Psychophysical data for evaluating this prediction are available. For achromatic intensity differences between comparison and standard stimulus patches, set in a common grey background, comparison error is doubled for standards that create a contrast of about 20% with the background; for isoluminant yellow–blue differences the cone contrast (for S cones in this case) at which error is doubled is again about 20%. But in the case of red–green isoluminant stimuli a standard L cone contrast of only 2% is enough to double the mean comparison error (Leonova and MacLeod, in preparation).

If the visual system adopts the encoding principle of the pleistochrome, we would therefore expect to find the probability density function $p(r)$ for natural colours dropping to $1/8$ of its peak value at r values that give a contrast of 2% with white, with half-heights at around 20% contrast for the other axes. This prediction is in one respect roughly upheld: in the distribution of environmental colours the ranges for luminance and for b are indeed wider than for r, by an order of magnitude or so (section 2). This could suggest that the operating ranges of the various relevant neurons are fairly well matched to the very diverse distributions of environmental inputs that they have to represent. In fact, however, the environmental distributions of b and (especially) of luminous reflectance are somewhat broader than would be expected for strict consistency with the pleistochrome principle. Why should the operating range of the visual system be narrower than 'optimal' in this way?

One answer appeals to visual adaptation. Since retinally stabilized images fade, vision is evidently sustained by the temporal transients that are generated when small eye movements scan spatial gradients in the image (Ditchburn 1973). It is therefore relevant to consider the distribution of the spatial *differences* in cone excitation in natural images. In the images of Ruderman *et al*, the differences in luminance, in b, and in r between adjacent 3 min arc pixels, have standard deviations of only 30, 13 and 0.6% respectively, and the PDFs drop to $1/8$ (where optimal differential sensitivity is halved) at about twice these values; these distributions are naturally somewhat tighter than those for the absolute values, owing to correlated variation in the values across the scene. The visual system can therefore advantageously employ an adaptively roving null point for the colour-opponent code (Krauskopf and Gegenfurtner 1992, Thornton and Pugh 1983), if its objective is the precise representation of local contrast (from which a metric representation of local brightness and colour can then be resurrected, perhaps in the way discussed by Land (1964) and Marr (1974)).

With a roving null point, the range of input values spanned by the neural response functions need only be wide enough to capture the relatively small *deviations* in the stimulus values from their time and space varying adapting levels—and the precision with which those values can be represented then becomes correspondingly greater. Local-contrast pleistochromes—contrast-response functions that lead to least error in the representation of pixelwise spatial differences in the images of Ruderman *et al*—are narrow enough to be fairly consistent with the cited psychophysical results in the case of the chromatic variables. For luminance, however, the contrast operating range implicit in the discrimination results remains narrower (by a factor of about three) than the theoretically optimal one.

A second limitation in our initial framework, also connected with the role of adaptation, may underlie this remaining discrepancy for luminance. We have taken for granted that the purpose of colour and lightness vision is to represent colours and lightnesses with the least possible error and allow these attributes of a surface to be estimated as precisely as possible. But of course, differences in lightness and colour are also indispensable for the detection of spatial features (Boynton 1980, Morgan *et al* 1992). For spatial vision it is desirable that local contrasts should be detected with the greatest possible sensitivity wherever they are present in the image. For this purpose, an all-or-none or categorical encoding scheme, with a step function nonlinearity at a small threshold offset from the adapting background stimulus is ideal (since the large, all-or-none spatial contrast signal resists obliteration by fluctuations in the output), and the graded response of the pleistochrome is not needed. Visual nonlinearity more step-like than the pleistochrome could therefore reflect a compromise in design between the conflicting requirements of surface identification and characterization on the one hand, and of detection of spatial features on the other. This is supported by the electrophysiological data to which we turn next.

The nonlinearity implied by the reviewed psychophysical data is quite severe. The gradient of the red–green response function, assumed in equation (4) to be directly proportional to differential sensitivity, is halved at a cone contrast of 2%. No physiological data suggest so severely compressed a response function for responses to chromatic stimuli: half-saturation L cone chromatic contrasts of around 10% appear to be more typical, for the red–green sensitive P cells of the parvocellular stream (Lee and others 1990). Thus, although the psychophysically estimated visual operating range along the red–green axis is efficiently matched to the range of environmental inputs, the physiological one apparently is not. Elsewhere (MacLeod and von der Twer, in preparation) we consider reasons for this, including the possibility that output noise increases with mean firing rate (as in a Poisson process), which makes the optimum physiological nonlinearity a more gentle one.

Along the achromatic axis of colour space, both the psychophysical operating range in cone contrast and the dispersion of the environmental inputs are (as noted above) at least tenfold greater than along the red/green one. The M cells of the magnocellular pathway (Livingstone and Hubel 1987) are a plausible substrate for achromatic discrimination. These cells, however saturate at very *low* achromatic cone contrasts, with half-saturation values of around 5% (Kaplan and Shapley 1986, Lee and others 1990, Wachtler *et al* 1996). The M cells thus deviate by an order of magnitude from the optimal behaviour embodied in the pleistochrome, and could not support the observed keen discrimination between test patches with relatively high achromatic contrast relative to their surrounds.

It is therefore likely that the M cells are not responsible for representing the achromatic attributes of surfaces in a continuous fashion, but serve instead as all-or-none detectors of spatial contrast and of change over time. Thus, although responses of P cells to red/green contrast are almost as nonlinear as those of M cells for achromatic contrast, comparison with natural scene statistics actually suggests different functions for these two systems. The red/green nonlinearity of the P cells is appropriate for a metric representation of their natural inputs, but the nonlinearity of the M cells is not. This is consistent with the common view that the luminance system (sometimes tentatively, though questionably, identified with the magnocellular pathway) is more concerned with form and with detection of spatial structure than are the chromatic ones (e.g. Boynton *et al* (1977), Gregory and Heard (1979), Livingstone and Hubel (1987)). The metric representation of the achromatic axis as well as the chromatic axes could be the job of the P cell system (Allman and Zucker 1990). In support of this idea, physiological investigations such as the ones cited have shown that the P cells have an almost linear response to achromatic contrast, roughly consistent with the pleistochrome for achromatic inputs.

Acknowledgments

We are grateful to Richard Brown, Anya Leonova, Dan Ruderman, Tom Cronin and C C Chiao for permission to use the data we cite from them (some if it not yet fully published). For useful conversations or comments we thank Brown, Leonova, Ruderman, Simon Laughlin, Markus Meister, and Harvey Smallman. A more complete account of this work, with more rigorous derivations, is in preparation. Supported by NIH grant EY01711.

References

Allman J and Zucker S 1990 Cytochrome oxidase and functional coding in primate striate cortex: a hypothesis *Cold Spring Harb. Symp. Quant. Biol.* **55** 979–82

Atick J J, Li Z P and Redlich A N 1992 Understanding retinal color coding from 1st principles *Neural Comput.* **4** 559–72

Barlow H B 1965 Optic nerve impulses and Weber's law *Cold Spring Harb. Symp. Quant. Biol.* **30** 539–46

——1972 Single units and sensation: A neuron doctrine for perceptual psychology? *Perception* **1** 371–94

Barlow H B, Levick W R and Yoon M 1971 Responses to single quanta of light in retinal ganglion cells of the cat *Vis. Res. Suppl.* **3** 87–101

Baylor D A 1987 Photoreceptor signals and vision. Proctor lecture *Invest. Ophthalmol. Vis. Sci.* **28** 34–49

Bialek W and Rieke F 1992 Reliability and information transmission in spiking neurons *Trends Neurosci.* **15** 428–34

Boynton R M 1980 Design for an Eye *Neural Mechanisms in Behavior* ed D McFadden (Berlin: Springer)

Boynton R M, Hayhoe M M and MacLeod D I A 1977 The gap effect: chromatic and achromatic visual discrimination as affected by field separation *Opt. Acta* **24** 159–77

Boynton R M and Kambe N 1980 Chromatic difference steps of moderate size measured along theoretically critical axes *Color Res. Appl.* **5** 13–23

Brown R O 1994 The world is not grey *Invest. Opthalmol. Vis. Sci. (Suppl.)* **35/4** 2165

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

Derrington A M, Krauskopf J and Lennie P 1984 Chromatic mechanisms in lateral geniculate nucleus of macaque *J. Physiol.* **357** 241–65

DeValois R L and DeValois K K 1975 Neural coding of color *Handbook of Perception* vol 5, ed E D Carterette and M P Friedman (New York: Academic) pp 117–66

——1993 A multi-stage color model *Vis. Res.* **33** 1053–65

Ditchburn R W 1973 *Eye-Movements and Visual Perception* (Oxford: Clarendon)

Donner K 1992 Noise and the absolute thresholds of cone and rod vision *Vis. Res.* **32** 853–66

Eisner A and MacLeod D I A 1980 Blue sensitive cones do not contribute to luminance *J. Opt. Soc. Am.* **70** 121–3

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

Fukurotani K 1982 Color information coding of horizontal cell responses in fish retina *Col. Res. Appl.* **7** 146–8

Gregory R L and Heard P 1979 Border locking and the Cafe Wall illusion *Perception* **8** 365–80

Kaplan E and Shapley R M 1986 The primate retina contains two types of ganglion cells, with high and low contrast sensitivity *Proc. Natl Acad. Sci. USA* **83** 2755–7

Kohonen Teuvo 1989 *Self-Organization and Associative Memory (Springer Series in Information Sciences, vol 8)* 3rd edn (Berlin: Springer)

Krauskopf J and Gegenfurtner K 1992 Color discrimination and adaptation *Vis. Res.* **32** 2165–75

Land E H 1964 *Sci. Am.* **52** 247

Laughlin S B 1981 A simple coding enhances a neuron's information capacity *Zeitschrift für Naturforschung* **36c** 910–12

——1983 Matching coding to scenes to enhance efficiency *Biological Processing of Images* ed O J Braddick and A C Sleigh (Berlin: Springer) pp 42–52

Lee B B, Pokorny J, Smith V C, Martin P R and Valberg A 1990 Luminance and chromatic modulation sensitivity of macaque ganglion cells and human observers *J. Opt. Soc. Am.* A **7** 2223–36

Lee B B, Wehrhahn C, Westheimer G and Kremers J 1993 Macaque ganglion cell responses to stimuli that elicit hyperacuity in man: detection of small displacements *J. Neurosci.* **13** 1001–9

Lennie P, Pokorny J and Smith V C 1993 Luminance *J. Opt. Soc. Am.* A **10** 1283–93

Livingstone M S and Hubel D H 1987 Psychophysical evidence for separate channels for the perception of form, color, movement, and depth *J. Neurosci.* **7** 3416–68

Luther R 1927 Auf dem Gebeit der Farbreizmetrik *Z. Tech. Phys.* **8** 540–58

MacLeod D I A and Boynton R M 1979 Chromaticity diagram showing cone excitation by stimuli of equal luminance *J. Opt. Soc. Am.* **69** 1183–6

MacLeod D I A and von der Twer T 2001 The pleistochrome: optimal opponent codes for natural colors *Color Perception: from Light to Object* ed R Mausfeld and D Heyer (Oxford: Oxford University Press)

Marr D 1974 The computation of lightness by the primate retina *Vis. Res.* **14** 1377–88

Morgan M J, Adam A and Mollon J D 1992 Dichromats detect color-camouflaged objects that are not detected by trichromats *Proc. R. Soc.* B **248** 291–5

Panter P F and Dite W 1951 Quantization distortion in pulse code modulation with non-uniform spacing of levels *Proc. IRE* **39** 44–8

Pratt W K 1978 *Digital Image Processing* (New York: Wiley)

Ritter H J, Martinetz T M and Schulten K J 1989 Topology-conserving maps for learning visuo-motor-coordination *Neural Networks* **2** 159–68

Ruderman D L, Cronin T W and Chiao C C 1998 Statistics of cone responses to natural images: implications for visual coding *J. Opt. Soc. Am.* A **15** 2036–45

Stockman A, MacLeod D I A and Johnson N E 1993 Spectral sensitivities of the human cones *J. Opt. Soc. Am.* A **10** 2491–521

Thornton J E and Pugh E N Jr 1983 Red/green color opponency at detection threshold *Science* **219** 191–3

Wachtler T, Wehrhahn C and Lee B B 1996 A simple model of human foveal ganglion cell responses to hyperacuity stimuli *J. Comput. Neurosci.* **3** 73–82

Zaidi Q 1997 Decorrelation of *L*- and *M*-cone signals *J. Opt. Soc. Am.* A **14** 3430–1