What does the eye see best?

Andrew B. Watson

Department of Psychology, Stanford University, Stanford, California 94305 and NASA-Ames Research Center, Moffett Field, California 94035, USA

H. B. Barlow & John G. Robson

Physiological Laboratory, Cambridge University, Cambridge CB2 3EG, UK

Our eyes see so much in such varied conditions that one might consider the question posed in the title to be meaningless, but we show here that, within the range that we have been able to test, there is a particular spatiotemporal pattern of light that is detected better than any other. At least two plausible theories of visual detection predict that a stimulus will be seen best (will have greatest quantum efficiency) when it matches the weighting function of the most efficient detector. We have measured quantum efficiency for detecting a wide variety of spatiotemporal patterns using foveal vision in bright light. The best stimulus found so far is a small, briefly exposed circular patch of sinusoidal grating having a spatial frequency of ~7 c deg^{-1}, drifting at ~4 Hz. We propose that this is the weighting function of the most efficient human contrast detector. We believe this answer to the question is unexpected and may have fundamental implications with regard to the mechanisms of visual perception.

A detector is a theoretical entity which maps each presentation of a visual signal into an internal representation on which the observer's decision is based. As a visual signal is distributed over space and time, any detector must collect together and appropriately combine information at different points in the image at different times. One important class of detectors performs this combination linearly: at each point in space and time, the signal is weighted by some coefficient and these values are added together. The weighting function specifying these coefficients completely characterizes the detector. If this weighting function is $w(x, y, t)$ and the signal is $I(x, y, t)$, then the response R of the detector, before the decision stage, is given by

$$R = \iiint w(x, y, t) I(x, y, t)\, \mathrm{d}x\, \mathrm{d}x\, \mathrm{d}t \qquad (1)$$

In the case of a visual neurone, the spatial weighting function corresponds to what is ordinarily called the receptive field sensitivity profile, and there is evidence for linear summation of small signals for many retinal ganglion cells[1-4] and neurones of the primary visual cortex[5,6].

The performance of a visual detector is ultimately limited by noise, part of which is inherent in the quantum fluctuations in the stimulus, and part of which may be due to the physical components of the detector. If this noise is independent from point to point and moment to moment, then there is a well known method for optimally combining information from the image. This optimum is achieved by weighting each point in proportion to its individual signal-to-noise ratio. Hence if the signal is $I(x, y, t)$, then the optimal linear weighting function is $kI(x, y, t)$ where k is any constant. This optimal detector is said to be 'matched' to the signal.

The best any visual detector can ever do is to meet the limit set by quantum fluctuations. This performance therefore represents an ideal, and performance of any detector relative to this ideal may be expressed as an efficiency. Specifically, suppose that the signal is some intensity perturbation, $I_s(x, y, t)$ that is added to an otherwise uniform steady background, $I_N(x, y, t) = I_N$. On signal trials, the intensity in the image is

$$= I_N[1 + c(x, y, t)] \qquad (2)$$

where $c(x, y, t)$ is the contrast at each point. Responses to signal-plus-noise and noise-alone trials are

$$\begin{aligned} R_{S+N} &= \iiint w(x, y, t) I_{S+N}(x, y, t)\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}t \\ &= I_N \iiint w(x, y, t)\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}t \\ &\quad + I_N \iiint w(x, y, t) c(x, y, t)\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}t \end{aligned} \qquad (3)$$

and

$$R_N = I_N \iiint w(x, y, t)\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}t \qquad (4)$$

The difference between these responses (the portion due to the signal alone) is

$$\begin{aligned} R_S &= R_{N+S} - R_N \\ &= I_N \iiint w(x, y, t) c(x, y, t)\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}t \end{aligned} \qquad (5)$$

Where c is small, as it is in these experiments, the variance of these responses will be approximately equal and given by the variance at each point (which is simply I_N when intensity is expressed in quanta deg^{-2} s^{-1}) multiplied by the square of the weight at that point and summed:

$$V = I_N \iiint w^2(x, y, t)\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}t \qquad (6)$$

However it is measured, performance will be monotonic with the ratio $R_S/\sqrt{V}$, usually called the signal-to-noise ratio or d'. As noted above, for the ideal detector, the weighting function is matched to the signal, in which case

$$\begin{aligned} d'_{\mathrm{ideal}} &= \left[I_N \iiint c^2(x, y, t)\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}t \right]^{1/2} \\ &= \sqrt{I_N E} \end{aligned} \qquad (7)$$

E is the contrast energy of the signal: the integral of the square of the contrast over all the dimensions in which it varies. Quantum efficiency (F) is the square of the ratio of empirical

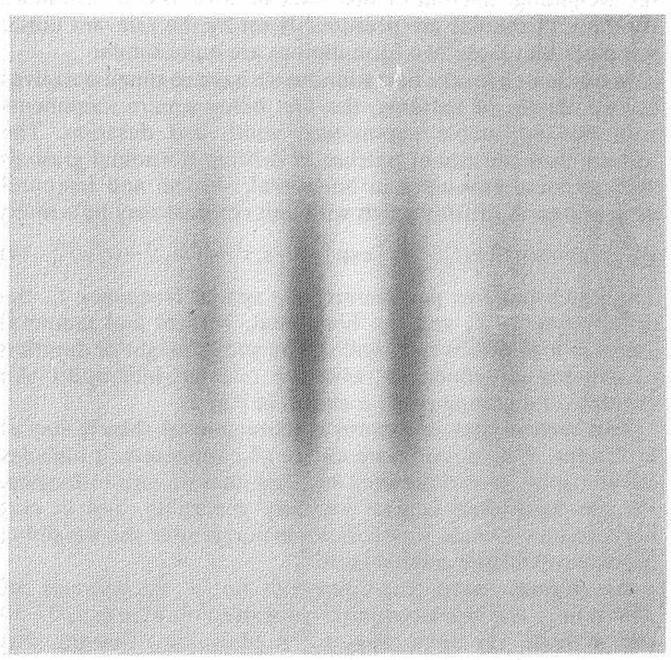

Fig. 1 An example of a grating patch. When viewed from a distance such that one cycle of the sinusoid subtends 1 deg of

and ideal d' values for a given contrast energy, or equivalently the ratio of ideal and empirical contrast energies, E_{ideal}/E_{actual}, at some fixed value of d' (refs 7–9),

$$F = (d'_{actual}/d'_{ideal})^2 \qquad \text{for constant } E$$
$$F = E_{ideal}/E_{actual} \qquad \text{for constant } d' \tag{8}$$

We use the latter relation as in our experiments we varied contrast to yield a fixed value of d'. This expression also makes it clear that for any visual signal on a fixed background, efficiency is inversely proportional to threshold contrast energy. Hence that signal is seen best (seen with greatest efficiency) for which threshold contrast energy is least.

Quantum efficiency for a particular signal will be less than one if the detector must contend with noise in addition to quantum fluctuations. However, provided that the noise is effectively uniform over the spatial and spectral extent of a signal, the optimal detector will still be one whose weighting function is matched to the signal.

Our principal motive in asking what the eye sees best is that the preceding observation may be inverted: if only a single detector exists, then the form of the signal that is seen best identifies the weighting function of the detector[10]. If on the other hand there are several detectors with different weighting functions, then the best signal would identify the weighting function of the most efficient detector—that least impeded by noise. Therefore the search for the best signal is a search for the weighting function of the most efficient visual detector.

In our experiments we used a two-interval forced-choice procedure. One interval contained I_N, the other I_{S+N}. To perform optimally, the observer would choose that interval in which the detector gave the larger response. An alternative decision rule that has interested us is to select that interval in which the response exceeds some threshold, and if this occurs in neither interval, to guess. Threshold is set so that it is never exceeded in both intervals. Fortunately, it is also true for this detector that contrast energy is least when the signal is matched to the detector weighting function[11]. Thus here too, when only one detector exists, that signal will be seen best that is matched to the detector weighting function. When several detectors exist, possibly with different thresholds, the best stimulus will identify the weighting function of the detector with lowest threshold. As these thresholds are presumably set by the relevant noise, it is clear that these two formulations are quite similar.

In our search for the best stimulus we have confined ourselves to two classes of patterns, the first being square increments with two adjustable parameters: width and duration. The second class consists of patches of drifting sinusoidal gratings with gaussian envelopes in horizontal, vertical and temporal dimensions. A drifting patch with unit contrast may be written

$$c(x, y, t) = \sin[2\pi(f_x x - f_t t)] \exp[-(x/s_x)^2 - (y/s_y)^2 - (t/s_t)^2] \tag{9}$$

The patch has five parameters: the spatial frequency f_x, the drift frequency f_t, and the horizontal, vertical and temporal gaussian half-widths s_x, s_y and s_t. The width, height or duration of a patch is defined as twice the relevant half-width. An example of a grating patch is shown in Fig. 1.

This second class of pattern is more general than it may at first seem. With appropriate choice of parameters, it includes circular spots of various sizes and thin lines of various lengths. Patches were used in part for their generality, and in part because they include waveforms which resemble the weighting functions of simple cortical cells[5,6].

All signals were superimposed on a background of 340 cd m^{-2}, and were computer-generated on a large (20×30 cm) cathode ray tube with a ^{31}P phosphor. Viewing was binocular with best optical correction and natural pupils from a distance of 228 cm. Observers fixated the centre of the square increment or of the stationary spatial gaussian window. We

For square increments, we measured thresholds for squares ranging in width from 0.075 to 1.2 deg, and varying in duration from 10 to 400 ms. The best square we have found has a 0.3 deg side, and a duration of 50 ms. For observer H.B.B., it has a threshold contrast energy of −5.6 log deg^2 s, which corresponds to a threshold contrast of 2.36%. Squares as small as 0.075 deg on a side were detected almost as well.

To find the best grating patch, we must hunt through a five-dimensional space. We have necessarily limited ourselves to a number of slices through this space, and can show here only

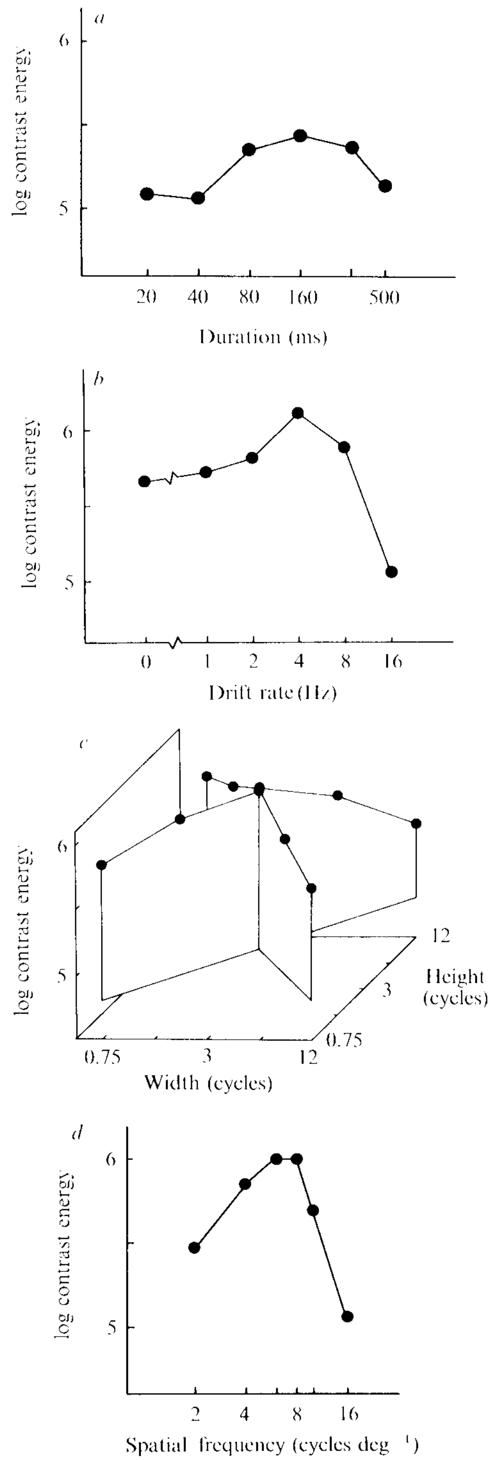

Fig. 2 Contrast energy thresholds for grating patches as a function of duration (a), drift rate (b), height and width (c), and spatial frequency (d). Height, width and duration of a patch are defined

a few of the many we actually examined. Figure 2*a* shows contrast energy thresholds for a 10 *c* deg^{-1} patch, 0.3 deg in height and width, drifting at 4 Hz. We have also measured thresholds for flickering, rather than drifting patches. For the stimuli in Fig. 2*a*, drifting and flickering patches have almost the same threshold contrast energy. A duration of ~160 ms is best, although the maximum is not very sharp. Figure 2*b* shows the dependence on temporal frequency of threshold contrast energy for a 6 *c* deg^{-1} patch, 0.5 deg in height and width, with an optimal duration of 160 ms. The best frequency is ~4 Hz. Figure 2*c* illustrates the effect of height and width on the threshold contrast energy of a 6 *c* deg^{-1} patch with optimal drift rate and duration of 4 Hz and 160 ms. Minimum threshold contrast energy occurs at a width and height of ~3 cycles. From comparable results at other spatial frequencies, we find the optimum to be more nearly a fixed number of cycles than a fixed size in degrees. Thus, in Fig. 2*d* we show contrast energy thresholds for patches of various spatial frequencies in which height and width are fixed at the optimal value of 3 cycles. Duration and drift rate are again set to optimal values of 160 ms and 4 Hz. The optimum spatial frequency is found to lie between 6 and 8 *c* deg^{-1}. Figure 1, when viewed from such a distance that one cycle of the sinusoid subtends about 1/7th of a degree, is a static picture of what the eye sees best.

This is the most efficiently detected stimulus we have discovered. If visual thresholds are set by linear detectors, this approximates the weighting function of the most efficient among them. It is likely that the threshold for a given stimulus is not set by a single detector at a single point in time, but rather by probability summation (or some other variety of nonlinear combination) over various detectors and over time[12,13]. A more precise estimate of the detector weighting function would take this into account.

For observer H.B.B., the best stimulus has a threshold contrast energy of -6.03 log deg^2 s, corresponding to a threshold contrast of 1.44% at 6 *c* deg^{-1}. This is about one-third the threshold contrast energy of the best square. We have not investigated all possible patterns, so a still more visible stimulus may exist at the luminance we have used. We invite those with candidate stimuli to test them against our best.

Note that at the moderately high luminance we have used (340 cd m^{-2}), the actual quantum efficiency is only ~0.0005 ($d' = 1.273$ and for both pupils together, $I_N = 4.2 \times 10^9$ quanta deg^{-2} s^{-1}). Much higher levels of quantum efficiency have been obtained at lower photopic luminances, and at low scotopic luminances the value approaches the fraction of quanta effectively absorbed[14]. As the foveal cones must surely absorb a much higher fraction of the available quanta, there are clearly limits to sensitivity beyond the quantum fluctuations. Recent measurements of thresholds for grating patches embedded in computer-generated noise[15] suggest near-perfect statistical efficiencies for signals like our best. Although noise added to the stimulus does not necessarily mimic internal noise, this observation reinforces our suspicion that the visual brain contains detectors with weighting functions of this form, and strengthens our belief that intrinsic noise limits detection on high luminance backgrounds.

The particular form of our best stimulus is of interest for several reasons. First, the detector spatial weighting function deduced here resembles the receptive field sensitivity profiles of many cortical neurones. The spatial frequency bandwidth of our best detector is about one half-octave, although stimuli with one-octave bandwidth are seen almost as well. Movshon *et al.*[5] and De Valois *et al.*[6] found that about 15% of simple cells in area 17 of cat and monkey have bandwidths of one octave or less. Second, we note that the detectors discovered here resemble those derived from other recent experiments on both detection[16,17] and discrimination[18,19] and may account for a broad range of previous psychophysical results.

Note also that quantum efficiency changes rather slowly with change of some stimulus parameters near the optimum. This not only makes the search for the best stimulus more difficult, but also suggests that the brain may contain many such detectors, with weighting functions located at different points and suited to stimuli of different parameters. In this regard we observe that the waveform of the best stimulus is one of a set of elementary basis functions devised by Gabor[20] and subsequently shown by Helstrom[21] to provide a complete and exact representation of an arbitrary waveform. Therefore a set of such detectors, when distributed appropriately over space and spatial frequency, are able to encode an arbitrary visual image, as suggested in several recent models[17,22,23]. Thus patterns like that in Fig. 1 may be among the elementary features of visual perception.

Received 19 November 1982; accepted 17 January 1983.

1. Hartline, H. K. *Am. J. Physiol.* **130,** 700 (1940).
2. Barlow, H. B. *J. Physiol., Lond.* **119,** 69 (1953).
3. Enroth-Cugell, C. & Robson, J. G. *J. Physiol., Lond.* **187,** 517–552 (1966).
4. Lennie, P. *Vision Res.* **20,** 561–594 (1980).
5. Movshon, J. A., Thompson, I. D. & Tolhurst, D. J. *J. Physiol., Lond.* **283,** 53–77 (1978).
6. De Valois, R. L., Albrecht, D. G. & Thorell, L. G. in *Frontiers of Visual Science* (eds Cool, S. J. & Smith, E. L.) 544–556 (Springer, New York, 1978).
7. Fisher, R. A. *Statistical Methods for Research Workers* (Oliver & Boyd, Edinburgh, 1925).
8. Rose, A. *Proc. Instn Radio Engrs* **30,** 293 (1942); *Vision: Human and Electronic* (Plenum, New York, 1973).
9. Tanner, W. P. Jr & Birdsall, T. G. *J. acoust. Soc. Am.* **30,** 922–928 (1958).
10. Barlow, H. B. *Vision Res.* **18,** 637–650 (1978).
11. Levinson, J. Z. *J. opt. Soc. Am.* **58,** 1558 (1968).
12. Watson, A. B. *Vision Res.* **19,** 515–522 (1979).
13. Robson, J. G. & Graham, N. *Vision Res.* **21,** 409–418 (1981).
14. Barlow, H. B. in *Vertebrate Photoreception* (eds Barlow, H. B. & Fatt, P.) 337–358 (Academic, New York, 1977).
15. Burgess, A. E., Wagner, R. F., Jennings, R. J. & Barlow, H. B. *Science* **214,** 93 (1981).
16. Robson, J. G. & Graham, N. *Vision Res.* **21,** 409–418 (1981).
17. Watson, A. B. *Vision Res.* **22,** 17–25 (1982).
18. Thomas, J. P. & Gille, J. *J. opt. Soc. Am.* **69,** 652–660 (1979).
19. Watson, A. B. & Robson, J. G. *Vision Res.* **21,** 1115–1122 (1981).
20. Gabor, D. *J.IEE London* **93**(III), 429–457 (1946).
21. Helstrom, C. W. *IEEE Trans. Inf. Theory* **IT-13**, 81–82 (1966).
22. Marcelja, S. *J. Opt. Soc. Am.* **70,** 1297–1300 (1980).
23. Sakitt, B. & Barlow, H. B. *Biol. Cybern.* **43,** 97–108 (1982).