Vision Res. Vol. 25, No. 11, pp. 1661–1674, 1985
Printed in Great Britain. 

0042-6989/85 $3.00 + 0.00


# A THEORY OF THE PRIMITIVE SPATIAL CODE IN HUMAN VISION

R. J. WATT[1] and M. J. MORGAN[2]

[1]M.R.C. Applied Psychology Unit, 15 Chaucer Road, Cambridge CB2 2EF and [2]Department of Psychology, University College, Gower Street, London WC1E 6BT, England

(*Received* 6 *July* 1984; *in revised form* 28 *March* 1985)

**Abstract**—MIRAGE, a theory for the primitive coding of the (1D) spatial distribution of luminance changes by the human visual system is developed from a theoretical examination of the practical problems associated with the characterization of such changes. The main novel feature of the theory is that the multiplicity of spatial filters in human vision is assumed to exist principally to transmit a broad bandwidth signal of considerable redundancy: the filters are not assumed to be marked with their centre frequency or bandwidth, and are not analyzed independently. The theory is largely independent of the particular filter transfer function form. MIRAGE is applied to a range of one-dimensional luminance patterns, and demonstrates several well-known brightness illusions, and a structured grouping principle. It is finally shown to be supported by a wide range of psychophysical data.

Edge/bar recognition  Central moments  Spatial filters  Noise  Spatial frequency interactions

## INTRODUCTION

The principal goal of the early stages of spatial vision is to characterize luminance changes in the retinal stimulus so that subsequent processes can derive information about intrinsic scene properties such as its surface discontinuities, slopes, reflectances and illumination (Marr, 1976; Barrow and Tenenbaum, 1978; Binford, 1981; Rosenfeld, 1983). For each change in luminance, measurements must be made of at least the location of the change, its spatial extent, and the amplitude of the luminance excursion. It seems most likely that the human visual system accomplishes this with a range of linear spatial-frequency selective elements or spatial filters that transduce the spatial information of the retinal image in an independent fashion (Campbell and Robson, 1968; Wilson and Bergen, 1979; for a review see Braddick *et al.*, 1978), and whose spatially extended outputs are then combined to allow inferences about that retinal image to be drawn.

This paper will describe a model (MIRAGE) of how this is accomplished by the human visual system. The model satisfies three criteria: it is computationally efficient; it agrees with a wide range of psychophysical data; and it successfully predicts the appearance of many different stimuli. The end result of the MIRAGE transformation is a symbolic representation (with continuously valued attributes). The symbols are themselves in a structured list, making some more rapidly accessible than others.

The model is derived from a theoretical examination of the computational difficulties associated with the characterization of *unknown* luminance changes in a noisy signal, particularly if the rate of information is to be limited. These restrictions are shown to have implications for the type of primitive measurement chosen to represent the stimulus and provide the basis for all subsequent processing. It is deduced that the central moments of zero-bounded distributions of activity in the filtered signal are the most suitable features for two reasons: (1) in the presence of response noise it is a simple and reasonable procedure to remove most spurious features; (2) the representational capacity (in terms of the ratio of data transmission to coding complexity) is particularly favourable. Both these arguments can apply when more than one filter is being used if, as we suggest, all the various filter responses are collapsed down into two spatial signals.

The model has as free parameters the range of filter space constants and their response gains. Spatially uncorrelated noise with a constant variance and zero-mean is assumed to be added to the linear convolution responses of each filter before any further processing. It is shown that psychophysical measurements of (i) the contrast sensitivity function, (ii) edge location precision for many different waveforms, and (iii) edge blur difference discrimination are all consistent with the same single set of filters.

## THEORETICAL EXPOSITION OF THE PROBLEM

Figure 1 shows a symmetric luminance change that is restricted and isolated in space and is produced by the convolution of a step function with a unit area Gaussian blurring function. This will serve as a sample stimulus for our discussion of the most suitable technique for edge characterization. *Edge location* we define as the discontinuous point in the step function; *edge blur* we define as the standard deviation of the blurring function; *edge contrast* we define

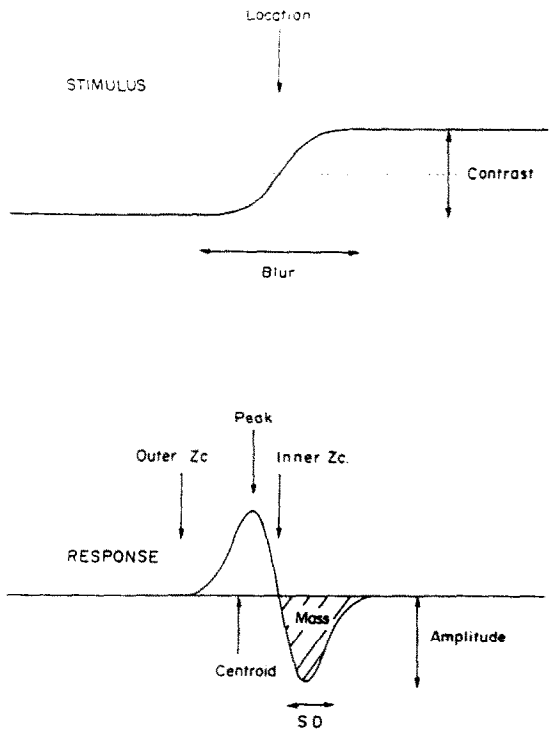


Fig. 1. At the top is shown the (1D) luminance profile of an edge and the terms location, blur and contrast are illustrated. Mean luminance is shown by a dotted line. Note that the luminance difference across the edge and contrast are linearly related at any one given mean luminance. In the text it is explained that contrast is probably the more correct measure. Beneath is shown the response of a triphasic spatial filter to the stimulus, together with the associated terms.

as the luminance difference across the edge divided by twice the mean luminance.

We assume the use of a range of bandpass spatial filters, with impulse functions that are at least triphasic. For convenience we shall adopt the normalized second derivative of a Gaussian

$$f(x) = (x^2/\sigma_f^2 - 1)\exp(-x^2/2\sigma_f^2)$$

with a range of different values for the space constant, $\sigma_f$. The theory that we shall now develop would apply directly for any other symmetric triphasic filter impulse function, and could be recast along the same lines for any other multi-phasic filter impulse function. Convolution of the filter function with the image luminance profile produces a spatially extended response, which has different properties from the image itself. The most notable change is in the mean signal level which is now near to zero, and the response will have both negative and positive values. Strictly the amplitude of the response of such filters is determined by the amplitude of the luminance differences in the image, but it seems likely that an initial non-linear energy transduction would cause edge contrast to determine response amplitude (Geisler, 1983). The spatially extended output image from each filter is subject to spatially uncorrelated noise, which may well be a function of mean luminance (Tolhurst *et al.*, 1981).

*(i) Edge characterization using one spatial filter*

Let us suppose that the edge is to be represented by a set of primitives, which are then used for all subsequent processing. The edge gives rise to a spatial response, $R(x)$, in a filter which contains many useful features (see Fig. 1), any of which could serve as primitives. The sequence of zero-crossings, $Zc_i$, in a filter response are a rich source of information (Marr and Poggio, 1979; Marr *et al.*, 1979; Marr and Hildreth, 1980), as are its peaks and troughs, where the rate of change of response is zero (Mayhew and Frisby, 1981; Watt and Morgan, 1983), and any one of these features could be used to assess edge location. The distance between peak and trough could be used to assess blur, and then edge contrast could be estimated from the response amplitude taking the blur into account (because blur would itself also affect the response amplitude). If the zero-crossing were to be used then both its slope and the response amplitude would also be needed in order to assess the blur and contrast of the edge.

Between adjacent zero-crossings, $Zc_i$ and $Zc_{i+1}$, in the response (inner and outer zero-crossings in the figure) is a distribution of activity which can be characterized by its central moments. Suppose that the centroid, $P_i$, was measured (that is the point in each zero-bounded distribution about which the first order moment is zero)

$$P_i = \frac{\int_{Zc_i}^{Zc_{i+1}} x \cdot R(x) \cdot \mathrm{d}x}{\int_{Zc_i}^{Zc_{i+1}} R(x) \cdot \mathrm{d}x}$$

The response to the edge has two zero-bounded distributions and the blur of the edge is a function of the separation of their centroids and the filter space constant. The rate of luminance change at the edge is proportional to the mass, $M_i$, which is the zero order moment

$$M_i = \int_{Zc_i}^{Zc_{i+1}} R(x) \cdot \mathrm{d}x$$

Three measurements are always needed to determine the three edge properties. However the various possible measurements are not equally affected by adding noise to the filter output. Noise introduces spurious zero-crossings and stationary points, and makes these features less reliable than the central moments, the measurement of which implies averaging and is therefore relatively noise resistant. Figure 2 plots the standard deviation of estimates of the values for the different features, as a function of noise amplitude relative to a fixed signal amplitude. Noise amplitude is the range within which the local instantaneous noise value may randomly occur. Since noise introduces spurious features, a rule in each case was devised to decide which instance of a given feature to

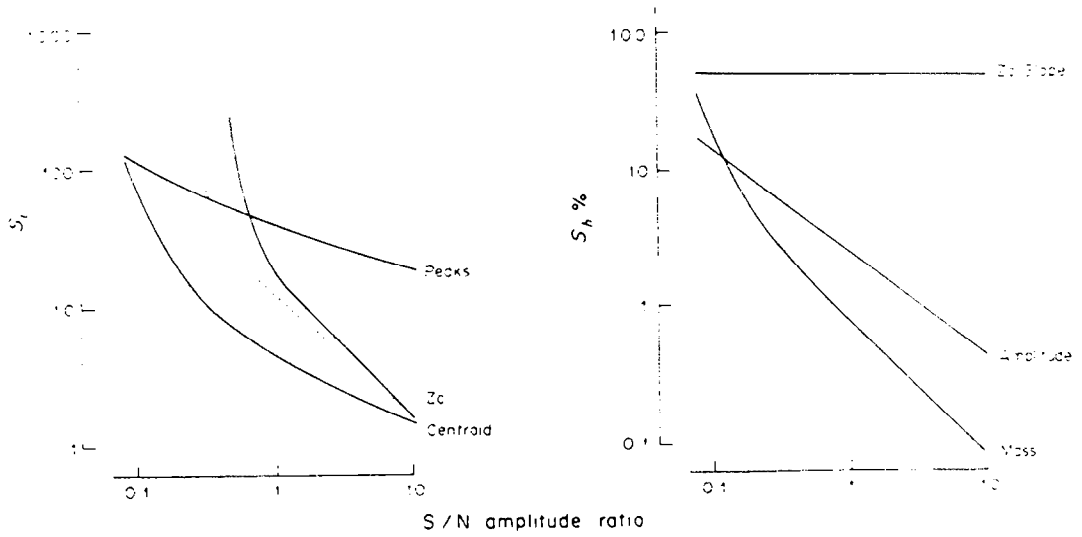


Fig. 2. The standard deviations ($S_l$ and $S_h$) of the distributions of estimates of the values of the various parameters in a filter response (in the presence of spatially uncorrelated noise) are plotted as a function of signal noise amplitude ratio. This is a measure of the precision with which the parameters can be determined, so that the lowest SD corresponds to the highest accuracy. On the left are shown spatial quantities, and on the right intensive quantities. For the spatial quantities the standard deviations are given in absolute distance units, for the intensive quantities a percentage scale is used. The values were determined by numerical modelling with a linear congruential random number sequence generator. Note that the moment-based properties, mass and centroid, are the most reliable.

use. In the case of peaks, that with the largest amplitude was taken. In the case of centroids, that with the largest mass was taken. In the case of zero-crossings, that with the largest slope was taken and linear interpolation between samples was employed. The zero-crossing chosen by this rule proved to be most unreliable at moderate and high noise levels, and so the figure also shows, as a dotted line, the effect of taking the steepest zero-crossing bounding the largest mass (a similar result occurs for taking the zero-crossing nearest to the largest amplitude peak).

The effects of noise shown in the figure arise because no knowledge of the stimulus waveform or position was employed (except in the case of the dotted line for zero-crossings). For a zero noise level, the standard deviation is zero for all features. For a signal/noise amplitude ratio greater than about 2 the noise did not usually introduce spurious extra zero-crossings about the true zero-crossing, and its only effect was to shift randomly the position of the zero-crossing. For this range a linear relationship between zero-crossing error and noise amplitude is found. For the same range, a square root relationship is found for peaks and centroids. At high noise levels the zero-crossing centroid and mass measures show an accelerating rise in variability; the peak and amplitude do not. The reason for this is that when the signal has less than half the peak amplitude of the noise, zero-crossings are likely at any place along the waveform, not just near the signal's zero-crossing, and the choice of which zero-crossing or mass is not always correct.

In conclusion, the centroid and mass measures are the most pragmatic choice in a circumstance where noise exists at the output of the filters.

*(ii) Using more than one filter*

For any given single bandpass filter of unit amplitude impulse function, there is a trade-off between positional accuracy, which is best for a large filter space constant (see Watt and Morgan, 1984, Fig. 2), and resolution, which is best for a small filter space constant. To have the best of both worlds, it might therefore be desirable to use several independent filters of different sizes, but this could cause a costly increase in the amount of information represented in the system. One way to limit the amount of information is to treat the various filter responses, $R_j$, as independent samples of the signal and to reduce noise by averaging them. Although acceptable for sparse textures (in which luminance changes occur not much more frequently than the centre frequency of the lowest filter), such averaging could reduce the resolution of the system considerably for denser textures. As is illustrated in Fig. 3, zero-crossings from the smallest filters would no longer necessarily cross zero in the average signal, resulting in the loss of certain response distributions and stimulus details altogether. Figure 3 also shows that this can be avoided if the individual filter responses are first split into positive and negative portions

$$R_j^+(x) = R_j(x);\ R_j(x) > 0$$

$$R_j^+(x) = 0;\ \text{otherwise}$$

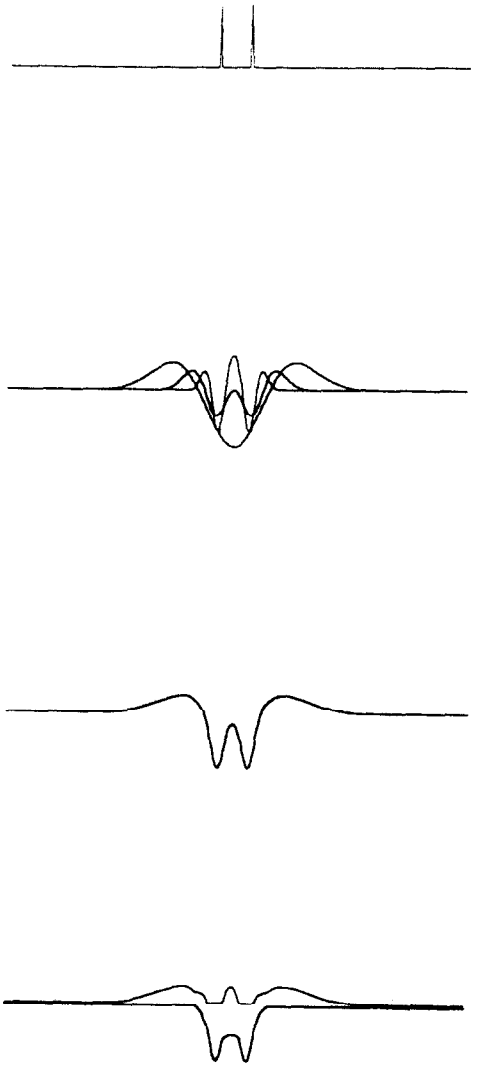

Fig. 3. The mechanism for visual resolution is illustrated. At the top is a stimulus comprising two bars separated by the minimum angle of resolution. In the second row the responses of a range of filters are shown superimposed. Note that only the smallest produces a central positive region of response, and thereby resolves the double. In the third row the sum of the filter responses is drawn. Although there is some effect of the central positive region in the response of the smallest filter in the averaged response, it does not cross zero, and would be difficult to detect in the presence of noise. In the bottom row, the sum of the positive portions of the filter responses is shown, as is the sum of the negative portions. The resolution of the smallest filter is preserved, and this is the basic mechanism of MIRAGE.

and

$$R_j^-(x) = R_j(x);\ R_j(x) < 0$$

$$R_j^-(x) = 0;\ \text{otherwise}$$

and then obtaining an average positive response, $T^+$, and an average negative response, $T^-$

$$T^+(x) = \sum_j g_j R_j^+(x)$$

and

$$T^-(x) = \sum_j g_j R_j^-(x)$$

which would then have between them at least one zero-bounded distribution for every monotonic change in luminance in the stimulus. The central *positive* response in the smallest filter is not averaged with the *negative* responses in all other filters, but is left apart and therefore survives. The analysis of each zero-bounded distribution into central moments (mass and centroid) would therefore adequately characterize such stimulus luminance changes. Each filter has its own independent gain, $g_j$, which would normally be set to unity but could be adaptive and responsible for the spatial frequency selective effects of adaptation upon contrast sensitivity (Blakemore and Campbell, 1969).

*(iii) Noise suppression*

The noise in each filter output is half-wave rectified along with the signals and this means that the noise processes in $R^+$ and $R^-$ each have non-zero means (e.g. $\pm\frac{1}{4}$ of the maximum noise amplitude, respectively, for a uniform probability of noise value). When the half-wave rectified signals are averaged, the noise variance is of course reduced but not its mean, and $T^+$ and $T^-$ will each have a nett d.c. shift corresponding to the mean noise level, as can be seen in the third row of Fig. 4. Because of this the signals $T$ will rarely be at zero and there will be relatively few zero-bounded response distributions, which will be rather uninformative. The remedy to this is to impose an opposite d.c. shift of amplitude, $t$, equal to the mean noise level: instead of zero-bounded distributions of activity in $T$, zero-bounded distributions of activity in $T - t$ should be detected and analyzed. The fourth row of Fig. 4 shows that this is quite effective. The noise also creates spurious distributions of activity where the noise locally exceeds $t$. It would be possible to set $t$ high enough to eradicate these, but this strategy would seriously reduce sensitivity to the signal, and a more efficient approach is to leave $t$ at the noise mean and ignore all distributions with a mass less than some threshold mass $m$.

*(iv) Regions of inactivity*

Watt and Morgan (1983) have argued that for more complex luminance waveforms, regions of inactivity have to be detected and represented, as well as distributions of activity in the filter response. All parts of the signals $T - t$ that were not represented as zero-bounded distributions would be regarded as zero-valued stationary points. The importance of the mass threshold, $m$, should now be clear. The noise is of some practical value in this respect, because it tends to truncate the ends of response distributions, which otherwise would only slowly approach zero.

*(v) Sampling rate*

The above discussion has been stated in terms of continuous functions, but in practice these would be represented by a series of discrete spatial samples forming implicitly continuous functions. Since the

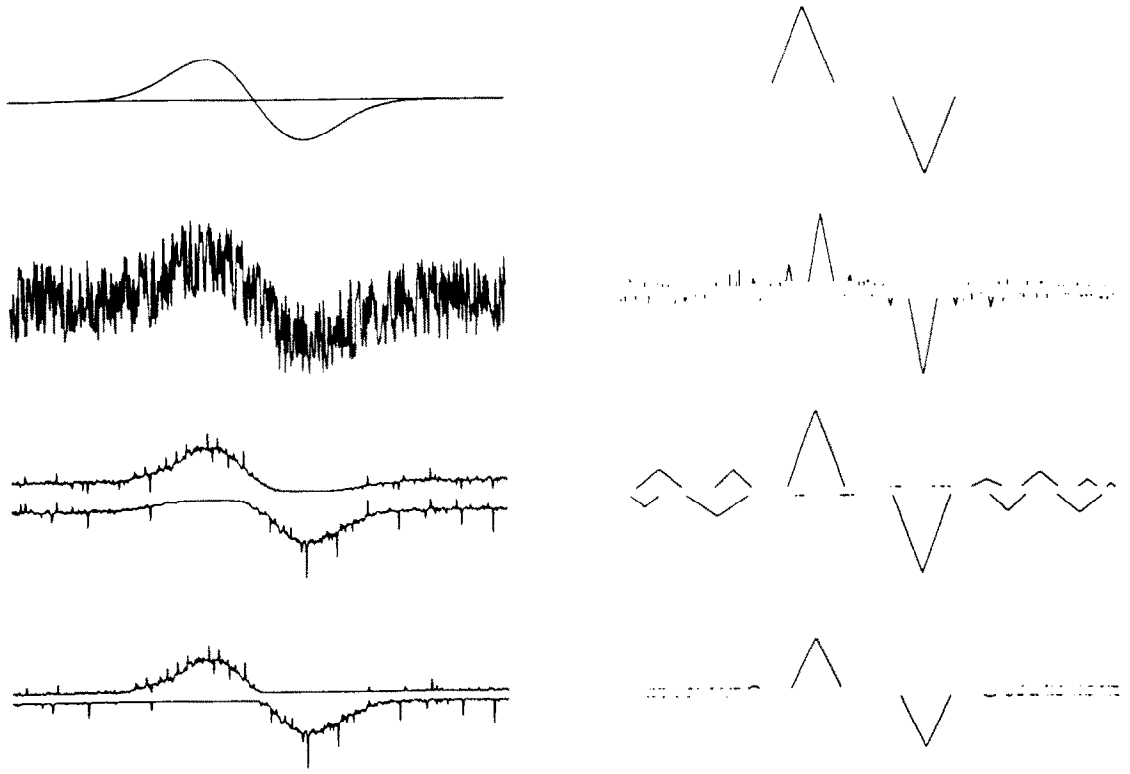

Fig. 4. The effects of noise on the estimates of centroid location and mass measurement are illustrated graphically. At the top left is shown the filter response (without noise). Opposite this, on the right, is shown the primitive characterization. The chevrons are placed with their apex at the location of the centroid, and are drawn with their height proportional to the mass, and their base width proportional to the SD of the response distribution they represent. The second row shows a case where the noise level in a single, unrectified filter output is high. The third row shows the result of averaging many independent filter outputs, after half-wave rectification. There is some improvement in the characterization over that from the single filter, but the spurious distributions still have an undesirably large standard deviation. The fourth row shows the effect of subtracting the mean noise level from the signals before measuring moments. The result is obviously a much more suitable representation of the signal than the previous two. Note that in the third and fourth rows the +ve and −ve signals are drawn slightly displaced from each other to help the reader distinguish them.

measurement of moments would not be improved by an explicit interpolation as suggested by Barlow (1979) and Crick *et al.* (1981), there would be no need for the various filter responses to be represented by samples at a rate much above their respective (and different) Nyquist limits. Furthermore, none of the stages we have described rely on the filter responses being represented at the same sampling rate, provided that the addition of signals after rectification were accompanied by a gain on each signal that was inversely proportional to the sampling rate for that signal to keep the signals in balance.

*(vi) The interpretation of the primitives*

Mirage transforms a continuous retinal image, by way of implicitly continuous intermediate transformations, into a spatially ordered list of discrete and symbolic primitives. The final aspect of the theory is the interpretation of the primitives, to allow an inference about the variation of luminance or brightness in the visual scene. The list contains two types of primitives, each of which has a number of associated continuously-variable attributes:

| | |
|---|---|
| (1) [RESP ($M$, $P$, $S$)] | A zero-bounded response distribution (with a measured mass, $M$, centroid, $P$, and SD, $S$). |
| (2) [ZVSP ($P_1$, $P_2$)] | A region of inactivity (between points $P_1$ and $P_2$). |

In order that local visual scene luminance or brightness may be inferred from the output of the MIRAGE transformation, three rules are needed to interpret the sequence of primitives:

($A$) *Null rule:* A [ZVSP] corresponds to a luminance plateau.

($B$) *Edge rule:* A [RESP] with a [ZVSP] on *only* one side, is interpreted as the boundary of an edge (monotonic luminance change). Luminance falls or rises away from the [ZVSP], depending on the sign of the measured mass.

($C$) *Bar rule:* A [RESP] with a [ZVSP] on *both* or *neither* side is interpreted as a dark or bright

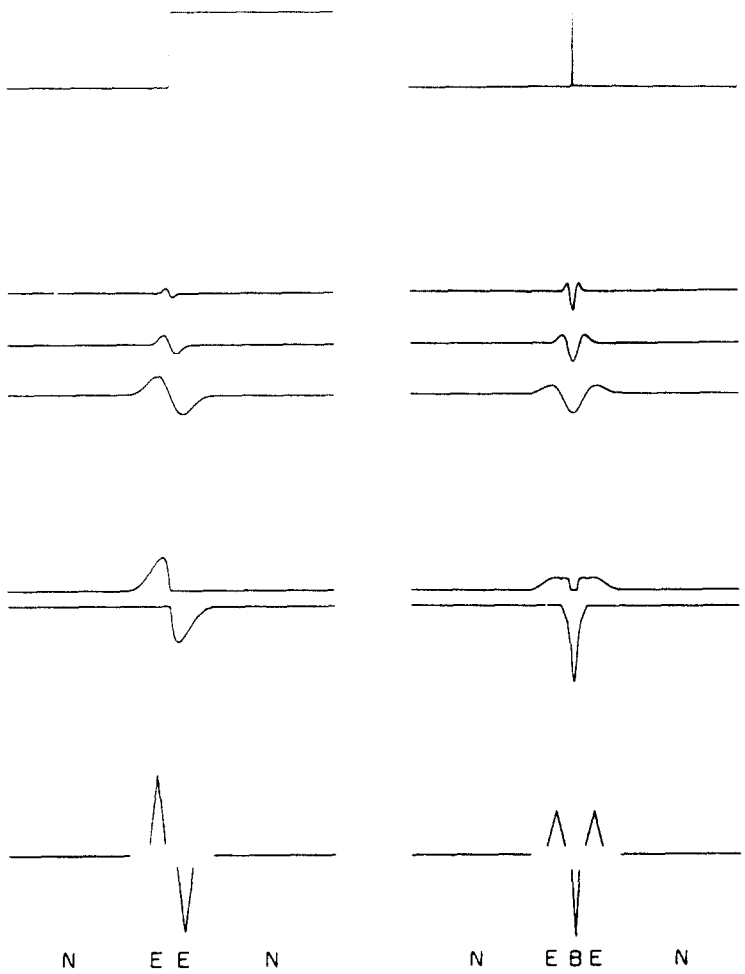


Fig. 5. Rules for interpreting the sequence of primitive measurements are illustrated. In the top row are shown the two basic stimuli: a step edge and a point. Beneath this are shown the responses of a range of filters to the stimuli. In the third row, the signals $T$ are plotted, and beneath these the list of primitives is shown. A [RESP] is shown by a chevron, of an amplitude proportional to its mass and a base width proportional to its SD, and located with its apex at the locus of the centroid. A [ZVSP] is shown by a horizontal bar. These are marked by the letters "N", "E", "B" to state which interpretative rule is used.

bar (local luminance maximum or minimum), depending on the sign of the measured mass.

We shall now use these three rules to interpret the response of MIRAGE to some complex luminance waveforms. In each case a figure is given showing the luminance waveform, $I$, itself at the top. Underneath this, the response of three filters, $R_j$, at octave intervals to the luminance waveform will be shown. The results of splitting the responses according to sign, and summation across the different filters to produce the signals, $T^+$ and $T^-$, are shown in the third row. Finally, in the bottom row will be shown a graphical representation of the primitive list, with an indication of which rule is used to interpret each element.

(*a*) *Basic waveforms.* Figure 5 shows the basic edge and bar rules in action for an edge and a point stimulus. We assume that the appearance of these is veridical, by asserting the three inferential rules.

(*b*) *Brightness errors.* The importance of the ZVSP is shown in Fig. 6, which shows what happens when its presence or absence is mis-interpreted. A luminance ramp of more than a critical width leads to a central ZVSP, as can be seen in the case on the right of the figure. This leads to the use of the bar rule to interpret the RESP primitives and spurious Mach bands result. On the left of the figure the response of MIRAGE to a pair of like polarity edges, separated by less than a critical width. The system fails to resolve a ZVSP between the two edges, and the bar rule is again used inappropriately. As a result bright and dark bars are inferred at the ends of the middle luminance region (this is called the Chevreul illusion). The critical width in each case is approximately the same at about 5 arc min (Ross *et al.*, 1981), a distance determined according to our theory by the space constant of the largest filter.

(*c*) *Grouping.* The stimulus on the left of Fig. 7 comprises two point sources separated by just 1 arc min, and that on the right by 5 arc min. In the latter case a straightforward representation is ac-

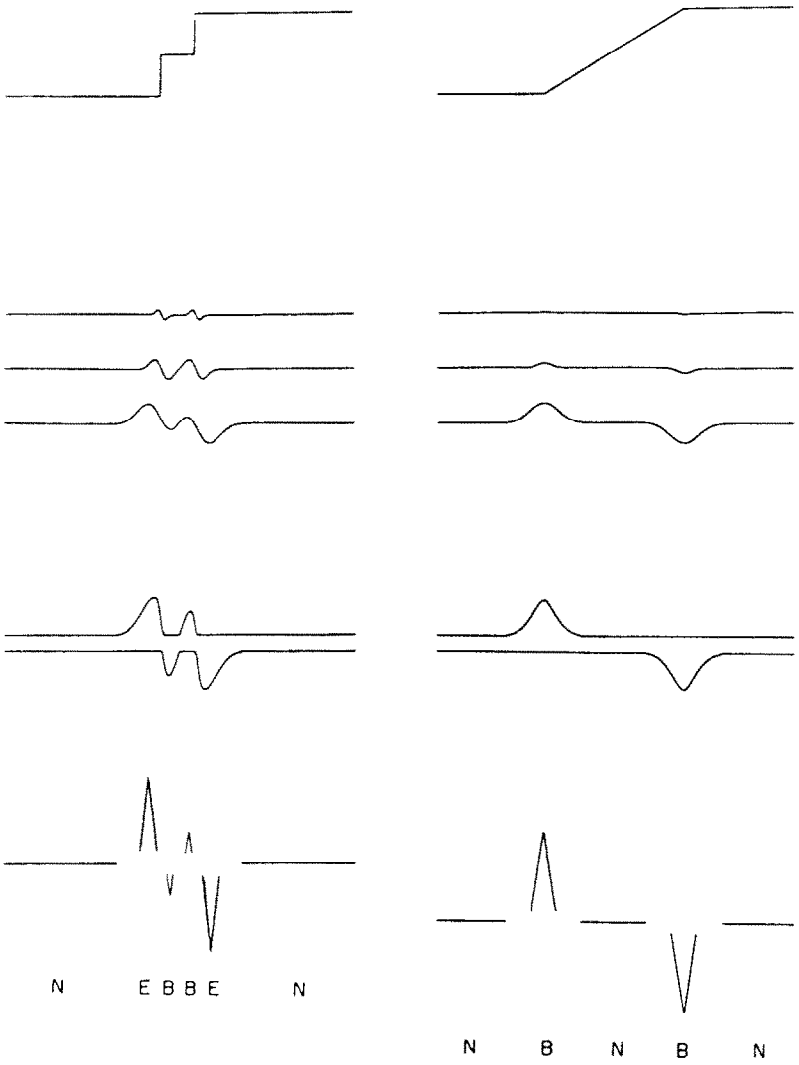


Fig. 6. Brightness errors in the performance of MIRAGE. On the left there are two like-polarity edges. Since the plateau region of luminance between them is not resolved into a ZVSP, a brightness illusion occurs following the mis-use of the bar rule for interpretation. On the right the response of MIRAGE to a luminance ramp is shown. The ZVSP in the response to the ramp leads to a misinterpretation by the bar rule and Mach bands result.

corded, but in the former case there are two primitives produced at the same place (they necessarily have opposite signs). Applying the bar rule to each, we have a broad bright bar with a narrower dark bar superimposed. There is an important principle suggested by this. In a sense MIRAGE can be said to have structured the representation of this stimulus, in that the stimulus to be described is two bright bars side by side, but the description produced has a different structure: finer detail is subordinated to a coarser grouping. The same idea can be observed in the following cases also.

(*d*) *Complex luminance modulation.* Figures 8 and 9 show some instances of complex luminance waveforms, and the actions of MIRAGE. From the first case, it can be seen that MIRAGE can be said to represent relative spatial phase, and from the second case, to accomplish spatial interpolation. Once again, where distributions in $T^+$ and $T^-$ overlap a structured representation results, with fine detail superimposed on coarser variations. In the case of the two cosines, one or two small bars are grouped by the large bars in the peaks add and peaks subtract phases respectively. It therefore follows that these could be discriminated one from the other by the grouping arrangement, whereas other phase differences which would not lead to a difference in the number of small bars grouped could only be discriminated by an examination of the individual bar locations. This has been forwarded by Watt (1985) as an explanation for the phase discrimination data of Rentschler and Treutwein (1985).

### (*vii*) *MIRAGE in two dimensions*

We have described MIRAGE as a one-dimensional transformation, and left aside any considerations of

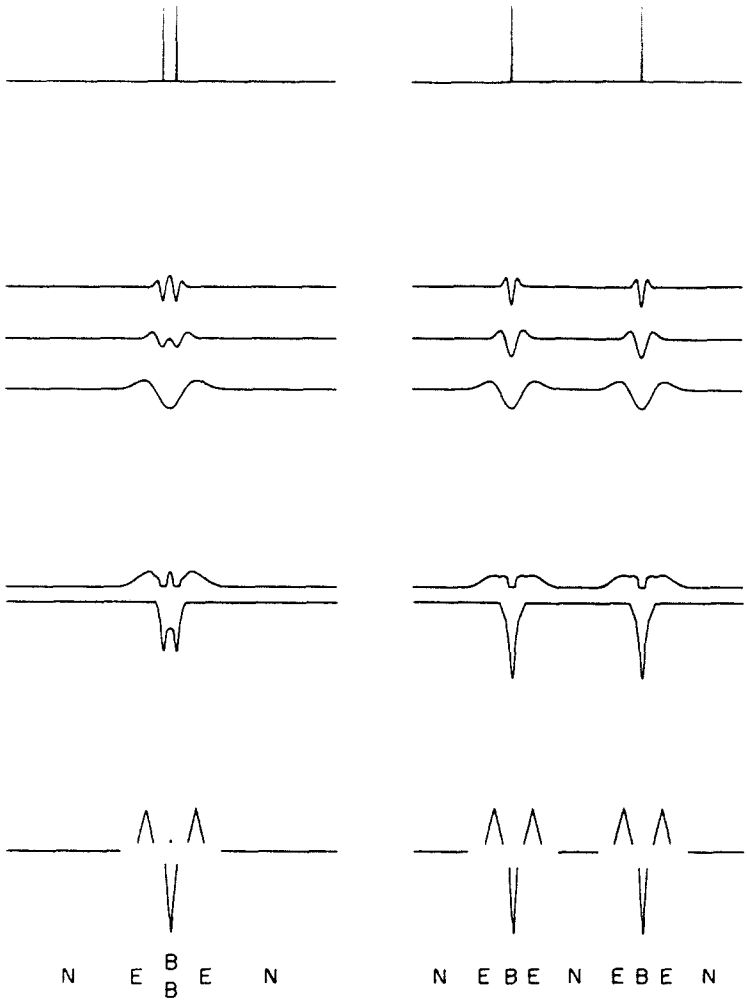


Fig. 7. The grouping behaviour of MIRAGE. The response of MIRAGE to a pair of point sources, close to resolution (on the left), and widely separated (on the right). Note that two primitives are found at the same locus in the former case. Thus the representation of the two side-by-side bars has a different structure: the stimulus is represented as a bright bar with a narrower dark bar superimposed on it.

how it may be used for two-dimensional images. The reason for this is the paucity of data. In theory MIRAGE could be extended into two dimensions in several different ways, for example by using elongated (orientation selective) filters and thereby reducing the two-dimensions to a set of one-dimensions. The finding of Daugman (1984) that the filters are not polar separable makes this unlikely. Alternatively, circularly symmetric filters could be used, with the centroid extraction process being orientation selective. In either case, it seems likely that further rules to interpret points, corners and terminations might be necessary.

The general principle that detail in the smallest filter is preserved but interactions due to the largest filter also occur has been used by Watt (1984) to explain several aspects of shape discrimination.

*(viii) MIRAGE: a summary*

We shall use the acronym MIRAGE to describe the transformation from the retinal image to the list of characterized primitives. It may be summarized thus: Multiple Independent filters, of various sizes and with both signs, half-wave Rectified before Averaging. The resultant signals are Gated between adjacent zeroes for the Extraction of the primitive code.

## DEVELOPMENT: SOME EXPERIMENTAL DATA

In the preceding section we have shown a means by which luminance changes can be characterized by the use of multiphasic spatial filters, and we have argued that the scheme is effective and reliable in the presence of intrinsic moise. It is offered as a model for human spatial vision, and in this section it is shown that several critical elements of MIRAGE are supported by empirical evidence as general principles.

*Theme 1: unknown stimulus.*

The major principle in the exposition was that the stimulus is unknown. This led to the adoption of zero-bounded response distributions as the features

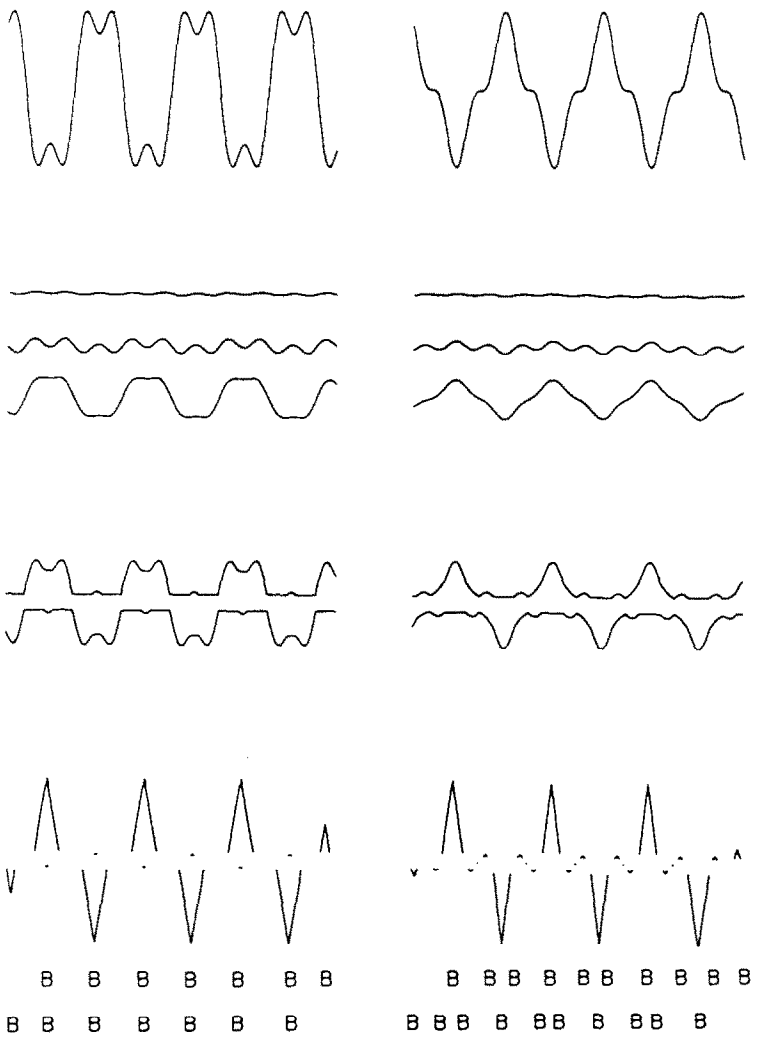

Fig. 8. The response of MIRAGE to two sections of compound grating. In each case there are two frequency components present, F1 and F3, on the left in peaks-add phase, and on the right in peaks-subtract phase.

to be characterized. By using zero-bounds the tails of response distributions are truncated by the spurious noise zero-crossings and small amount of the mass is removed. Although at high contrasts this has little effect, at low contrasts it has three important consequences.

The truncation removes mass, to an amount which is dependent on the signal amplitude. This results in a small nonlinearity in the function relating mass to contrast, which is shown in Fig. 10. This, in conjunction with a mass threshold, which is itself necessary because regions of inactivity are equally unknown, results in a psychometric function for contrast detection which is not a cumulative normal, but rather has the same form as the contrast transduction function of Fig. 10. This function is in good agreement with the data and proposals of Nachmias and Sansbury (1974) and Foley and Legge (1981). In fact, the extent and exact form of the nonlinearity depend on the stimulus waveform. Fig. 10 shows the nonlinearity for a compound grating ($F + 3F$) in either peaks-add or peaks-subtract phase, as well as the fundamental on its own. It can be seen that if a mass threshold is set below the 0.01 mark, then thresholds for the fundamental or either compound will all be nearly the same, as found by Graham and Nachmias (1971).

The second effect of spurious zero-crossings that arises from the unknown stimulus postulate is that there is a greater variability at low contrasts in the value obtained for the mass (see Fig. 2). This means that contrast discriminations will be less accurate at low contrasts. Contrast difference discrimination involves the measurement of two masses and their comparison, and therefore the errors of judgement made should have a standard deviation, $S_c$, determined by the standard deviation of the errors due to the measurement of mass, $E_m$, and that of the errors made in comparing massess, $W_m$

$$S_c^2 = E_m^2 + W_m^2.$$

We suppose that a Weber's law applies for the

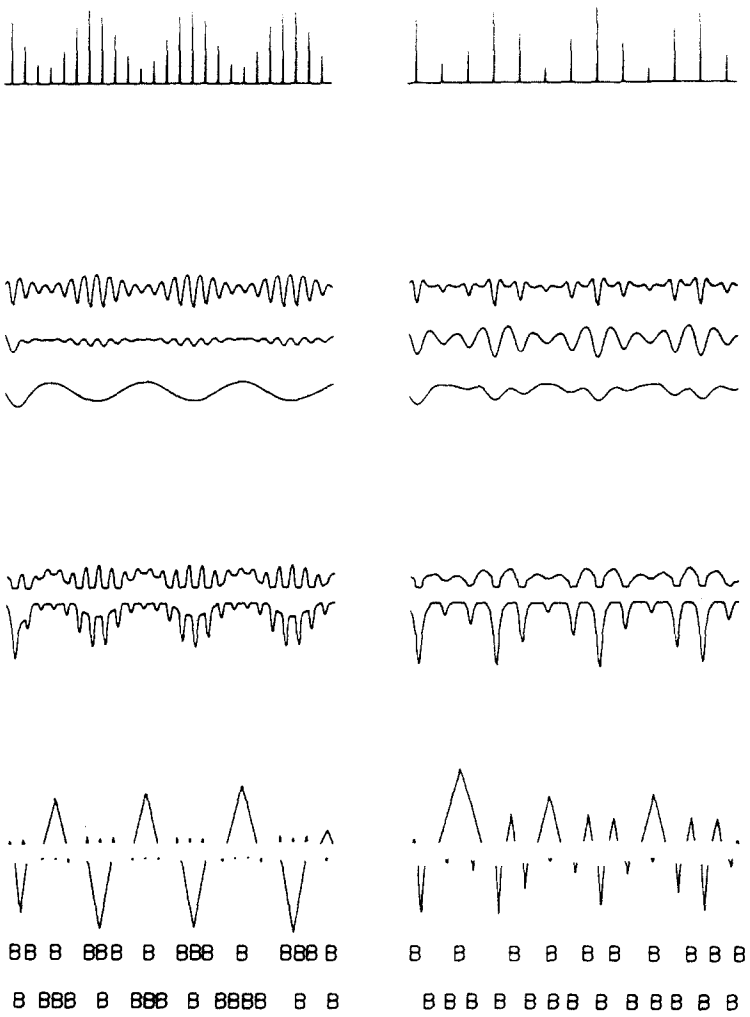

Fig. 9. This figure illustrates the response of MIRAGE to a discretely modulated waveform and shows how it is able to simultaneously interpolate between the individual samples and to resolve them. On the left the samples are 2 arc min apart, on the right they are 4 arc min apart. The interpolation is only successful in the former case. The samples are themselves grouped according to the phase (positive or negative) of the cosine.

judgement process of comparison (see Laming, 1982 for a detailed consideration of what this implies), and Fig. 11 shows the effect of a Weber's law for mass comparisons when expressed in terms of contrast. The figure also shows the uncertainty, $E_m$, with which the zero-bounded mass may be measured. The joint effect, $S_c$, of these two independent sources of error is a non-monotonic function as shown in Fig. 11 as a function of $C$. As can be seen, this compares well with the data of Legge (1982) and Legge and Kersten (1983). The function drawn has two free parameters, namely the Weber fraction for mass, and the noise level in the signals $T$. These influence the vertical position of the functions shown, and the point at which $E_m$ flattens out, but the shape of the function for $S_c$ would be little changed by relatively large variations in these parameters.

The third effect of the spurious zero-crossings is the rapidly increasing variability in location accuracy at low contrasts (see Fig. 2). Watt and Morgan (1984) found that as contrast is reduced, location precision first falls with the square root of contrast but at low contrasts it falls more rapidly. As can be seen from Fig. 2, zero-bounded centroids do indeed behave in this fashion, and furthermore, no other feature that we have considered does so.

*Theme 2: multiple filters*

We declare that the filters span a range of space constants (SDs) from 0.35 to 2.8 arc min, and all have the same amplitude impulse function and maximum noise amplitude. This range fits the edge location and blur discrimination data of Watt and Morgan (1983, 1984), and their two sets of results are compared with predictions in Fig. 12. The use of multiple filters was suggested to achieve both high location accuracy and high resolution, and Fig. 12 shows how well matched these two are. The prescribed range of filters also fits the contrast sensitivity function of Campbell and Robson (1968), as is shown in Fig. 13. These three

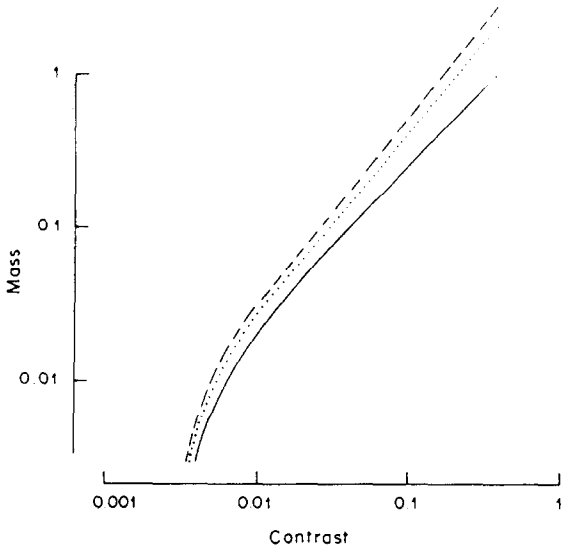


Fig. 10. The non-linear relationship between mass and contrast. We have assumed that the signals $R$, and $T$ do not saturate in amplitude. If they did, then the functions would climb progressively less steeply at higher contrasts, but would not reach a plateau. The solid curve shows the function for a cosine wave stimulus; the dotted line shows the function for the difference of two cosine waves (F − 3F i.e. in peaks-subtract phase); and the dashed line shows the function for the sum of two cosine waves (F + 3F i.e. in peaks-add phase). A mass threshold below 0.01 (in our units) would lead to the contrast thresholds for all three being similar.

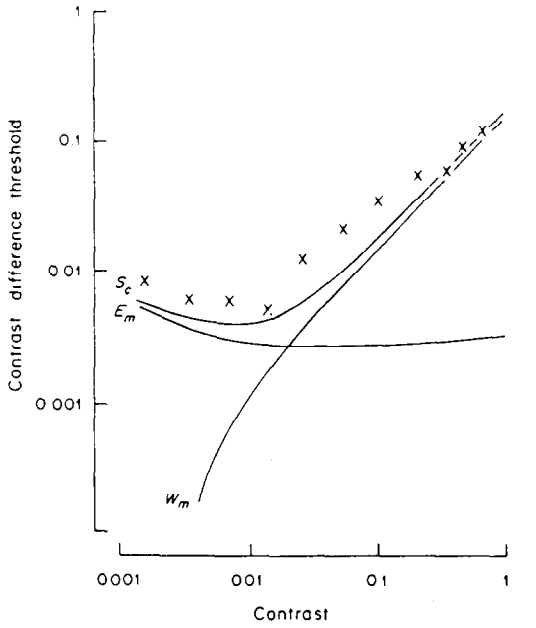


Fig. 11. The expected variation in contrast difference threshold, $S_c$, as a function of criterion contrast, according to MIRAGE. It also shows, $W_m$, the result of a Weber's law for mass expressed as an effect of contrast (according to the transformation of Fig. 10), and the precision with which the mass is measured, $E_m$, as a function of contrast. The data points are taken from Fig. 4 (Subject D.K.) of Legge and Kersten (1983), and fit the expected function reasonably closely. If the non-linearity of Fig. 7 also allowed for the saturation in amplitude of the signals $R$ and $T$, then a closer fit could be obtained.

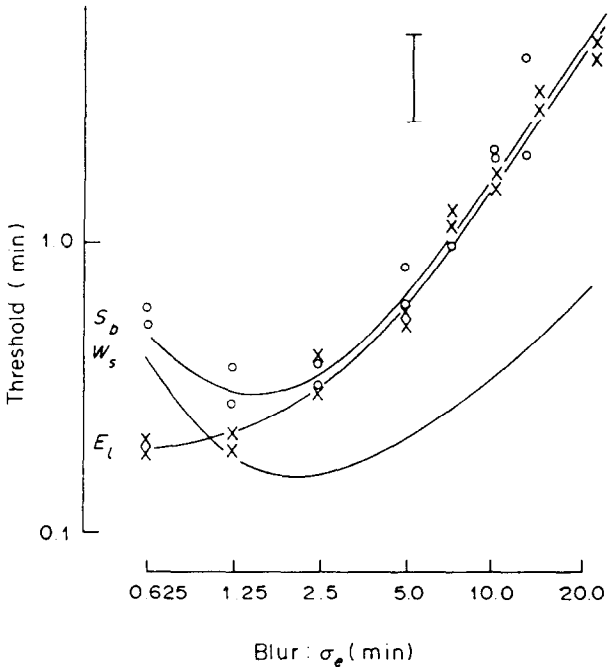


Fig. 12. This figure shows how thresholds for blur difference, $S_b$, and relative location, $E_l$, of edges should vary with blur, according to MIRAGE. The line marked $W_c$ shows how a Weber's law for centroid separation appears in the blur metric. The circles show the blur difference discrimination thresholds for two subjects from Watt and Morgan (1983). The crosses show the relative location thresholds for two subjects from Watt and Morgan (1984). The error bar shows two SEMs.

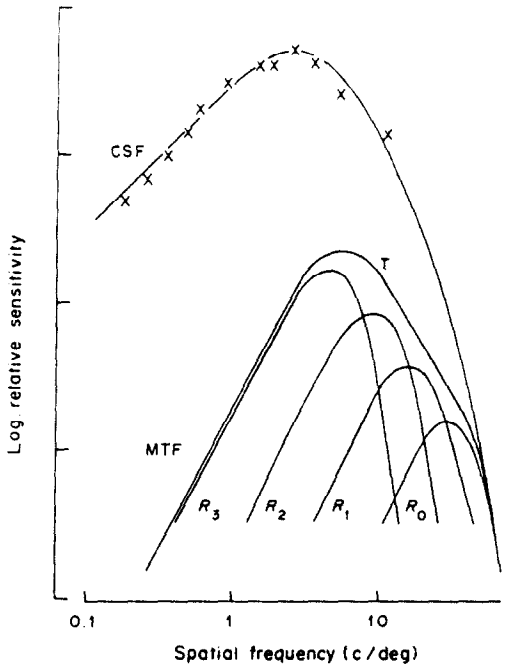


Fig. 13. The contrast sensitivity function (CSF) for MIRAGE, and the amplitude (MTF) of the different filter responses, $R_j$, and their sum, $T$. Four filters are shown at octave intervals covering the entire range, although the density of filters in the spatial frequency domain may well be greater. The modulation transfer function for $T$ is the linear sum of the individual filter MTFs (note that a logarithmic scale obscures this), and not their envelope. The CSF for MIRAGE is similar to that obtained by Campbell and Robson (1968), whose data points are plotted as crosses on the figure. Note that the CSF is not equivalent to the MTF however. This is because the CSF is based on spatial integration between zeroes (to measure mass), whereas the MTF is based on amplitude alone. The improvement in sensitivity and bandwidth is clearly seen.

experiments are sufficiently different to lead to the suggestion that all filters are used for all tasks. It should be noted that, although the figure shows four filters at octave intervals, this is for graphical simplicity, and we are not making any assertions or assumptions about the actual number involved within the overall range.

There is an important principle behind the contrast sensitivity function calculations which is also illustrated on the figure. The filters, marked $R_0$ to $R_3$, have peak frequencies in their modulation transfer functions (MTF: response amplitude to a cosine wave as a function of its spatial frequency) at 27.3 down to 3.4 c/deg. The signals $T^+$ and $T^-$ each have a MTF which is the sum of the $R_j$ MTFs (not their envelope). These signals therefore have a broader bandwidth and a centre frequency at 4.3 c/deg (it should also be borne in mind that they are not linear).

A luminance change becomes visible according to our theory, not when one of these signals reaches some threshold amplitude, but when the mass of a zero-bounded response distribution exceeds the threshold level $m$. Contrast detection sensitivity for cosine waves as a function of spatial frequency (CSF) is thus proportional to the mass in the signals $T^+$ and $T^-$ of MIRAGE. The value of the mass in a cosinusoidal response is the same as the peak amplitude of the integral of the signal. Integrating the signal raises its amplitude in inverse proportion to frequency and therefore results in peak sensitivity being at 2.6 c/deg, lower than the peak response in $R_j$ or $T$. The greater sensitivity and broader bandwidth of MIRAGE, compared with the individual filters is clearly seen in the figure, as is the excellent match with the experimental measurements of Campbell and Robson (1968).

*Theme 3: limiting the amount of information*

The analogue summation of the signals from the different filters is another important aspect of MIRAGE, and arose primarily to limit the amount of information in the system. The filters themselves are of course independent before the analogue summation, and the demonstrations of independent filters do not challenge the theory. If the reason for the analogue summation is to limit information, then the visual system should behave somewhat inappropriately under the rare circumstances when useful information in one frequency band is mixed with irrelevant information in another, such as the Harmon and Julesz (1973) demonstration.

In line with that demonstration, we have found (Watt and Morgan, 1984; Morgan and Watt, 1984) that the precision with which the relative position of a blurred edge (predominantly low-frequency) could be judged was markedly reduced by superimposing an irrelevant patch of high frequency grating over the centre of the edge, and enhanced by superimposing it only at the ends of the edge (and not at the centre). This implies that low frequency filters cannot be accessed independently of the high, and supports the notion of averaging between filters. The same principle applies to blur discrimination (Watt and Morgan, 1983), the rise in which for nearly sharp edges, as Watt and Morgan (1984) have shown, implies an inability to access high frequency information independently of low. The phase-dependent interactions between widely separated spatial frequencies of Arend and Lange (1979) are clearly another instance. Arend and Lange only considered models with thresholds on each filter and then could not explain the results in terms of a broad-band response following narrow ones. MIRAGE has the threshold later in the system and therefore behaves rather as though the broad-band stages, $T$, were in parallel to any independent narrow-band stages.

Whilst these manifestations of the summation across spatial frequencies appear to be disadvantageous, the reduction of information is a considerable saving, as is illustrated by the phenomenon of grouping, which arises as a direct consequence of the rectification and summation. Watt (1985) has suggested that the position of elements within a group is

not accurately (or perhaps not at all) represented in the periphery of vision, leading to a great saving in data rate without seriously affecting perception. This is also similar to the report of Nachmias and Rogowitz (1983) in another study of cross-frequency interactions.

The various filter responses could only be averaged after they had been split into negative and positive portions unless severe loss of resolution were to result. It is therefore interesting and encouraging to learn that the on- and off-centre retinal ganglion cells in the cat branch on different levels of the inner plexiform layer (Nelson *et al.*, 1978; Waessle *et al.*, 1983) and that the on- and off-pathways remain separate in the LGN (Schiller, 1984).

## RECAPITULATION

By way of summary, we shall describe the consequences of the MIRAGE transformation. Some of these are advantageous for a visual system, and some are rather detrimental. Both types of consequence appear to be properties of human vision.

### (*i*) *The benefits of MIRAGE*

MIRAGE takes any (one-dimensional) luminance waveform and explicitly (but symbolically) represents all monotonic changes in luminance. The position of each such event is recorded, as is the spatial extent of the change. To within the d.c. luminance level and nett ramp across the whole field of view, brightness differences between any two points in the field are known. By virtue of using more than one spatial filter, a relatively broad bandwidth is achieved and intrinsic noise is reduced.

In psychological terms, MIRAGE transforms the retinal image to create a representation in which the locus and contrast of each edge and bar is explicit. Furthermore, to a degree, dense textures in such an image are separated into relatively dark regions with associated bright markings, and *vice versa*. This is the beginning of a structured representation, and provides a means of information rate control.

### (*ii*) *The costs of MIRAGE*

Because MIRAGE is based on the propositions of unknown stimulus and information limitation, there are, in an abstract and theoretical sense, a number of circumstances where MIRAGE appears to behave in a less than ideal fashion. Examples are the brightness errors and the non-linear transformations of edge blur and luminance contrast, although these non-linearities could be compensated for by appropriate inversion, based on either some limited knowledge about the system (e.g. the largest filter space constant for blur) or some information about the expected distributions of blur and contrast in the image. These are all properties of the human visual system and rarely lead to sensory or perceptual difficulties. Indeed, it is only by noting these limitations, such as the non-linearities in the measurement of blur and contrast for example, that MIRAGE can be tested by psychophysical data. Many of the cross-frequency interactions could be regarded as costs of the MIRAGE transformation, and it is easy to conceive circumstances where independent access to each filter would be useful. It is our thesis that such access would be, on balance, a disadvantage because of the information rate that it implies.

## CODA

We close by posing two critical questions that distinguish MIRAGE from other schemes.

To specify fully the filter output space requires two co-ordinate systems: one for retinotopic position and one for filter characteristics. Thus any subsequent analysis could use two different types of information: response amplitude as a function of retinotopic space and as a function of the filter space.

The two apparently different schemes of Fourier analysis (Campbell and Robson, 1968; DeValois and DeValois, 1980; Robson, 1983) and zero-crossing analysis (Marr and Poggio, 1979; Marr *et al.*, 1979; Marr and Hildreth, 1980) each use both of these sources of information and in a similar fashion. In the Fourier theory, response amplitude as a function of filter scale is used to derive the power spectrum, and response amplitude as a function of position is used (at each scale) to determine the phase. In the zero-crossing theory, response amplitude as a function of position is used (at each scale) to determine the location of edges, and response amplitude as a function of scale is used to determine the degree of blur.

These two schemes are also in accord in that a discrete code is generated for each filter output, either as a discrete spatial frequency transform, or as a list of zero-crossing occurrence.

The two greatest differences between MIRAGE and these other schemes (and many like them) are drawn by asking the following two questions:

> "Is note taken of the centre frequency of each filter?"

In MIRAGE it definitely is not. Indeed, in principle, the filters do not need to be consistent across time and space: provided that each point in the retinal image is covered by a reasonable spread in the frequency domain, the properties of the individual filters are unimportant. In terms of evolution and development, adding extra filters to widen the bandwidth can be achieved without any reorganization of the higher analytic process. Likewise, the demise of filters with old age or under-use will not affect the competence of the remaining system.

> "Is a separate symbolic code generated for the continuous output of each filter?"

Once again, in MIRAGE, definitely not. By postponing the transition to a symbolic code until the filter signals have been combined, a simplification is

achieved (compare the three rules of MIRAGE with the spatial co-incidence and various other parsing rules of Marr and Hildreth, 1980).

*Acknowledgements*—This work was supported by grants G979/870/N and G822/0517/N from the Medical Research Council. We thank H. B. Barlow, G. Legge and D. Pelli for illuminating discussions. Many other colleagues have helped us in the expression of our ideas, and we are most grateful to them. We are also grateful to the referees whose incisive comments have helped us to improve the text.

## REFERENCES

Arend L. E. and Lange R. V. (1979) Phase-dependent interaction of widely separated spatial frequencies in pattern discrimination. *Vision Res.* **19,** 1089–1092.

Barlow H. B. (1979) Reconstructing the visual image in space and time. *Nature* **279,** 189–190.

Barrow H. G. and Tenenbaum J. M. (1978) Recovering intrinsic scene characteristics from images. In *Computer Vision Systems* (Edited by Hanson A. and Riseman E.), Academic Press, New York.

Binford T. O. (1981) Inferring surfaces from images. *Artif. Intell.* **17,** 205–244.

Blakemore C. B. and Campbell F. W. (1969) On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *J. Physiol.* **203,** 237–260.

Braddick O., Campbell F. W. and Atkinson J. (1978) Channels in vision: basic aspects. In *Handbook of Sensory Physiology* (Edited by Held R., Leibowitz H. W. and Teuber H. L.), Vol. 8. Springer, Berlin.

Campbell F. W. and Robson J. G. (1968) Application of Fourier analysis to the visibility of gratings. *J. Physiol.* **197,** 551–566.

Crick F. H. C., Marr D. C. and Poggio T. (1980) An information processing approach to understanding the visual cortex. In *The Cerebral Cortex* (Edited by Schmidtt F. O. and Worden F. G.). MIT Press, Cambridge, Mass.

Daugman J. G. (1984) Spatial visual channels in the Fourier plane. *Vision Res.* **24,** 891–910.

DeValois R. L. and DeValois K. K. (1980) Spatial vision. *Ann. Rev. Psychol.* **31,** 309–341.

Foley J. M. and Legge G. E. (1981) Contrast detection and near-threshold discrimination in human vision. *Vision Res.* **21,** 1041–1053.

Geisler W. S. (1983) Mechanisms of visual sensitivity: backgrounds and early dark adaptation. *Vision Res.* **23,** 1423–1432.

Graham N. and Nachmias J. (1971) Detection of grating patterns containing two spatial frequencies: a comparison of single-channel and multiple-channel models. *Vision Res.* **11,** 251–259.

Harmon L. and Julesz B. (1973) Masking in two-dimensional recognition: effects of two-dimensional filtered noise. *Science* **180,** 1194–1197.

Laming D. R. J. (1982) Differential coupling of sensory discriminations inferred from a survey of stable decision models for Weber's Law. *Br. J. Math. Stat. Psychol.* **35,** 129–161.

Legge G. E. (1982) A power law for contrast discrimination. *Vision Res.* **21,** 457–467.

Legge G. E. and Kersten D. (1983) Light and dark bars: contrast discrimination. *Vision Res.* **23,** 473–483.

Marr D. (1976) Early processing of visual information. *Phil. Trans. R. Soc.* **B275,** 483–524.

Marr D. and Hildreth E. C. (1980) A theory of edge detection. *Proc. R. Soc. Lond. B* **207,** 187–217.

Marr D. and Poggio P. (1979) A computational theory of human stereo vision. *Proc. R. Soc. Lond.* **204,** 301–328.

Marr D., Poggio T. and Ullman S. (1979) Bandpass channels, zero-crossings, and early visual processing. *J. opt. Soc. Am.* **69,** 914–916.

Mayhew J. E. W. and Frisby J. P. (1981) Psychophysical and computational studies leading towards a theory of human stereopsis. *Artif. Intell.* **17,** 349–387.

Morgan M. J. and Watt R. J. (1984) Spatial frequency interference effects and interpolation in vernier acuity. *Vision Res.* **24,** 1911–1919.

Nachmias J. and Rogowitz B. E. (1983) Masking by spatially modulated gratings. *Vision Res.* **23,** 1621–1629.

Nachmias J. and Sansbury R. (1974) Grating contrast: Discrimination may be better than detection. *Vision Res.* **14,** 1039–1042.

Nelson R., Famigletti E. V. and Kolb H. (1978) Intracellular staining reveals different levels of stratification for On- and Off-centre ganglion cells in cat retina. *J. Neurophysiol.* **41,** 472–483.

Rentschler I. and Treutwein B. (1985) Loss of spatial phase relationships in extrafoveal vision. *Nature* **313,** 308–310.

Robson J. G. (1983) Frequency domain visual processing. In *Physical and Biological Processing of Images* (Edited by Braddick O. and Sleigh A.). Springer, Berlin.

Rosenfeld A. (1983) "Intrinsic images": deriving three-dimensional information about a scene from single images. In *Fundamentals in Computer Vision* (Edited by Faugeras O. D.). Cambridge Univ. Press.

Ross J., Holt J. J. and Johnstone J. R. (1981) High frequency limitations on Mach bands. *Vision Res.* **21,** 1165–1167.

Schiller P. H. (1984) The connections of the retinal on and off pathways to the lateral geniculate nucleus of the monkey. *Vision Res.* **24,** 923–932.

Tolhurst D. J., Movshon J. A. and Thompson I. D. (1981) The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. *Expl Brain Res.* **41,** 414–419.

Waessle H., Peichl L. and Boycott B. B. (1983) A spatial analysis of on- and off-ganglion cells in the cat retina. *Vision Res.* **23,** 1151–1160.

Watt R. J. (1984) Towards a general theory of the visual acuities for shape and spatial arrangement. *Vision Res.* **24,** 1377–1386.

Watt R. J. (1985) Structured representations in low-level vision. *Nature* **313,** 266–267.

Watt R. J. and Morgan M. J. (1983) The recognition and representation of edge blur: evidence for spatial primitives in human vision. *Vision Res.* **23,** 1457–1477.

Watt R. J. and Morgan M. J. (1984) Spatial filters and the localization of luminance changes in human vision. *Vision Res.* **24,** 1387–1397.

Wilson H. R. and Bergen J. R. (1979) A four mechanism model for spatial vision. *Vision Res.* **19,** 19–32.